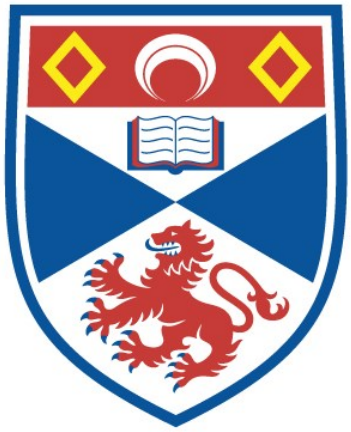
Behavioural responses by seals to offshore energy activities

Katherine Fae Whyte

A thesis submitted for the degree of PhD
at the
University of St Andrews



2022

Candidate's declaration

I, Katherine Fae Whyte, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 52,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in January 2017.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date 15/10/21 Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date 15/10/21 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Katherine Fae Whyte, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

Embargo on all of print copy for a period of 3 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for printed embargo request

Publication of the thesis would hinder further publication of this work. The work presented in this thesis is also of high commercial interest and sensitivity (to offshore energy developers), and so publishing the thesis prior to a scientific peer-reviewed publication is undesirable.

Electronic copy

Embargo on all of electronic copy for a period of 3 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for electronic embargo request

Publication of the thesis would hinder further publication of this work. The work presented in this thesis is also of high commercial interest and sensitivity (to offshore energy developers), and so publishing the thesis prior to a scientific peer-reviewed publication is undesirable.

Title and Abstract

- I agree to the title and abstract being published.

Date 15/10/21 Signature of candidate

Date 15/10/21 Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Katherine Fae Whyte, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date 15/10/21 Signature of candidate

Acknowledgements

First of all, I would like to thank my supervisors: Gordon Hastie, Debbie Russell, Len Thomas, and Carol Sparling. Thank you for your unwavering encouragement, patience, trust, and support over the past four years. Thank you for always helping me see the bigger picture, as well as carefully consider all the important scientific details. I could not imagine having a better team.

Special thanks also go to my "lab" group (DanGER), and all members past and present: Gordon, Debbie, Joe, Matt, James, Clair, Chris, Laura, "honourary member" Monica, Lauren, Izzy, Hannah, Claire, Jules. Thank you for being such a knowledgeable and genuine support bubble, especially during the uncertainty of a global pandemic and lockdown. Thanks also to all of my office mates in my different offices over the years: the old top-floor SMRU office, the new SOI open-plan office, the CREEM office, and of course my working-from-home office. Only one of these offices had a cat (who is responsible for any typos found in this thesis), but in all of these offices I was lucky enough to work with people who made day-to-day PhD life an enjoyable experience.

To everyone at SMRU and CREEM, thank you for always making me feel welcome. I cannot even begin to list names here, partly out of fear I will forget someone, but also because I struggle to think of anyone who has not helped shape my career in some way. Thank you to companions at (often much-needed) lunch, coffee, and cake breaks. In particular, thank you to those who offered and supported me in opportunities to learn and grow as a scientist. I am grateful to have gained these experiences, and all of them have helped to make me a better scientist: from teaching and public engagement, to collaborative research projects, to fieldwork tagging seals, to publishing my first paper, to attending (and organising) conferences. In particular, special thanks to Catriona Harris and Stacy DeRuiter for many long chats about Mahalanobis distance.

Above all, thank you to my family. Thank you for laughs, for adventures, for always keeping me sane, and for reminding me that a PhD is only a small part of life.

Funding

This work was supported by the Department for Business, Energy and Industrial Strategy's (BEIS, UK) Offshore Energy Strategic Environmental Assessment programme (OESEA-16-74), administered through Hartley Anderson Ltd.

This work was supported by the University of St Andrews (School of Biology).

Thanks also to:

- Race Bank Offshore Wind Farm Ltd for funding the sound propagation modelling.
- Funders of the original data collection in 2012:
  - Harbour seal (pv42) tag deployment in The Wash: Tags and their deployment were funded by the Department for Business, Energy and Industrial Strategy's (formerly the Department of Energy and Climate Change, UK) Offshore Energy Strategic Environmental Assessment programme (OESEA-11-24), with additional resources from National Capability funding from the Natural Environment Research Council to the Sea Mammal Research Unit (NE/R015007/1).
  - Harbour seal (pv40) tag deployment in The Thames: Tags and their deployment were in collaboration with ZSL, funded by SITA Trust and BBC Wildlife Fund.

Data and collaborators

Thank you to:

- Bas Binnerts and Sander von Benda-Beckmann at TNO (Acoustics and Sonar Expertise Group, TNO, The Hague, Netherlands) for producing the acoustic propagation models of pile driving used in this thesis.
- Centrica plc for the provision of detailed pile driving data at Lincs wind farm.
- SMRU Instrumentation and those who helped with any aspect of the original data collection and seal tagging in 2012.

All capture and handling protocols were carried out under UK Home Office License 60/4009 in accordance with the Animals Scientific Procedures Act 1986, with additional licence approval from the University of St Andrews Animal Welfare and Ethics Committee. Seals were captured under licence from Marine Management Organisation. All appropriate permissions with regard to designated sites, with any necessary mitigation measures in place, were obtained from the relevant authority. The data used in this thesis are published in Hastie et al., 2015, Russell et al., 2016, and Whyte et al., 2020.

Abstract

To effectively manage interactions between industrial activities and wildlife populations in increasingly urbanised environments, it is essential to understand how animals may be affected by different anthropogenic activities. In this thesis, I used biologging devices to investigate the potential effects of sound disturbance on seals. By simulation study, I evaluated the use of statistical tools (Mahalanobis distance) to detect unusual instances of movement and dive behaviour in seal biologging data. The results of these simulations were used to produce recommendations for future studies aiming to detect behavioural changes in animal movement data. Building on the findings of this work, I examined the movement and dive behaviour of 24 harbour seals (*Phoca vitulina*) during pile driving construction at an offshore wind farm in the UK. Using GPS location data collected on animal-borne tags, I identified statistically unusual horizontal movement events during pile driving, typically consisting of increases in speed, the cessation of horizontal movement, or the sudden initiation of travel. Using dive data from animal-borne tags, I identified statistically unusual groups of dives, and also characterised the effect of pile driving activity on behaviour-switching between different dive types (by hidden Markov models). Seals were found to switch dive behaviours more often during pile driving (compared to baseline periods), and the identified unusual dives were typically shorter and shallower, with longer post-dive surface intervals. For both horizontal and diving responses, dose-response curves were produced to estimate the relationship between the received sound level of pile driving and the probability of a behavioural change. By examining behaviour at the individual-level, improved insights of at-sea seal behaviour during disturbance events were gained. The results of this thesis also inform future offshore activities, enabling the renewable energy industry to develop in a timely and environmentally-responsible manner.

Contents

Chapter 1

General Introduction

Part of the work in this chapter (See Appendix F for full paper) is published as:

Whyte, K.F., Russell, D.J.F., Sparling, C.E., Binnerts, B., & Hastie, G.D. (2020). Estimating the effects of pile driving sounds on seals: Pitfalls and possibilities. *The Journal of the Acoustical Society of America*, 147(6): 3948-3958.

1.1 Animal behaviour and human interactions

Animals must make decisions as they move through and interact with their environment (Davies et al., 2012; Owen et al., 2017). They must decide where to go, what to eat, when to rest, and which conspecifics to interact with. The consequences of these decisions may affect their immediate survival, their long term health and condition, and their breeding success. To make successful decisions, sensory cues are fundamental for informing them of particular resources or risks within their environment (Sih, 2013; Smith et al., 2021). As human presence and activities continue to expand across the globe (Venter et al., 2016), humans are modifying the physical environment and producing sensory pollution (Halfwerk and Slabbekoorn, 2015) which may alter or impact animal decision-making.

Sensory pollution is often produced as a by-product of human activities, and can be visual, acoustic or chemical (Dominoni et al., 2020). This pollution may affect animals by masking the signals of other important stimuli, distracting or disrupting the animal from its intended behaviour, or eliciting animals to interpret and react to these signals as if they were natural cues (Barber et al., 2010; Dominoni et al., 2020). For example, marine turtle hatchlings can be disorientated by artificial light near breeding sites (Kamrowski et al., 2013), great tits (*Parus major*) spend more time on vigilance behaviours and less time feeding during exposure to aircraft noise (Klett-Mingo et al., 2016), and moths (*Manduca sexta*) can have a reduced ability to locate flowers when background odours are present (Riffell et al., 2014). Over time, such individual behavioural changes have the potential to reduce the fitness of individuals, populations, or species, for example through missed foraging opportunities or maladaptive behaviour. Wider ecosystem functions and processes (e.g. trophic cascades, nutrient cycling, seed dispersal) may also be disrupted if animal behaviours, distributions or abundances are significantly altered (Wilson et al., 2020). Understanding how and why animals may alter their behaviour in response to human activity is therefore an important step in predicting how populations and ecosystems may be affected by human activities, and in determining how mitigation may be effectively used (Smith et al., 2021).

1.2 Urbanisation of the marine environment

Over the last century, technological advances have enabled a vast increase in the type and extent of human activities being conducted in the marine environment. Although it is likely that only some of these activities will have acute impacts on marine wildlife

(e.g. the removal of individuals) or habitats (e.g. degradation), many will produce sound pollution which is released into the marine environment. The number of ships worldwide increased by a factor of 3.5 between World War II and 2008 (Erbe et al., 2019; Frisk, 2012), and shipping lanes now extend across the globe, including regions of the Arctic previously inaccessible due to ice (Smith and Stephenson, 2013). This expansion has led to commercial shipping now being considered as a major source of low-frequency (5–500 Hz) background noise in the world's oceans (Duarte et al., 2021; Hildebrand, 2005). In the fishing industry, substantial noise pollution is produced from bottom trawling activity (Daly and White, 2021), and in aquaculture acoustic deterrent devices are used to deter seals from farming facilities (Findlay et al., 2018). Energy extraction activities, in particular seismic surveys to locate oil and gas reserves, use high intensity low-frequency sounds to detect and map the extent of resources underwater (Estabrook et al., 2016). Seismic surveys are thought to be a significant component of underwater noise in the North Atlantic (Nieukirk et al., 2004). In addition, the marine environment is also widely used for tourism and recreation (Marley et al., 2017), military activities (Goldbogen et al., 2013), construction (Culloch et al., 2016), and harvesting and mining of natural resources (Christiansen et al., 2020), all of which produce noise from increased vessel activity, sonar, drilling, or other areas of activity.

For aquatic animals, sound is an important sensory modality. In marine habitats, the availability of light reduces with depth, and water turbidity limits underwater vision (e.g. during extreme weather, or in areas with high silt). Relative to air, sound travels approximately 4–5 times faster in seawater ($\sim$1,500 ms^{-1} vs. $\sim$340 ms^{-1}) and can travel long distances underwater. For example, the calls of blue whales (*Balaenoptera musculus*) can be detected from over 200–600 km away (Širović et al., 2007; Stafford et al., 1998). Sound propagation underwater depends on a variety of environmental conditions including water temperature, depth (pressure) and salinity, as well as the frequency characteristics of the sound itself (Au and Hastings, 2008). Many marine animals are highly reliant on their ability to make and detect sounds in order to communicate, navigate and interact with their environment. The production of additional sound from human activities therefore presents a potential environmental concern for marine habitats, as many of the sounds that are produced overlap highly with the sound frequencies heard, produced, and used as cues and signals by marine animals (Figure 1.1). The extent to which wildlife is affected will depend on the characteristics of the sound source (e.g. loudness, frequency), the pattern of its use, and the spatial and temporal overlap of these sounds with marine animal vocalisations, behaviours, and distributions. As marine habitats are increasingly being filled

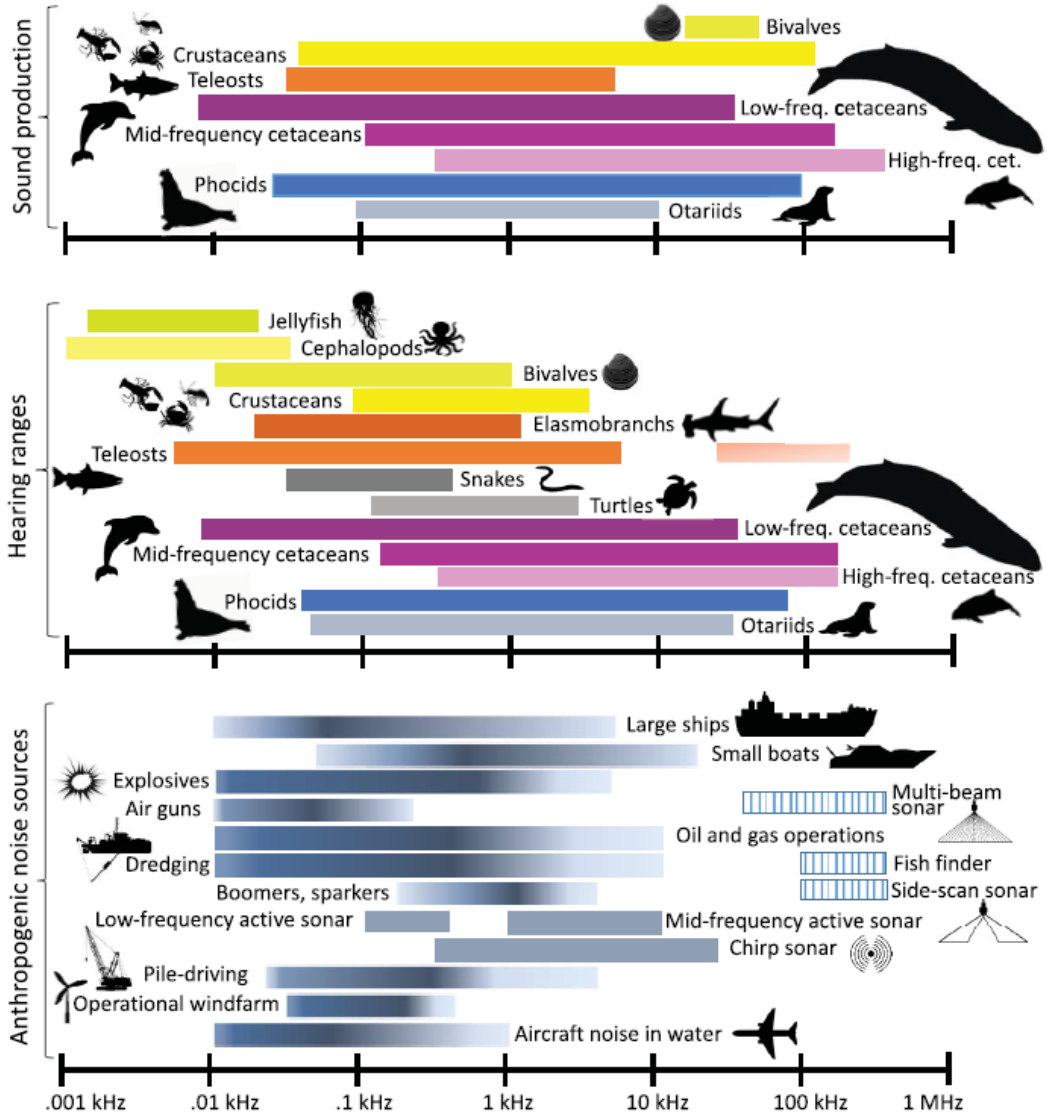


Figure 1.1 Sources and receivers of sound in the marine environment. Shown are the approximate sound frequency ranges of sound production by different animal taxa, the estimated hearing ranges of each animal group, and the frequency ranges of sound produced from selected anthropogenic noise sources. Colour shading denotes the dominant frequencies of the anthropogenic sound sources. Figure from Duarte et al. (2021).

with sounds from human activities (Duarte et al., 2021), it is important to document and understand the potential impacts of anthropogenic sounds, both as individual stressors, and as additional contributions to an already complex ocean soundscape.

1.3 The transition to renewable energy

1.3.1 Renewable energy trends

The need to combat climate change is driving a global transition from carbon-based energy (oil, coal, natural gas) to renewable energy sources (e.g. wind, solar, geothermal, hydropower). In the decade of 2011–2020, global surface temperatures were 1.09°C higher than the 1850–1900 average, and it is predicted that global warming will exceed 2°C by 2100 unless severe reductions in global greenhouse gas emissions are made (IPCC, 2021). Concern regarding the consequences of this warming for global climate, extreme weather events, sea level rise and ecosystem functioning (IPCC, 2021; Walther, 2010) is increasing pressure on countries to reduce their emissions and switch to more environmentally-sustainable sources of energy. For example, the United Kingdom (UK) is committed to reaching net-zero emissions by 2050, and reaching a 40 GW capacity (equivalent to the electricity required to power all UK households) in offshore wind by 2030 (The United Kingdom Government, 2020). Worldwide, although the use of renewable energy sources has been increasing annually over the last half-century (Figure 1.2(a)), the rate of increase in total renewable energy generation has been substantially higher within the last 20 years (average annual increase of ∼54 GW pre-2003 vs. ∼268 GW post-2003; Ritchie and Roser, 2020).

1.3.2 Offshore wind and pile driving

The wind energy sector has been one of the fastest growing renewable energy industries in the last 20 years (Figure 1.2(b)), and is increasingly moving towards offshore developments, despite the increased financial costs of construction and operation in these environments (Breton and Moe, 2009). Wind farms built offshore can have larger arrays, larger turbines, benefit from stronger and more predictable offshore winds, and can be built with reduced limitations caused by visual and operational noise concerns from nearby residents (Breton and Moe, 2009; Díaz and Guedes Soares, 2020). With this expansion of the sector comes the urgent need to understand how these new technologies may affect marine habitats and

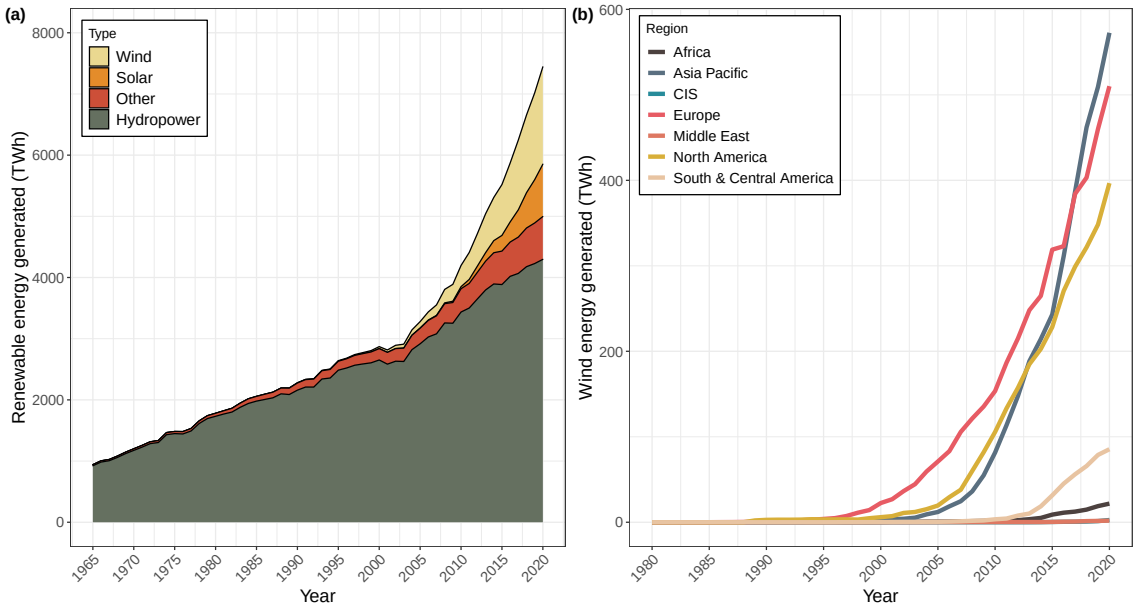


Figure 1.2 Trends in renewable energy generation from 1965 to 2020. (a) Total annual global renewable energy generated by source type from 1965–2020. "Other" includes renewable energy generated from geothermal, biomass, waste, wave and tidal sources. (b) Annual wind energy generated by geographic region from 1980–2020, including both onshore and offshore wind. "CIS" refers to the Commonwealth of Independent States and has a similar trend to that of the Middle East region. Data from Ritchie and Roser (2020) and BP (2021).

.

marine wildlife. Environmental impact assessments (EIAs) are conducted to predict the potential impacts of each development, but these assessments require robust information on the effects of construction and operation on each species. For example, turbines (both onshore and offshore) present a potential risk of direct collision between the fast-moving turbine blades and in-flight bats (Ahlén et al., 2009; Arnett et al., 2008) and birds including migrants and seabirds (Furness et al., 2013; Hüppop et al., 2006).

For marine animals that spend time underwater, a key environmental concern from offshore wind is underwater noise. During operation, offshore wind turbines produce a near-continuous low-frequency (< 1 kHz) sound which is radiated into the water (Tougaard et al., 2008; Yang et al., 2018). As the operational sound produced is of low intensity (< 120 dB re 1 μPa (RMS) at ∼100 m from the turbine), it is considered unlikely that this sound has a significant effect on marine mammals (Madsen et al., 2006). Notably, Yang et al. (2018) reported that the operational noise produced by two types of turbine was so low-level that it was challenging to measure and was often masked by wave and tidal flow noise. Underwater noise is highest during the construction phase of the wind farm, which can encapsulate a wide range of activities including increased boat activity, dredging and pile driving (Madsen et al., 2006). Pile driving, used to hammer wind turbine foundations into the seafloor, produces very high intensity sounds (up to 250 dB re 1μPa @ 1 m (peak-peak); Bailey et al., 2010). Given that pile driving to install a single turbine can require repeated strikes (approximately every 1–2 seconds) for multiple hours (Bailey et al., 2014; Graham et al., 2019), understanding of how this sound may affect the hearing (Hastie et al., 2015), behaviour (Brandt et al., 2011), and long-term population health (Thompson et al., 2013) of marine animals present in the vicinity of the wind farm is urgently required.

1.4 Seals: ecology, behaviour, and population status

Pinnipeds (comprising the true seals (Phocidae), eared seals (Otariidae), and walruses (Odobenidae)) are long-lived mammals with a large body size and a semi-aquatic lifestyle (Berta, 2018). The life of a pinniped is divided between time spent hauled out on land (or ice), and time spent at-sea. In this thesis, the focus is largely on investigating the at-sea portion of pinniped lifecycles. At sea, pinnipeds must acquire the food they need to survive and may encounter a variety of threats including natural predators and human activities. This is typically when pinnipeds are also hardest to observe directly. Understanding at-sea

behaviour is therefore a challenging but important area of research; however, it is essential to remember that the requirement of pinnipeds to regularly haulout (e.g. to rest, breed, digest) remains an intrinsic part of their movement ecology, distribution and life history.

1.4.1 Movement ecology of seals

Seals are considered to be central-place foragers (Bell, 1990), making foraging trips out to sea from a central location on land. In practice, while seals typically show high site-fidelity to haulout sites (or groups of sites), seals can change haulout locations during their lifetime (Dietz et al., 2013; Sharples et al., 2012). The distance that seals travel during these trips can be 10's to 1000's of km, with differences between species and between geographic regions. For example, southern elephant seals (*Mirounga leonina*) in South Georgia have been observed to travel up to ∼3,000 km away from their breeding beaches (McConnell and Fedak, 1996). In contrast, a study of harbour seals (*Phoca vitulina*) in west Scotland (Cunningham et al., 2009) reported maximum trip extents of only 46 km from the haulout site, with trips typically lasting from 12–24 hours; however, there is considerable variation in harbour seal foraging trip behaviour between different regions of the UK (Sharples et al., 2012).

1.4.2 Diving behaviour of seals

While at sea, seals spend the majority of their time underwater (Bekkby and Bjørge, 2000; DeLong and Stewart, 1991). The physiology of pinnipeds is highly adapted to enable this, with animals possessing thick layers of blubber to aid in thermoregulation, large oxygen stores and specialised cardiovascular systems to cope with long dives and high pressure at depth, and sensory systems that can function well both in-air and underwater (Blix, 2018; Crocker and Champagne, 2018). Pinnipeds must dive to locate and capture their prey, but many species also dive while travelling to foraging locations and during low-activity periods in the water. Within each species, the presence of dives with different biological functions leads to the observation of dives with different measured properties, such as variable dive depths, durations, and shapes (e.g. Baechler et al., 2002; Kuhn et al., 2009; Photopoulou et al., 2014). This third (vertical) dimension of seal movement is therefore an important consideration when attempting to understand the patterns and drivers of seal behaviour.

1.4.3 Harbour seals in the UK

Harbour seals are one of two pinniped species that breed in the UK, alongside the larger and more abundant grey seal (*Halichoerus grypus*) (Thomas et al., 2019; Thompson et al., 2019). Mature adult male harbour seals can reach maximum body lengths of approximately 1.6 m and weights of ∼75–104 kg (Teilmann and Galatius, 2018). Females are typically smaller, reaching maximum sizes of approximately 146 cm and 67–83 kg (Teilmann and Galatius, 2018). Both sexes reach sexual maturity at 4–6 years, with maximum recorded ages of 31 and 36 for males and females respectively (Härkönen and Heide-Jørgensen, 1990; Teilmann and Galatius, 2018). In the UK, the harbour seal pupping and breeding season lasts from June to July (Thompson et al., 1997, 2019), during which time females give birth on land to a single pup. The lactation period lasts an average of ∼24 days, with females continuing to forage during this period and harbour seal pups being able to swim from birth (Arso Civil et al., 2021; Thompson et al., 1994).

The estimated total population size of harbour seals in the UK is 43,450 (95% CI: 35,550–57,900; Thompson et al., 2019). Approximately 80% of the UK population is in Scotland, but significant numbers are also found in south-east England and in Northern Ireland (Thompson et al., 2019). In particular, aerial surveys of the south-east England Seal Management Unit area (Figure 1.3), which includes The Wash & North Norfolk Special Area of Conservation (the focus of the tagged seals in this thesis), counted 5,061 harbour seals at haulout sites in the most recent survey in 2016 (Thompson et al., 2019). Despite populations of harbour seals increasing in the south-east of England, there have been substantial regional declines (east Scotland, Northern Isles) in some harbour seal populations (Lonergan et al., 2007; Thompson et al., 2019). The cause of these regional declines is not fully understood; however, there are several possible factors including competition with grey seals (Wilson and Hammond, 2019), predation (Brownlow et al., 2016), and algal toxins (Jensen et al., 2015). Given the projected increases in offshore wind developments around UK coastlines, it is vital to understand how increased industrial activity may affect this protected species, particularly for populations which may already be under pressure.

1.5 Biologging as a tool to track individuals

Biologging, the practice of recording and/or relaying data from animal-attached tags (Hooker et al., 2007), is a rapidly growing area of scientific research (Rutz and Hays, 2009). Developments in the miniaturisation of sensors and mobile electronic devices are

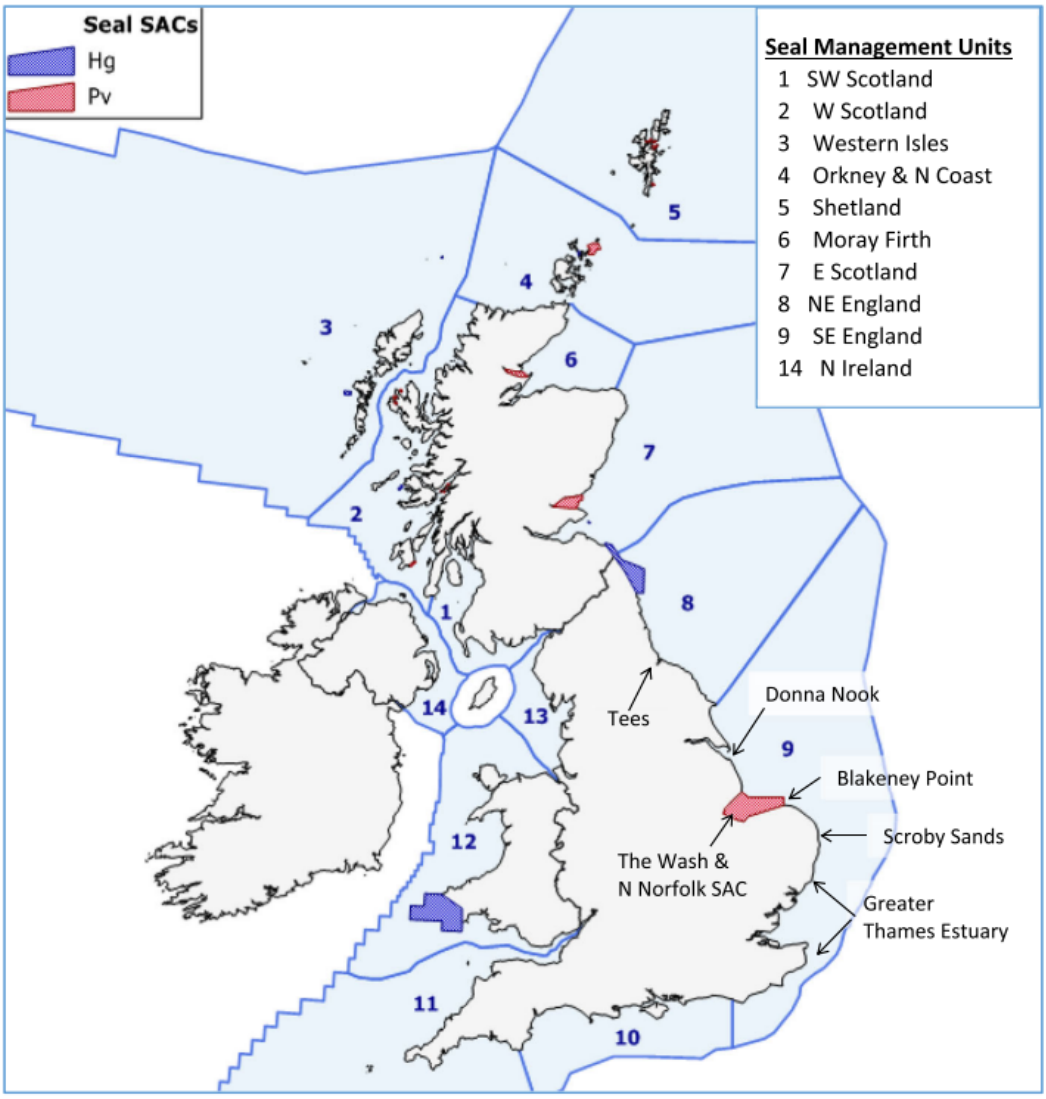


Figure 1.3 Map of Seal Management Units in the UK for both harbour (Pv) and grey (Hg) seals. Also shown are Special Areas of Conservation (SACs) for which seals were a primary reason for the protected site. Figure from Thompson et al. (2019)

.

enabling scientists to attach devices to animals to record different aspects of their movement, behaviour, physiology, and environment (Wilmers et al., 2015). These developments have been particularly important for aquatic environments, as tags can now record data on animals which were previously challenging to observe, e.g. during diving or extended periods at-sea (Hooker et al., 2007; Hussey et al., 2015).

As biologging devices are deployed on free-ranging animals, careful consideration must be given to how the recorded data will be recovered at the end of the study. The first option is to store all recorded data on the tag (often termed "archival" tags; Cooke et al., 2004), and recover the physical tag itself. In practice, guaranteeing the recovery of tags can be challenging, in particular for studies with long-term deployments (months–years), wide-ranging species, or inaccessible habitats. The second option is therefore to deploy tags which are able to record and remotely transmit data recorded on the animal back to a receiver (often termed "telemetry" tags). Multiple options now exist for transmitting data, including VHF radio, ARGOS satellites, and the GSM phone network (Hooker et al., 2007). The need to transmit data often places constraints on the type, resolution, and volume of data that can be reliably recovered from the telemetry tag (Carter et al., 2016). Therefore, data collected from telemetry tags may need to be pre-processed on-board the tag to summarise the data before transmission (e.g. Cox et al., 2018; Photopoulou et al., 2015). Understanding the limitations of this abstracted data is an important consideration when interpreting telemetry data; however, in practice, all biologging tags must necessarily be a careful compromise between obtaining data appropriate for the study goals (e.g. data resolution, study duration, number of individuals) and technological constraints (e.g. battery life, data storage, physical size of the tag) (Holton et al., 2021).

One of the main benefits of using biologging devices is the ability to continuously track individuals while they interact with their environment, without the requirement or potential disturbance of human observers (Cooke et al., 2004). These devices therefore also enable detailed data collection on how wild animals interact with human activities in their environment. For example, Ordiz et al. (2019) investigated the responses of wild brown bears (*Ursus arctos*) in Sweden to groups of human hikers, using GPS devices to track the movements of both bears and humans. Two common concepts in behavioural biology were also investigated: habituation and sensitisation (Peeke, 1984). During repeated exposures to a stimulus, habituation is considered to occur when the occurrence or severity of the animal's response to the stimulus decreases over time (McSweeney et al., 1996). In contrast, sensitisation occurs when the occurrence or severity of the animal's response increases over time (McSweeney et al., 1996). During the experimental trials in Ordiz et al. (2019), bears

consistently reacted to human presence by moving away from the hikers, with no evidence of sensitisation or habituation during repeated trials (n≤8) on the same animals (n=29) over 1–3 months. Tracking the bears' movements continuously during the study revealed that the increased movement activity observed during the responses to hikers was followed by a period of 4–5 hours where bears exhibited reduced movement (relative to baseline), highlighting the potential longer-term effects of disturbance on their movement activity. Pirotta et al. (2018b) modelled the movement behaviour of GPS-tagged northern fulmars (*Fulmarus glacialis*) in relation to fishing boat activity off the coast of Scotland. It was estimated that fulmars may be attracted to fishing vessels up to 35 km away, potentially using olfactory cues to inform them of scavenging opportunities from discarded fish and offal. The use of biologging devices in Pirotta et al. (2018b) enabled animal movements to be studied over geographic scales which would be impossible to achieve by other means. Many studies also use biologging devices to monitor how animals move through human-altered landscapes (e.g. urban environments, fragmented habitats, roads). The interest in these studies has largely focussed on animals in terrestrial ecosystems (e.g. large cats, Tigas et al., 2002; squirrels, Stevenson et al., 2013; rattlesnakes, Tracey et al., 2005), where human developments are perhaps most apparent.

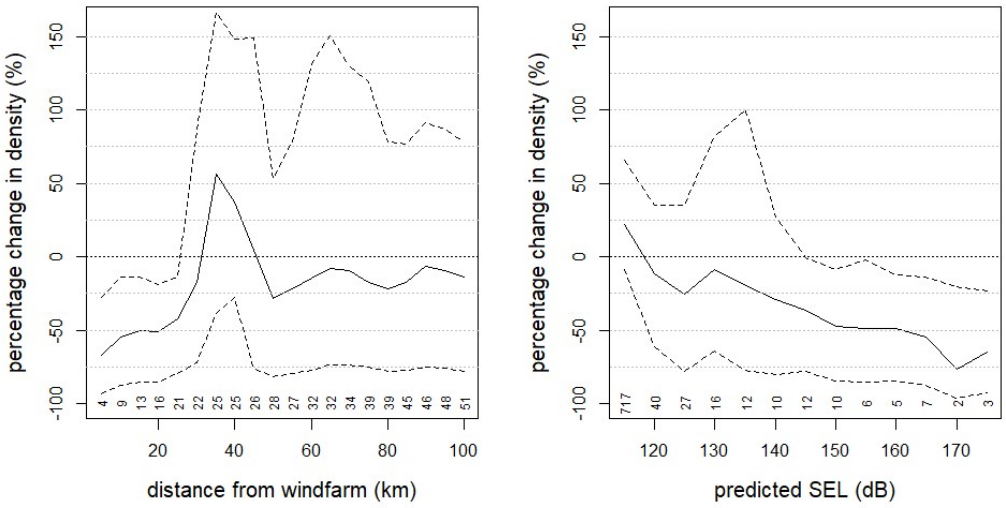
1.6 Seals and noise: from population to individual

To date, research into the effects of noise on seals has largely focussed on the analysis of visual observations (e.g. Harris et al., 2001), surfacing and haul-out behaviour (e.g. Edrén et al., 2010), abundance estimates (e.g. Culloch et al., 2016) and captive experiments (e.g. Kastelein et al., 2006). Biologging devices, despite being first deployed on wild pinnipeds over 30 years ago (Kooyman, 1964; Thompson et al., 1991), have had a relatively limited use in studies investigating the potential effects of anthropogenic sound. These technologies present an often undervalued resource in further understanding how individual seals may move and behave in response to human activities. Some studies have conducted playback experiments to wild tagged seals, investigating the behaviour and distribution of seals when exposed to artificial tidal turbine sounds (Hastie et al., 2018), predator calls (Fregosi et al., 2016; Gordon et al., 2019), and a variety of other acoustic deterrent and human-associated sounds (Costa et al., 2003; Fregosi et al., 2016; Gordon et al., 2019). Within the last 5–10 years, biologging data has been used to examine the effects of some human activities and sound *in situ*, with studies predicting the sound exposure of shipping noise to tagged grey and harbour seals (Jones et al., 2017; Trigg et al., 2020) and the movement

and distribution of tagged harbour seals in the presence of operational tidal turbines (Joy et al., 2018; Onoufriou et al., 2021; Sparling et al., 2018).

For pile driving sounds in particular, biologging devices have been used to investigate the predicted sound exposure (Hastie et al., 2015; Whyte et al., 2020) and density (Russell et al., 2016; Whyte et al., 2020) of harbour seals during offshore wind farm construction. Out of the 24 tagged harbour seals, it was predicted that four (17%) exceeded the estimated onset threshold for temporary auditory damage (in the form of a temporary threshold shift (TTS)), using the updated Southall et al. (2019b) thresholds for phocid seals in water (Whyte et al., 2020). The at-sea density of harbour seals was also compared between times of piling and non-piling (Russell et al., 2016; Whyte et al., 2020). Predicted seal density significantly decreased during pile driving within 25 km of the wind farm, and above estimated single-strike sound exposure levels (SELss) of 145 dB re $1\mu\text{Pa}^2 \cdot \text{s}$ (Figure 1.4; Whyte et al., 2020). Russell et al. (2016) used a cumulative approach to examining the relationship between seal density and wind farm distance (or SELss), where results were quantified and presented by including all spatial cells within each specified distance (e.g. seal density at distance x_1 includes all spatial cells between 0 and x_1 km; seal density at distance x_2 includes all spatial cells between 0 and x_2 km). In addition to this, Whyte et al. (2020) quantified the relationships using an annulus approach, where spatial cells are included based on distinct bands or zones as distance increases away from the wind farm (e.g. seal density at distance x_1 includes all spatial cells between 0 and x_1 km; seal density at distance x_2 includes all spatial cells between x_1 and x_2 km). By quantifying seal density change in annulus zones around the wind farm, the updated estimated relationships in Whyte et al. (2020) highlighted that changes in seal density were present in each distance (and SELss) increment. They also improved the applicability of these results to environmental impact assessments, as these assessments typically require data collected using the annulus approach.

Population-level redistribution studies are a key step in determining the presence and magnitude of potential disturbance effects; however, there are limitations to this approach, and there remain a number of uncertainties in understanding the effects of pile driving on seals. For example, to examine changes in seal density, these studies (Russell et al., 2016; Whyte et al., 2020) conducted a binary comparison between the distribution of all tagged seals during non-piling periods, and the distribution of all tagged seals during pile driving. Multiple piling events were necessarily combined together, averaging over multiple pile driving locations and different temporal patterns of piling blow energies (and corresponding sound levels). By this approach, it was also not possible to quantify differences



(a) Distance annulus

(b) SEL annulus

Figure 1.4 Predicted changes in seal density as a function of (a) distance from the centre of the wind farm and (b) estimated sound exposure level (b)(SEL, dB re $1\mu\text{Pa}^2 \cdot \text{s}$), with SEL averaged across all water depths and piling locations. (a) Seal density in annulus 5 km increments: plotted density change at distance x is the change in all spatial cells between $x - 5$ and x km. (c) Seal density in annulus 5 dB increments: plotted density change at SEL x is the change in all spatial cells between x and $x + 5$ dB. Annotations denote the number of spatial grid cells in each distance/SEL category. The dashed lines represent 95% confidence intervals. Figure adapted from Whyte et al. (2020).

between and within individual animals, or to investigate the effect of different internal (e.g. behavioural motivation) and external (e.g. time of day) contexts on the nature or severity of disturbance. Furthermore, although it is known that seal density is predicted to be lower near to pile driving activity, the mechanism by which this density change occurs is not understood. For example, changes in density could arise through seals travelling away from piling, through seals delaying entering an area of piling, or through a combination of both. Studying the behaviour of individuals, and accounting for the specific context of each disturbance event, may therefore provide greater insights into the effects of sound on marine mammals such as seals (Southall et al., 2021), and ultimately improve efforts to quantitatively model effects at the population level.

1.7 Statistical methods for behavioural response studies

To further understand the potential mechanisms of disturbance, and to understand individual variation, it is important to use data and analytical methods which provide information at the individual level. For example, research into the effects of sound on cetaceans have shown that, by tracking tagged individuals, it is possible to gain insights into how disturbance may affect horizontal movements (van Beest et al., 2018), dive behaviour (Falcone et al., 2017), and activity budgets (Isojunno et al., 2017). Much of this research has been driven by a need to understand and predict the effects of military sonar sounds on cetaceans. The observations of multiple mass strandings of beaked whales coincident with military sonar exercises has prompted concern that some animals may respond so severely to these sounds that they dive beyond their normal physiological limits, resulting in an increased risk of gas bubble emboli or decompression sickness (Cox et al., 2006; Fahlman et al., 2014; Hooker et al., 2009). Thus, behavioural response studies investigating the effects of sonar on the behaviour of different cetacean species have been conducted and, in tandem, the most appropriate statistical methods to analyse these data have been developed and utilised (by projects such as MOCHA; Harris et al., 2016). The challenges of analysing data of this sort include accounting for baseline variability in behaviour, modelling repeated observations on the same (few) individuals over time, and maximising the use of the multivariate metrics of movement recorded on the biologging devices (Harris et al., 2016). Whilst, in practice, a variety of statistical methods can be used in behavioural response studies, there are two main approaches increasingly being used to detect changes in individual movement behaviour.

1.7.1 Mahalanobis distance

The first approach is based on the multivariate statistic Mahalanobis distance (Mahalanobis, 1936). Mahalanobis distance measures the distance between two points in multivariate space, such as two data observations where multiple metrics are measured on both occasions. Correlation between the different metrics measured (the different dimensions of the multivariate space) are accounted for by including the covariance matrix of the data within the distance calculation. A full description of the statistic and its application to these studies is provided in Section 2.2.2. In this application, Mahalanobis distance can be used to reduce multivariate time-series measurements of movement behaviour (such as those recorded on a tag) to a single number summarising behaviour change over time (Figure 1.5). By examining the distribution of Mahalanobis distance values observed during baseline (or non-exposure) periods, understanding is gained on what values are normal to observe; in practice, this is usually approached by determining a quantitative threshold, above which behaviour is considered to be statistically unusual. By comparing Mahalanobis distance values observed during known sound exposures to the values observed in baseline, it is possible to identify times when individuals exhibited statistically unusual movement behaviour. These calculations are usually conducted within each individual to account for likely differences between animals. The focus of this approach is the identification of extreme or highly unusual movement behaviours ("change-points"), where individuals exhibit unusual changes in one or a combination of movement metrics.

While Mahalanobis distance-based methods have been used in studies of cetacean responses to sonar (DeRuiter et al., 2013; Miller et al., 2014), they have largely been used on very high-resolution data (e.g. accelerometer tags, that measure movement and body posture multiple times per second) and have not yet been applied to other taxa or other types of potential sound disturbance. Additionally, they have only been applied in controlled exposure experiments, which typically involve a standardised experimental protocol for exposing few (< 10) animals to set durations and levels of sound over a short time period (hours–days). The method has not yet been applied to observational studies, typically involving larger numbers of animals which are exposed to a wider variety of sound durations and levels as they move around their environment over a long study period (weeks–months). Furthermore, despite the reliance of many sonar behavioural response studies on this method, and its importance in producing quantitative estimates for naval activities (Harris et al., 2016), there has been limited quantitative assessment of how this approach performs, and in how different choices made in method implementation may affect the conclusions drawn.

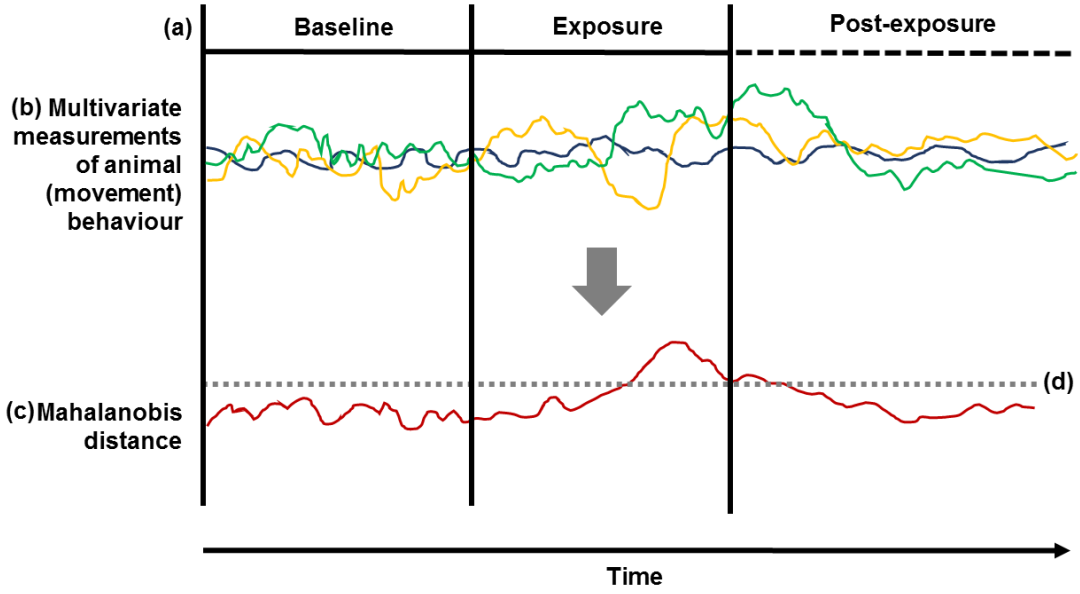


Figure 1.5 Mahalanobis distance can be used to reduce multivariate measurements of animal behaviour (b) to a single number summarising behaviour change over time (c). In behavioural response studies, data are often available before (baseline), during and after exposure to a particular disturbance (a), and so this method can be used to determine if the animal's behaviour exceeded a pre-defined threshold (d) in response to this disturbance.

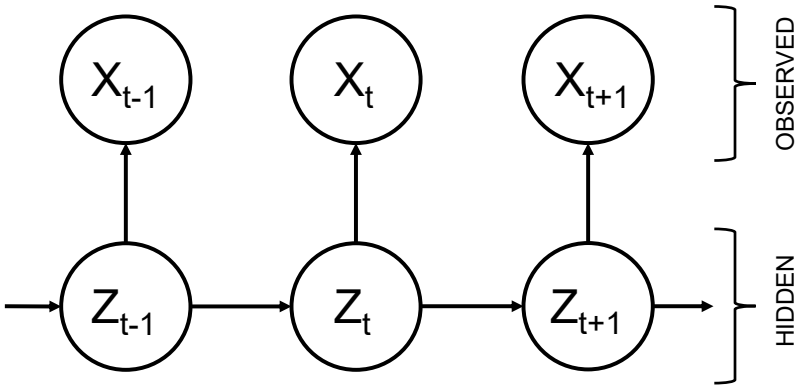


Figure 1.6 Dependence structure of a basic hidden Markov model (HMM). The observations $X_1, \ldots, X_T$ arise from an unobserved sequence of underlying states $Z_1, \ldots, Z_T$. Figure adapted from McClintock et al. (2020).

1.7.2 Hidden Markov models (HMMs)

State-space models are hierarchical models consisting of (1) a state process (an unobserved time series of interest) and (2) an observation process which produces data related to the underlying time series (Auger-Méthé et al., 2021). The second approach in behavioural response studies uses a type of state-space model (Auger-Méthé et al., 2021) called a hidden Markov model (HMM). HMMs are used to model and understand time series, where some system switches between a discrete number of "states" over time (Zucchini et al., 2016; Figure 1.6). The states themselves are not directly observed, and so are often described as latent or "hidden". Data observations are made which can then be used to infer the underlying state-switching process; these observations may be univariate (a single type of data observation over time) or multivariate (multiple types of data observation over time). The statistical model assumes that the state Z_t at time t depends only on the previous state Z_{t-1} at time $t-1$ (the first-order "Markov property"), and that each observation X_t depends only on the current state Z_t (the "conditional independence property") (Figure 1.6; Zucchini et al., 2016). This therefore assumes that there is no additional correlation between observations made at successive time steps, i.e. observation X_t is independent from observation X_{t-1}. Extensions and modifications to these relatively simple assumptions are increasingly being developed (McClintock et al., 2020), for example by incorporating longer-term memory into the state-switching process and by relaxing the assumption that sequential observations are independent.

HMMs are a practical approach to model ecological time series, where indirect (often imperfect) observations are made about a time series of interest (McClintock et al., 2020). In particular, HMMs are often used in animal movement studies, where observations of the animal's horizontal movement (e.g. steps, turns) are used to estimate the sequence of underlying animal behaviours (e.g. travelling, foraging, resting) (Langrock et al., 2012). Similarly, HMMs have been used to examine sequences of different dive behaviours in a variety of species including macaroni penguins (*Eudyptes chrysolophus*; Hart et al., 2010), American mink (*Neovison vison*; Bagniewska et al., 2013), and narwhals (*Monodon monoceros*; Ngô et al., 2019). In behavioural response studies, HMMs can be used to model the normal sequence of animal behaviours observed during baseline periods, and examine how this may be altered during potential disturbances (DeRuiter et al., 2017; Durbach et al., 2021; Isojunno et al., 2017). The focus of this approach is to obtain a broad estimate of the effects of disturbance on the overall movement activity of the animal, through exploring its effect on the type, duration, and/or nature of different exhibited behaviours.

1.8 Thesis overview

Overall, the objective of this thesis is to improve understanding of how individual seals react to *in situ* anthropogenic sound disturbance, with a particular focus on renewable energy applications. This thesis presents an in-depth observational study of wild seal behaviour in response to an anthropogenic sound source. This thesis also aims to generate quantitative estimates of behavioural disturbance for environmental impact assessments, and evaluate the statistical methodology of current behavioural response studies.

In **Chapter 2**, I design and run a simulation study to evaluate the performance of a statistical method used in cetacean behavioural response studies: Mahalanobis distance. I develop two novel simulators to generate realistic harbour seal biologging data for: (1) horizontal movement tracks (via a multi-state correlated random walk, with bias to and from spatial features), and (2) dive records (via a Markov chain and state-dependent multivariate distributions). The Mahalanobis distance method is applied to data from both simulators to quantify the performance of the method in a variety of simulated scenarios. This work aims to evaluate the performance of the method when applied to seal biologging data, compare different implementations of the method, and provide insights and recommendations for future studies using this method in any behavioural response study.

In **Chapter 3**, I investigate the horizontal movement of 24 GPS-tagged harbour seals and examine the movement tracks taken by individuals during pile driving at an offshore wind farm. Building on the findings of Chapter 2, I identify times of unusual movement behaviour (using Mahalanobis distance) and quantify the relationship of these events with piling sound level (using a Bayesian hierarchical model). A number of behavioural, individual and exposure-related factors are investigated as potential contextual drivers of behavioural response. The effective response range and total number of seals predicted to respond within the study area are estimated. This work provides direct information for offshore industries using pile driving, and insights into how individual seals may react to sound disturbance.

In **Chapter 4**, I examine the dive behaviour of harbour seals exposed to pile driving sounds during offshore wind farm construction, using dive records from 24 tagged individuals. Building on the findings of Chapter 2, I identify times of unusual dive behaviour during pile driving (using Mahalanobis distance) and quantify the relationship of these behaviours with piling sound level (using a Bayesian hierarchical model). I also investigate how pile driving affects behaviour-switching between different dive types (using a hidden Markov model).

In the **General Discussion**, I bring the results from Chapters 2, 3 and 4 together and present the wider implications of this work. In particular, the key insights gained for the renewable energy sector, for studies of behavioural disturbance, and for seal movement ecology are discussed. Knowledge gaps and future research directions are highlighted to aid developments in statistical, technological, ecological and applied research in this area. Together, this thesis presents an in-depth study of wild seal behaviour during anthropogenic noise exposure, provides quantitative information for environmental impact assessments, and generates concrete advice for future studies aiming to explore changes in animal movement behaviour.

As part of the work of this thesis, several `R` functions (R Core Team, 2020) were written which allow reproducibility of the data computation and analysis undertaken. A summary of these key functions is presented in Appendix A.

Chapter 2

Detecting behavioural responses in biologging data: an evaluation of Mahalanobis distance

2.1 Abstract

Behavioural response studies aim to investigate the effects of stimuli on animal behaviour. In studies of cetacean responses to sonar, an increasingly common approach is to use biologging devices to record aspects of individual movement behaviour, and use statistical approaches based on Mahalanobis distance to quantitatively identify behavioural responses to disturbance. Despite the growing application of this approach, there is currently limited information on (1) the detection and false positive rates of the method, (2) how different implementations of the method may affect performance, and (3) the performance of the method with lower resolution biologging devices. In this chapter, I designed a simulation study to quantify the detection and false positive rates of Mahalanobis distance approaches when aiming to detect behavioural responses in seal biologging data. To do this, I developed two simulators to generate both realistic harbour seal GPS locations and harbour seal dive summary records. A variety of different types of behavioural response were simulated, and different implementations of Mahalananobis distance were applied to the simulated data. In the horizontal simulations, strong and long movement responses were likely to be detected (up to 79% of simulations), whereas weak and short movement responses were detected less reliably (up to 43% of simulations). In the dive simulations, performance differed between the type of behavioural response simulated, with detection being generally higher for simulated responses with short unusual post-dive surfacing behaviour (∼55–100% of simulations), than responses with longer-term modifications to travelling behaviour (∼5–100% of simulations). False positive rates were generally low (∼5%) when at least five days of baseline data were available. Overall, both simulations provided important insights and recommendations for future behavioural response studies aiming to use this methodological approach. In particular, future studies should aim to (1) maximise the collection of baseline data prior to disturbance, (2) calculate the covariance matrix (used in calculations of Mahalanobis distance) on baseline data only, and (3) consider carefully the duration and type of behavioural response that is of interest prior to analysis.

2.2 Introduction

Behavioural response studies aim to identify and characterise the effects of particular stimuli or stressors on animal behaviour. Studies typically measure one or more aspect of animal behaviour, and attempt to assess whether the stimulus has caused the animal to exhibit a substantial change in this behaviour. For example, animals may change their behaviour in response to cues that inform them of predators. Jara and Perotti (2010) found that three species of tadpoles (*Pleurodema thaul*, *Pleurodema bufoninum*, *Rhinella spinulosa*) reduced their movement activity when presented with visual and chemical cues of predators; however, the responses differed between tadpole species, predator species, and the size of the tadpole. Animals may also alter their behaviour in response to environmental cues. For example, brown trout (*Salmo trutta*) can become nocturnal feeders during winter in response to low water temperatures (Heggenes et al., 1993). As human activities continue to expand across the globe (Venter et al., 2016), a key consideration in many recent studies are the behavioural responses of animals to humans (e.g. Doherty et al., 2021). Animal reactions may be due to human presence directly (Braimoh et al., 2018; Ordiz et al., 2019), or due to a stimulus produced by human activity, such as sound (Weilgart, 2007). An increasingly popular method to collect data on animal behaviour is to use biologging devices, which are deployed directly onto the study animals and can contain a variety of sensors to record aspects of behaviour (e.g. location, acceleration, posture) and the environment (e.g. depth, temperature, salinity). This approach can be useful as the study animals do not have to be visible to observers at all times during the study, allowing data to be collected remotely on wild animals in their natural habitat, over long periods (e.g. months), and in habitats where visibility is limited (Wilmers et al., 2015). This is particularly advantageous when considering diving animals, which are often unavailable to observe at the water surface (Hooker et al., 2007). As such, many studies of marine mammal behaviour have used biologging devices to help understand the potential impacts of human disturbance on their behaviour.

2.2.1 Use of Mahalanobis distance in practice

Despite developments in biologging technology allowing data to be recorded on many different species, statistical methods to effectively analyse these data and identify disturbance events are relatively limited. One approach that has been used to quantify behavioural responses in marine mammals, particularly in studies of cetacean responses to sonar, is

based on Mahalanobis distance (DeRuiter et al., 2013). This approach has three key benefits which have led to its growing use in this context. First, it can be used to objectively and clearly categorise an animal's behaviour during disturbance as either a response or a non-response. Second, the method identifies the precise timing of the behavioural response, allowing identified responses to be linked to secondary data which may explain its occurrence, such as estimated sound level. The relationship between the probability of an animal responding and the strength of a particular stimulus can then be quantified. Third, the method is relatively straightforward to implement, and does not require any further model fitting or goodness-of-fit assessments.

Mahalanobis distance has been applied in studies assessing the impact of sonar on a variety of cetacean species, including blue whales (*Balaenoptera musculus*, Southall et al., 2019a), minke whales (*Balaenoptera acutorostrata*, Kvadsheim et al., 2017), killer whales (*Orcinus orca*, Miller et al., 2014), long-finned pilot whales (*Globicephala melas*, Antunes et al., 2014), and beaked whales (e.g. DeRuiter et al., 2013, Stimpert et al., 2014). Biologging data have been a key component in the majority of these studies, but there has been variation in how the data have been considered within the Mahalanobis distance approach. In general, the approach involves (1) deciding on the animal behaviour metrics of interest (e.g. movement speed, dive depth), (2) dividing the data time series into discrete time windows, and (3) performing Mahalanobis distance calculations between pairs of time windows. (A detailed introduction to the specifics of the approach is provided in Section 2.2.2.) The chosen animal behaviour metrics differ between studies, and some studies have also subdivided the measurements from biologging tags into "movement" and "energetic" metrics to perform separate Mahalanobis distance analyses, attempting to detect different types of response (Kvadsheim et al., 2017; Miller et al., 2015; Wensveen et al., 2019). Most previous studies have used the approach to divide the data into time windows corresponding to durations of time (e.g. every x minutes). Other studies have divided the data into dive-by-dive units for analysis (e.g. DeRuiter et al., 2013, Kvadsheim et al., 2017), where the start of each underwater dive (or group of dives) represents the start of a new time window (e.g. every x dives). The size of the resulting time windows chosen (a necessary step in the analysis) varies, from 5 minutes (e.g. Stimpert et al., 2014) to 15 minutes (e.g. Wensveen et al., 2019), and the overlap between time window calculations also differs. During the Mahalanobis distance calculation stage, the majority of studies compare animal behaviour during disturbance to their average behaviour during non-disturbance periods (baseline), but it is also possible to compare the animal's current behaviour to its most recent behaviour to detect behavioural responses (Antunes et al., 2014).

2.2.2 Mahalanobis distance: an introduction

A multivariate statistic

Mahalanobis distance (Mahalanobis, 1936) is a statistical measurement of the distance between two points in multivariate space. These two points should be from the same statistical distribution, and contain observations of the same set of variables (e.g. body length, body weight). In this application, the method is used to calculate the distance between the means of two sets of multivariate data. By including the covariance matrix of the data within the calculation, the distance accounts for correlation between variables within the dataset. Hence, Mahalanobis distance can be used to identify unusual combinations of variables (outliers) in multivariate data (for a two-dimensional example see Figure 2.1). The Mahalanobis distance $D_{(\boldsymbol{x},\boldsymbol{y})}$ between two sets of observations is calculated by

$$D_{(\boldsymbol{x},\boldsymbol{y})} = \sqrt{(\boldsymbol{x}-\boldsymbol{y})^T \boldsymbol{S}^{-1} (\boldsymbol{x}-\boldsymbol{y})} \tag{2.1}$$

where $\boldsymbol{x} = (x_1, x_2, x_3, ..., x_N)$ is a vector containing the mean of multivariate data observations for variables 1 to N, $\boldsymbol{y} = (y_1, y_2, y_3, ..., y_N)$ is a vector containing mean observations of the same variables, and $\boldsymbol{S}^{-1}$ is the inverse covariance matrix of the data. The resulting Mahalanobis distance is a measurement of how far apart the means of the two observations are, given the variability and correlations of the dataset. Mahalanobis distance is both unitless and scale-invariant (i.e. is not affected by the magnitude of the data observations used).

Data inputs

Any number of behaviour measurements can be used as inputs in the Mahalanobis distance calculation. The choice of inputs depends on the data available, the research question, and some judgement on which variables may be appropriate and informative in detecting a particular type of behavioural response. Some studies have carried out multiple Mahalanobis distance analyses (with different inputs) to look for different types of response, e.g. Miller et al. (2015) investigated the behaviour change in "movement parameters" and "energetic parameters" separately.

Time window structure

In order to use Mahalanobis distance to investigate change over time series of data, observations need to be divided into time windows. There are several different approaches to do this; here I describe these approaches and outline the terminology I will use to differentiate them.

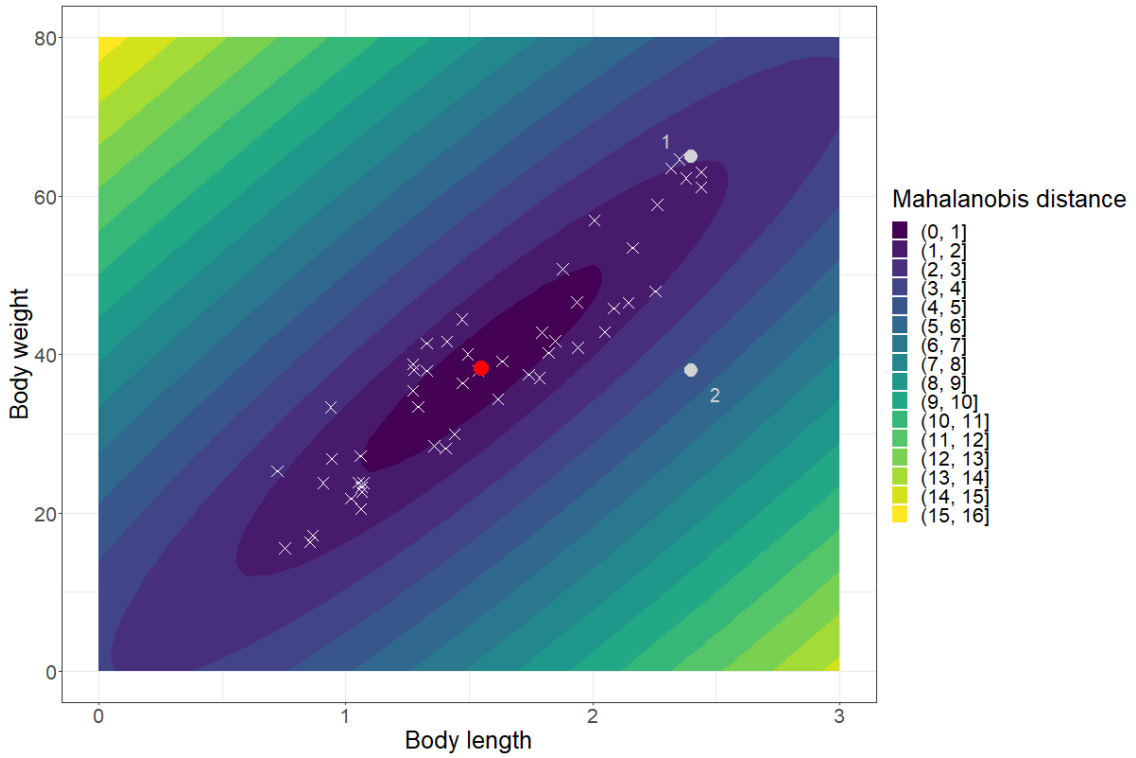


Figure 2.1 Example of Mahalanobis distance for two variables: animal body length and weight. Shown are a simulated dataset (crosses) and the values of Mahalanobis distance (colour-scale) relative to the mean (red dot) of this dataset. The new data observation 1 (annotated white dot) would have a lower Mahalanobis distance (value of 2.1), whereas observation 2 (annotated white dot) would have higher Mahalanobis distance (value of 4.3). Although observation 1 has a high body weight, it is not an unusual value for the corresponding observed length. Observation 2 has a weight equal to the mean of the dataset and the same body length as observation 1, but this is an unusual combination of the two variables.

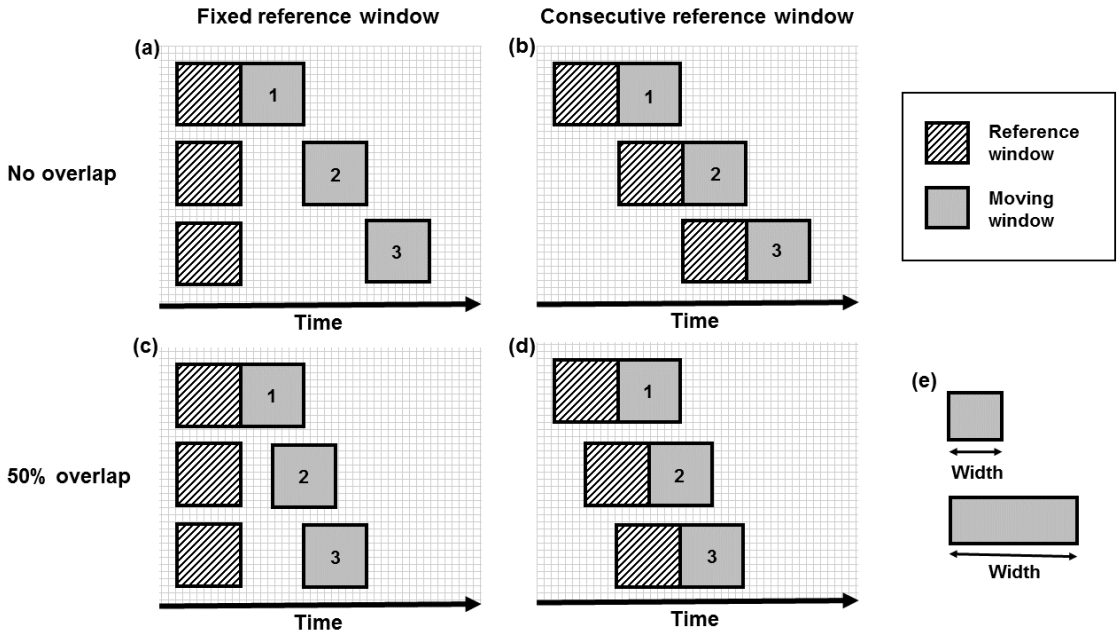


Figure 2.2 Time window options for Mahalanobis distance calculations. All calculations take place between a reference window (striped) and moving window (solid colour). The reference window can be fixed at a particular point in the data timeline, with a moving reference window for comparison (a); or both windows can slide consecutively together over the data timeline (b). When the windows are moved over the data timeline, they can have overlap with the previous calculation (c)(d) or no overlap (a)(b). The width of the time window can also be varied for different studies (e).

Every Mahalanobis distance calculation takes place between two groups of data observations (e.g. $\boldsymbol{x}$ and $\boldsymbol{y}$ in Equation 2.1). Here, these two groups are two time windows placed within the dataset: a moving window, and a reference window. Vector $\boldsymbol{x}$ contains the mean of the observation(s) contained within the moving window, and vector $\boldsymbol{y}$ contains the mean of the observation(s) contained within the reference window. The moving window slides along the timeline of the data, and at each position (e.g. 1, 2, 3; Figure 2.2) the Mahalanobis distance between the moving and reference window is calculated.

The reference window itself can either be fixed or consecutive (Figure 2.2). If the reference window is fixed (Figure 2.2(a)(c)), it is usually positioned immediately prior to the exposure event or covers the entire baseline period. By this approach, the moving window is comparing the current behaviour to the animal's general behaviour prior to exposure. If the reference window is consecutive (Figure 2.2(b)(d)), it moves forward in time alongside the moving window. By this approach, the moving window is comparing the current behaviour to the animal's behaviour immediately prior to this.

Each time the window is moved, the amount of window overlap with previous calculations can vary. If the time windows have no overlap (see Figure 2.2(a)(c)), at each time step the moving window contains a completely new set of data observations. This provides a Mahalanobis distance calculation at the same resolution as the window size. If the moving windows overlap (i.e. the new moving window position still contains some of the observations from the previous moving window calculation; Figure 2.2(c)(d)), Mahalanobis distance can be calculated at a finer resolution than that of the window size. Different amounts of overlap will provide Mahalanobis distance measurements at different resolutions, and the width (duration) of the windows can be varied to contain different numbers of observations (Figure 2.2(e)).

Alternatively, the window overlap is sometimes presented as the slide amount of the windows. The two are linked by the formula

$$s = w - o_t \tag{2.2}$$

where s is the slide amount (in time or data units, e.g. seconds, number of dives), w is the window width (in time or data units) and o_t is the amount of overlap (in time or data units). The proportion of overlap (0 to 1) with previous windows o_p can be calculated by $o_p = o_t/w$.

The covariance matrix

In addition to the observations from the time windows ($\boldsymbol{x}$ and $\boldsymbol{y}$), the other input required for the Mahalanobis distance calculation is a covariance matrix (denoted $\boldsymbol{S}$ in Equation 2.1). There are two different possible approaches to calculating this: (1) the covariance matrix can be calculated across the entire dataset for the individual (including baseline, exposure, and post-exposure periods), or (2) the covariance matrix can be calculated across the individual's baseline data only. In general, including all of the data (approach 1) is considered to be a more conservative approach (Stimpert et al., 2014) to limit false positive detections, especially when there is limited baseline data available.

Identifying responses by randomisation test

Once the chosen inputs, time windows and covariance matrix have been specified, Mahalanobis distance can be calculated across the time windows by Equation 2.1. To identify unusual behaviour changes and possible behavioural responses, a randomisation test is typically used to compare behaviour change during the sound exposure to behaviour during the baseline period. For each known exposure period, a number of "mock exposure periods"

(of the same duration as the true exposure) are sampled randomly from the baseline data for that animal. The maximum Mahalanobis distance during the exposure period is then compared to the Mahalanobis distances observed across each of the mock exposure periods. The mock exposure period measurements are used to determine a response threshold (e.g. the 95^{th} percentile of Mahalanobis distances). The probability of a behaviour change of this magnitude occurring in the baseline data is low, and therefore if the animals' behaviour exceeds this threshold during the exposure period it is interpreted as being in response to the disturbance.

2.2.3 Knowledge gap

Despite widespread use of the Mahalanobis distance approach in studies of cetacean responses to sonar, and the application of these results to manage future human activities, there has currently been no formal assessment of how well these methods perform. There is a lack of information on how often these methods may exhibit a false positive (detect a behavioural response when none is present) or how often behavioural responses present in the data are undetected by this methodological approach. Different studies have also implemented the method in different ways (e.g. different time window structures), and so it may be unclear which approach to use and whether these differences affect performance. This knowledge gap limits the ability of current studies to robustly identify responses to stimuli, and could limit the potential for this method to be applied with confidence to other species, data types and disturbance sources. With the exception of one study, which used focal follow methods to determine the animal's location approximately every two minutes (Antunes et al., 2014), the data used for this method has typically been based around high-resolution accelerometer tags. These tags collect information on the animal's movement and posture multiple times per second. It is unclear how these methods may perform with lower resolution biologging data, such as individual dive summaries or location data collected at ∼10–20 minute intervals.

2.2.4 Aims

The majority of pinniped biologging data consists of long-term deployments with dive data from time-depth recorders, or location data from tags which use GPS or ARGOS satellites. In this chapter, I aim to evaluate the performance of Mahalanobis distance as a method to detect behavioural responses in two types of seal biologging data: GPS locations and

dive records. Specifically, this chapter will develop a simulation study to (1) estimate the detection and false positive rates of the method, (2) compare different implementations of the method, and (3) provide recommendations for future use. To do this, performance will be quantified by applying Mahalanobis distance to both simulated GPS location data and simulated dive records, and a variety of implementation options will be compared to test detection of different simulated behavioural responses. The results will be used to inform advice for future studies using Mahalanobis distance in behavioural response detection.

2.3 Methods

Two simulation studies were designed and conducted to test the performance of Mahalanobis distance in detecting behavioural responses. The first simulation considered seal horizontal movement, focussing on GPS location data. The second simulation considered vertical movement, focussing on the dive-by-dive measurements (e.g. dive depth, duration) recorded by most marine mammal biologging devices. In each case, the aim was to simulate data that replicate those recorded by the tag. By this approach, I was able to simulate data both with and without behavioural responses, and then use this simulated data to quantify the ability of the method to pick out these behaviour changes.

In both simulations, it was assumed that the behavioural response of the animal did not alter the quality of the (simulated) data itself. In theory, it is possible that behavioural changes may alter the duration or frequency of seal surfacings (or movement/behaviour at the surface), leading to different availabilities and qualities of location data between different behaviours. With developments in GPS technology (such as Fastloc GPS), this difference is likely to be minimal, but could be more substantial for ARGOS tracking data which is highly reliant on achieving multiple satellite links to obtain a high-quality location estimate (Carter et al., 2016). At present, there is no evidence to suggest substantial differences in biologging data quality exist between different animal behaviours; however, this is an important area for further research development.

2.3.1 Horizontal movement: Seal location data

A framework for simulating seal locations

A simulator for seal location data was developed, building on the general framework presented in McClintock et al. (2012). Here, the simulator was a (correlated and biased)

Table 2.1 Summary of key parameters and notation used in horizontal movement simulation.

Parameter	Definition
Behavioural states	
$\psi_{k,i}$	State transition probability (from state k to i)
η	Switch strength (from state 1 to 2) near attraction points
Attraction and repulsion	
$d_{a,t}$	Distance to attraction point (at time t)
$d_{r,t}$	Distance to repel point (at time t)
ω_a	Attraction source strength
ω_r	Repel source strength
α_t	Attraction weighting (effect on animal's behaviour)
β_t	Repel weighting (effect on animal's behaviour)
π	Probability of retaining the same attraction point
Step lengths (speed)	
s_t	Step length (at time t)
$v_{i,t}$	Expected mean speed (at time t and behavioural state i)
ϵ_i	Standard deviation of speed (for behavioural state i)
v_i	Mean speed for behavioural state i
$v_{max,i}$	Maximum mean speed during a response (for behavioural state i)
κ_i	Correlation with previous step length (for behavioural state i)
Headings (movement direction)	
ϕ_t	Heading direction (at time t)
$\lambda_{i,t}$	Expected mean heading direction (at time t and behavioural state i)
$\rho_{i,t}$	Heading concentration (at time t and behavioural state i)
ρ_z	Mean heading concentration (for each behavioural state)
μ_t	Direction towards current attraction point (at time t)
σ_t	Direction away from repulsion point (at time t)
γ_a	Concentration constant for attraction strength
γ_r	Concentration constant for repel strength

discrete-time, multi-state random walk, where animal tracks were able to exhibit attraction to points in space, as well as repulsion from a disturbance point. Below, I outline the general structure of the simulator, before describing how the simulator was parameterised and how it was used in the simulation study. A summary of the key notation used is provided in Table 2.1.

Behavioural states

At any time t, an individual is assumed to be in one of Z behavioural states, which describe its general movement behaviour. Here, I considered two states:

- State 1: Travel (characterised by larger movement steps and lower variability in heading);
- State 2: Encamped (characterised by smaller movement steps and higher variability in heading).

The sequence of behavioural states over time for an individual is a first-order Markov process, where animals switch between states based on transition probabilities:

$$\psi_{k,i} = \Pr(z_t = i | z_{t-1} = k)$$

where $\psi_{k,i}$ is the probability of an individual switching to state i at time t, given it was in state k at time $t-1$. The transition probability matrix G_0 contains the initial transition probabilities (at time $t = 0$) between the two behavioural states:

$$G_0 = \begin{pmatrix} \psi 0_{1,1} & \psi 0_{1,2} \\ \psi 0_{2,1} & \psi 0_{2,2} \end{pmatrix}.$$

I include attraction points in the simulated landscape to mimic the expected behaviour of a central-place forager, where animals travel between different points of attraction (e.g. foraging areas, haul-out sites) to search for prey and/or haul out. Therefore, seals may be more likely to switch into state 2 (encamped) as they approach one of these attraction points. In order to simulate this, at every timestep t I update the transition probability matrix:

$$G_t = \begin{pmatrix} \psi_{1,1}(t) & \psi_{1,2}(t) \\ \psi 0_{2,1} & \psi 0_{2,2} \end{pmatrix}$$

where the probability of encamped animals changing state remains constant, but the probability of travelling animals changing their behavioural state varies over time. The probability of switching from state 1 to 2 ($\psi_{1,2}(t)$) increases as distance $d_{a,t}$ to the attraction point at time t decreases:

$$\psi_{1,2}(t) = \text{logit}^{-1}\left(\text{logit}(\psi_{1,2}) + \frac{\eta}{d_{a,t}}\right)$$

where η, the switch strength, controls the relationship between distance to the attraction point ($d_{a,t}$) and the probability of switching to state 2 ($\psi_{1,2}$), logit denotes the logit function (i.e. $\text{logit}(p) = \log(p) - \log(1-p)$), and logit^{-1} denotes the inverse logit (or logistic) function. Thus, the minimum probability of switching into state 2 is the baseline transition

rate $\psi 0_{1,2}$, and this increases to 1 as distance $d_{a,t}$ decreases to zero. The corresponding probability of staying in state 1, $\psi_{1,1}(t)$ is the complement of $\psi_{1,2}(t)$.

Attraction and repulsion

The importance of attraction can vary with distance $d_{a,t}$ from the attraction point. In this case, one may expect the strength of attraction towards a location to be strong further away (to bring the animal close, even from long distances) and weaker nearby (to allow animals to move around loosely within the vicinity of the attraction point). The attraction weighting α_t (effect of attraction on the animal's behaviour) increases with distance $d_{a,t}$ from the attraction point:

$$\alpha_t = \tanh(\omega_a d_{a,t})$$

where ω_a is the strength of the attraction source in the landscape. This attraction weighting ranges from 0 to 1, where if $\alpha_t = 0$ there is no effect of attraction and if $\alpha_t = 1$ the animal's movement is completely dependent on its attraction to a source. The effect of attraction on the animal's behaviour is therefore a function of both the source strength and the distance from the source. In this case, I consider two sources of attraction, where only one attraction point is active at a time. In order to emulate a seal residing at an attraction point for a reasonable duration, every time the seal switches from state 2 (encamped) to state 1 (travel), there is a probability π that it keeps the same attraction point, and probability $1 - \pi$ that it switches.

Similarly, one would expect the importance of repulsion (where points of repulsion represent noise sources) to vary with distance. The strength of repulsion would be highest close to the point of repulsion, and lower further away from the source. The repulsion weighting β_t (effect of repulsion on the animal's behaviour) decreases with distance $d_{r,t}$ from the point of repulsion:

$$\beta_t = \tanh\left(\frac{\omega_r}{d_{r,t}}\right)$$

where ω_r is the strength of the source of repulsion in the landscape. The repulsion weighting ranges from 0 to 1, where if $\beta_t = 0$ there is no effect of repulsion and if $\beta_t = 1$ the animal's movement is completely dependent on its repulsion from a source. Whilst an attraction point is always present in the simulation, a point of repulsion can appear and disappear over time. During a response simulation, the source of repulsion appears a set distance $d_{r,0}$ away from the animal (at a random location around the circumference of the animal's current location) and remains in this location for the duration of the simulated exposure. By this approach, repel sources could appear when animals were near or far from their sources

of attraction, enabling the simulations to average over multiple possibilities of response scenario.

Step lengths (speed)
At each timestep t I simulate a movement step length s_t for the animal:

$$[s_t \mid s_{t-1}, z_t = i] \sim Gamma(a_{i,t}, b_{i,t})$$

where $a_{i,t}$ and $b_{i,t}$ are the shape and scale parameters (respectively) of a gamma distribution of possible step lengths. The simulated step length depends on the step length in the previous timestep s_{t-1}, the behavioural state i, and the effect of repulsion at the animal's current location. The shape and scale parameters $(a_{i,t}, b_{i,t})$ of the Gamma distribution are calculated from an expected mean speed $v_{i,t}$ at time t and a standard deviation ϵ_i for each behavioural state i. (The conversions are $a_{i,t} = (v_{i,t}/\epsilon_i)^2$ and $b_{i,t} = \epsilon_i^2/v_{i,t}$.)

The expected mean speed $v_{i,t}$ depends on the behavioural state i, the repulsion weighting β_t at the animal's location, and the correlation κ_i between sequential movement steps:

$$v_{i,t} = \overbrace{(1-\beta_t)(\underbrace{\kappa_i s_{t-1}}_{\text{Correlated movement}} + \underbrace{(1-\kappa_i)v_i}_{\text{Behavioural state}})}^{\text{Normal behaviour}} + \overbrace{\beta_t v_{\max,i}}^{\text{Response}} \tag{2.3}$$

where s_{t-1} is the step length at the previous timestep, v_i is the mean speed for each behavioural state i, and $v_{\max,i}$ is the maximum mean speed during a behavioural response in state i. In this formulation, if the repulsion weighting $\beta_t = 1$ the animal is responding fully and exhibits its maximum response speed, and if $\beta_t = 0$ (there is no centre of repulsion or the animal is very far away from the source) the animal's expected speed reduces to its mean speed for the behavioural state it is in (i.e. normal behaviour). In this way, I simulate a seal increasing its movement speed during a behavioural response. As attraction is a long-term process influencing the animal's simulated movement between geographically separated points, here it assumed that attraction only influences movement direction and behavioural state, and not the speed of movement within each of the behavioural states.

Headings (movement direction)
At each timestep t I simulate a heading direction ϕ_t from a wrapped Cauchy distribution:

$$[\phi_t \mid \phi_{t-1}, z_t = i] \sim \text{wCauchy}(\lambda_{i,t}, \rho_{i,t})$$

where $\lambda_{i,t}$ is the mean ("expected") heading and $\rho_{i,t}$ is the heading concentration (di-

rectedness of movement). The wrapped Cauchy is a circular distribution and so enables simulation of absolute headings in any direction in space (relative to points of attraction and repulsion). The simulated heading therefore depends on the heading in the previous timestep ϕ_{t-1}, the behavioural state i, and the effects of attraction and repulsion at the animal's current location.

The expected heading $\lambda_{i,t}$ depends on the relative weightings of attraction α_t and repulsion β_t at the animal's location:

$$\lambda_{i,t} = \overbrace{(1-\beta_t)(\underbrace{\alpha_t\mu_t}_{\text{Attraction}} + \underbrace{(1-\alpha_t)\phi_{t-1}}_{\text{Correlated movement}})}^{\text{Normal behaviour}} + \overbrace{\beta_t\sigma_t}^{\text{Response}}$$

where μ_t is the heading towards the current attraction point, σ_t is the heading away from a repulsion point, and ϕ_{t-1} is the heading in the previous timestep. In this formulation, if the repulsion weighting $\beta_t = 1$ the animal is responding fully and its expected heading is directly away from the centre of repulsion, and if $\beta_t = 0$ the animal's expected heading reduces to its normal behaviour, which depends on attraction and correlation with the animal's previous heading. In this way, I simulate a seal exhibiting correlated movement paths, with attraction towards possible foraging or resting sites, and repulsion away from centres of possible disturbance.

The concentration parameter $\rho_{i,t}$ determines how directed the animal's movement is. If $\rho_t = 1$, there is no variability in heading and I simulate a movement heading ϕ_t equal to the expected heading λ_t. If $\rho_t = 0$, there is no directness and any heading is equally likely. One would expect differences in concentration between different behavioural states, with higher concentration in state 1 (travel). One would also expect higher concentration further from the attraction point (to bring animals towards the centre, and allow exploration near the attraction site), and higher concentration closer to the repulsion point (more directed movement in response to disturbance). Therefore, the concentration $\rho_{i,t}$ is calculated by:

$$\rho_{i,t} = \begin{cases} \text{logit}(\text{logit}^{-1}(\rho_z) + \gamma_a d_{a,t}) & \text{if } \beta_t = 0 \\ \text{logit}(\text{logit}^{-1}(\rho_z) + \frac{\gamma_r}{d_{r,t}}) & \text{if } \beta_t > 0 \end{cases}$$

where ρ_z is the mean concentration for each behavioural state, γ_a is a concentration constant for attraction strength, and γ_r is a concentration constant for strength of repulsion. In this way, the variability in heading depends mainly on behavioural state, but can be altered slightly by attraction or repulsion.

Parameterising normal (baseline) behaviour

The parameters for the seal dive simulator were based on GPS location data collected from SMRU Instrumentation GPS telemetry tags (hereafter GPS/GSM tags; SMRU Instrumentation, University of St Andrews, Fife, UK) on 24 harbour seals tagged in The Wash, south-east England, UK. These data were collected during intermittent pile driving as part of construction of a nearby offshore wind farm (Hastie et al., 2015; Russell et al., 2016; Whyte et al., 2020), and so location data that overlapped with the times of pile driving construction activity were removed from the dataset.

All location data were linearly interpolated to regular 15-minute intervals. Known data gaps (due to pile driving and haulouts) were removed from the interpolated data for each individual, to produce separate complete segments of movement tracks. Only track segments containing at least 100 data points (∼25 hours) and observations of both types of movement behaviour (travel, encamped) were retained to inform simulation parameters. For each individual, a hidden Markov model (HMM; Zucchini et al., 2016) with two behavioural states was fit to the track segments to obtain parameter values for each of the behaviours, using the R package `momentuHMM` (McClintock and Michelot, 2018). Movement step lengths were fit using a Gamma distribution and turn angles using a wrapped Cauchy distribution with a mean of 0. Estimated mean step lengths, standard deviations of step length, and concentrations of heading angle were averaged across models to obtain an overall mean across individuals for each behavioural state. These values were used as the simulation parameters for the mean speed v_i, standard deviation of speed ϵ_i and mean heading concentration ρ_i for each behavioural state i. The fitted HMMs were also used to obtain transition rates between the two states, used for the initial transition probability matrix G_0.

The step length correlation parameter κ_i and the parameters involved in simulating attraction (behavioural state switch strength η, attraction source strength ω_a, probability of retaining the same attraction point π, and concentration constant for attraction strength γ_a) were obtained by gradually increasing the influence of each parameter until the simulated tracks resembled those of the real data used in this study and other studies of seal tracking data.

Parameterising a response

For this simulation study, I aimed to simulate a fleeing response, where the animal generally speeds up and moves away from the disturbance source. A study of tagged harbour seals exposed to acoustic deterrence devices detected modest increases in speed and directed movement away from the sound source in some animals (Gordon et al., 2019). In practice, simulating this response involved the following.

- **Increasing the probability of switching into the "travelling" state**.
  I increased the probability $\psi_{1,2}$ of switching from state 1 to state 2 during a response.
- **Increasing the speed (simulated step lengths) of the animal**.
  This was implemented by increasing the maximum mean speed $v_{\max}$ which is used in Equation 2.3 to generate the speed at each timestep.
- **Simulating movement which is more directed, and moving away from the disturbance source**.
  This was implemented by increasing the strength ω_r of the simulated repel source in the landscape, which influences the animal's movement direction away from the repulsion point. The directedness of the animal's movement during the response is implemented by the concentration parameter γ_r, which makes the movement heading less variable.

These simulation input parameters were modified to generate possible responses that might be observed in real data. I ran a series of different simulations by modifying the strength of the disturbance point (weak or strong), the initial starting distance of the animal from the disturbance point (5, 15 or 25 km), and the length of the response (0.5 or 3 hours).

Simulation setup

Timeline setup

I ran 500 simulations of individual seal horizontal movement data for each type of response, with locations simulated at a 15-minute time resolution. For each simulation, I generated 30 days of baseline (pre-exposure) data, followed by one day of data containing both the exposure and post-exposure period. As above, for each simulated response type, movement data after the baseline period were modified to emulate responses of different types and durations. In order to estimate the false positive rate of the method, a control set of

Table 2.2 Mahalanobis distance options for horizontal simulations. All 48 combinations of options were compared.

Mahalanobis distance specification	**Values and options used**
Inputs (same for all simulations):	Speed (m s^{-1})
	Easting
	Northing
	Circular heading variation
Baseline data availability:	5 days; 20 days
Covariance matrix used:	All data; Baseline only
Reference window position:	Fixed; Consecutive
Window size (minutes):	15; 20; 25; 30
Window overlap (proportion of window size):	0.1; 0.5; 0.9

500 simulations was also generated in which no response was inserted (i.e. no simulation parameters were modified from baseline). In each simulation, the simulated seal began at the x, y coordinate of (0, 0) and two attraction points were positioned at (0, 0) and (30 km, 30 km). In response simulations, the location of the centre of repulsion is randomly generated at the start of the simulated response and remains static throughout the rest of the simulation. The centre of repulsion is randomly placed on a circumference of a circle around the animal's current location, where the radius is specified by the initial starting distance $d_{r,0}$.

Inputs for Mahalanobis distance

In all analyses of the simulated location data, the metrics used as inputs for the Mahalanobis distance analyses were: speed, easting, northing, and heading variation (Table 2.2). After calculation of the movement metrics, the metrics were interpolated to a five-minute resolution (Antunes et al., 2014) for use in analysis. A standard set of input variables were chosen for all simulations and it was not investigated how using different inputs may affect the performance of the method. Instead, it is recommended that individual studies carefully consider the inputs they choose to use for the type of behaviour change they are trying to detect.

Comparison of different method implementations

For each simulated dataset, a variety of implementations of Mahalanobis distance were applied. I varied the size of the windows used (15–30 minutes), the overlap proportion of the windows (0.1–0.9), the position of the reference window (fixed or consecutive), and the covariance matrix used (all data or baseline data only) (Table 2.2). In total, this resulted

Table 2.3 Summary of key parameters and notation used in dive simulation.

Parameter	Definition
Behavioural states	
$\psi_{k,i}$	Dive transition probability (from dive type k to i)
Dive metrics	
m_t	Maximum depth of dive (for dive t)
d_t	Descent duration of dive (for dive t)
b_t	Bottom duration of dive (for dive t)
a_t	Ascent duration of dive (for dive t)
s_t	Surface duration after dive (for dive t)

in 48 different implementations of Mahalanobis distance that were investigated. I also investigated the effect of having different amounts of baseline data available, by running analyses with 5 days and 20 days of baseline data for all response simulations, and 1–30 days for the control simulation.

2.3.2 Vertical movement: Seal dive data

A framework for simulating seal dives

Here, a simulator for seal dive data was developed. The simulator was structured as a Markov-chain of different dive types, with the dive metrics simulated from a multivariate gamma distribution for each dive type. This simulator was based on the dive data collected from SMRU GPS-GSM tags, but the framework and metrics used are applicable to a wide range of time-depth recorder tags. Here, I first describe the structure of the simulator, before detailing how the simulator was parameterised and how it was used in the simulation study. A summary of the key notation used is provided in Table 2.3.

Behavioural states

Each individual's data consists of a sequence of dives over time. At each dive t, individuals can be in one of K dive types (behavioural states), which describe the general properties of the dive. The sequence of dive types over time is a first-order Markov process, where animals switch between dive types based on transition probabilities:

$$\psi_{k,i} = \Pr(z_t = i | z_{t-1} = k)$$

where $\psi_{k,i}$ is the probability of an individual switching to dive type i in dive t, given its previous dive $t-1$ was of dive type k. The transition probability matrix G_n contains the transition probabilities between the K dive types during normal behaviour:

$$G_n = \begin{pmatrix} \psi_{1,1} & \psi_{1,2} & \cdots & \psi_{1,K} \\ \psi_{2,1} & \psi_{2,2} & \cdots & \psi_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{K,1} & \psi_{K,2} & \cdots & \psi_{K,K} \end{pmatrix}$$

State-dependent dive metrics
For each dive t, I simulate five properties of the dive:

- m_t – Maximum dive depth in metres;
- d_t – Descent duration in seconds;
- b_t – Bottom duration in seconds;
- a_t – Ascent duration in seconds;
- s_t – Surface duration (post-dive) in seconds.

All of the dive properties depend on the current dive type i, and are simulated jointly using a multivariate gamma distribution

$$[m_t, d_t, b_t, a_t, s_t \mid z_t = i] \sim MVGamma \begin{bmatrix} \alpha_{m,i} & \beta_{m,i} \\ \alpha_{d,i} & \beta_{d,i} \\ \alpha_{b,i} & \beta_{b,i} \\ \alpha_{a,i} & \beta_{a,i} \\ \alpha_{s,i} & \beta_{s,i} \\ \kappa_i & \end{bmatrix}$$

where α and β are the shape and rate parameters respectively of a gamma distribution for each of the five dive properties in each dive type i. For example, $\alpha_{m,i}$ and $\beta_{m,i}$ are the shape and rate parameters for the maximum depth of a dive of type i. κ_i is the correlation matrix between all five dive properties in dive type i. The R package `lcmix` (Dvorkin, 2012) was used to simulate the multivariate gamma distributions. This package uses a normal (Gaussian) copula (Song, 2000) to combine the univariate gamma distributions. In

this way, I was able to simulate dives with different properties for each unique dive type, whilst retaining the realistic correlations that exist within each dive.

Simulating a response

During a behavioural response, the transitions between different dive types are likely to change. Thus, the transition probability matrix changes from G_n to G_r, a matrix containing the probabilities of switching between the K dive types during a response to disturbance. The properties of the dive types are also likely to be different during a response. In order to simulate different types of response, the shape parameter α, rate parameter β, and correlation matrix κ used to simulate particular dive properties were modified during the simulated response. Full details of responses simulated and values of the parameters used are described later in this chapter.

Parameterising normal (baseline) behaviour

The parameters for the seal dive simulator were based on dive data collected from the same 24 harbour seals described in the horizontal simulation above. Dives which overlapped with pile driving construction activity were removed from the dataset. The dive data used consisted of 686,151 dives in total. Dives were recorded when the tag was submerged below 1.5 m for eight seconds, and each dive record contained a measurement of the maximum dive depth reached (in metres), duration of the dive (in seconds) and duration of the post-dive surface interval (time shallower than 1.5 m, in seconds). There were also nine depth points recorded per dive, giving more detailed information on the approximate shape of the dive.

Dividing dives into phases

To quantify the overall shape of the dives, each dive was divided into four phases of different durations: descent, bottom, ascent and surface. Interpolation between the recorded depth points was used to determine the overall dive profile at a time resolution of 0.1 seconds. Seals were considered to start the bottom phase of a dive when their dive depth was $\geq$85% of the maximum depth of that dive (Baechler et al., 2002; Wilson et al., 2014). It was assumed that the bottom phase ended when the seal swam shallower than this 85% threshold for the last time in that dive, i.e. if a seal went up and then back down again, this was all considered to be part of the bottom phase. The descent and ascent phases were considered to be, respectively, the time periods before and after this identified bottom phase (Figure 2.3). The surface phase was the post-dive surface interval which is already recorded separately on the tag.

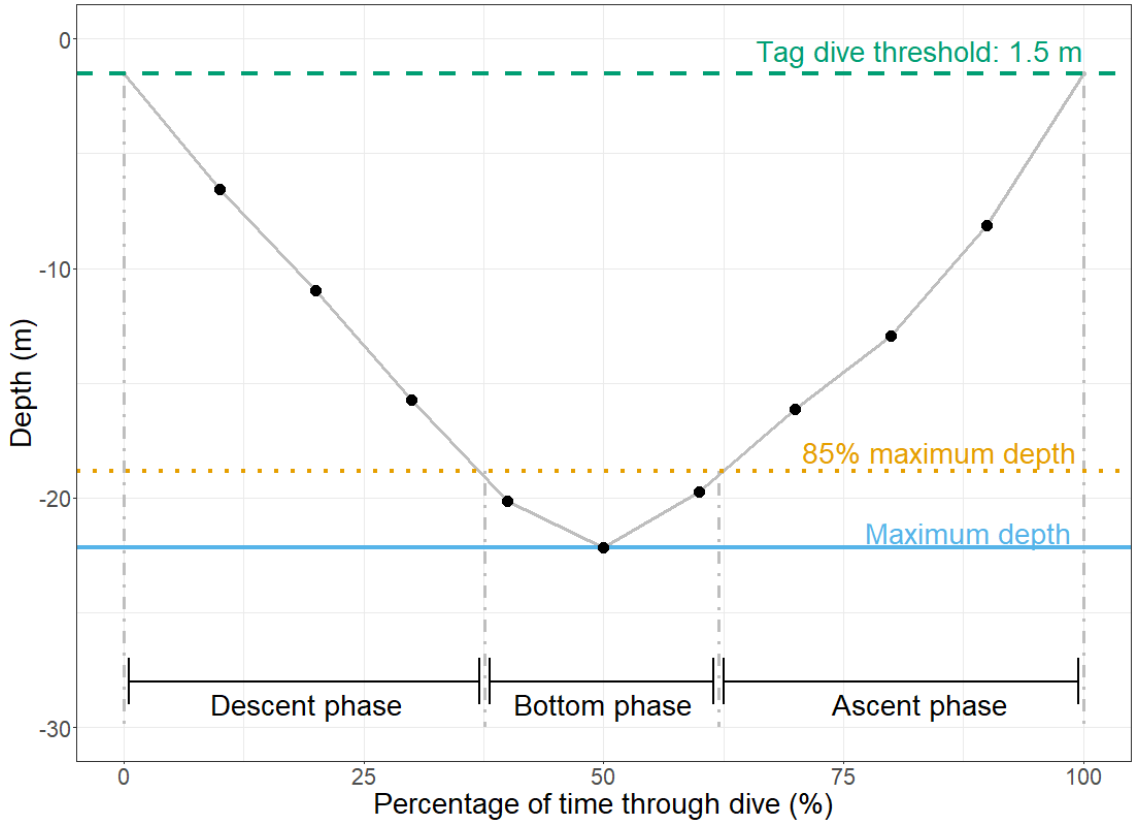


Figure 2.3 An example seal dive (from individual pv42-162-12) divided into phases: descent, bottom and ascent. Shown are the nine recorded depth points (black dots) and the resulting interpolated dive profile between them (grey solid lines). The maximum depth of the dive was 22.0 m (blue solid line), and the bottom phase occurs at ≥85% of the maximum at 18.8 m (orange dotted line). The tag begins and ends recording the dive at 1.5 m depth (green dashed line).

Clustering to identify dive types

Once the dives in the example dataset were divided into phases, a series of metrics were chosen to summarise the overall dive shape. These chosen metrics were based on the data available and previous studies of pinniped diving behaviour and dive classification (Baechler et al., 2002; Blanchet et al., 2015; Lesage et al., 1999; Wilson et al., 2014). The metrics that were calculated for each dive were:

- Maximum depth of the dive (m);
- Total dive duration (s) – sum of descent, bottom and ascent phases;
- Surface phase duration (s);
- Descent rate (m s^{-1});
- Ascent rate (m s^{-1});
- Proportion of the dive in the bottom phase.

These six metrics were then used to perform K-means clustering to split the dives into different groups, using the R package `cluster` (Maechler et al., 2019). K-means clustering aims to group data observations into separate clusters based on minimising the multivariate distance between each of the observations and the mean of the proposed nearest group (Hartigan and Wong, 1979). To determine the optimal number of clusters, the overall within-cluster sum-of-squares was compared between runs with different numbers of clusters (from 2–10 clusters).

Extracting the parameters of each dive type

Once dives were divided into clusters, the parameters of each identified cluster were extracted for use in the dive simulator. Within each cluster, a gamma distribution was fit to each dive metric required for the simulator, using the R package `fitdistrplus` (Delignette-Muller and Dutang, 2015). This provided a shape and rate parameter for each of the following metrics, for each dive type:

- Maximum dive depth (m);
- Descent duration (s);
- Bottom duration (s);
- Ascent duration (s);

- Surface duration (post-dive) (s).

The correlation matrix between the metrics within each cluster was also extracted for use in the simulator. The time series of transitions between the identified dive types was used to extract the transition rates between each pair of dive types, accounting for known gaps in the data set (due to missing data, removed dives during wind farm construction). This was used to parameterise a full transition probability matrix for the simulator.

Parameterising a response

In the simulation study, I aimed to simulate three different types of possible behavioural response. Each type was considered to be a potential response one may expect to see in the dive data of a seal exposed to noise disturbance.

1. **Surface dive type.**
   During the response, the seal exhibits one "surface-type" dive (one of the identified dive clusters) with an extended surface duration. I simulated both a weak response, where the surface duration is extended slightly (∼40% increase), and a strong response where the surface duration is more extreme (∼80% increase). Captive studies have observed increased time at the surface in harbour seals exposed to playbacks of pile driving sounds (Kastelein et al., 2018a) and acoustic deterrent sounds (Kastelein et al., 2017), and hooded seals exposed to sonar signals (Kvadsheim et al., 2010). During these responses, seals may be stationary at the surface or actively swimming.

2. **Surfacing after a normal dive.**
   During the response, the seal carries out a dive (of any dive type) as normal, but afterwards has an extended surface duration for one dive. I simulated both a weak response, where the surface duration is extended slightly (∼40% increase on the surface dive type mean surface duration), and a strong response where the surface duration is more extreme (∼80% increase on the surface dive type mean surface duration).

3. **Travelling.**
   Compared to baseline periods, during the response the seal is more likely to switch into the "travel-type" dive (one of the identified dive clusters), and the travel dives have a shallower depth, lower descent rate, and a lower proportion of the dive in the bottom phase (relative to travel dives in baseline periods). The aim was to

emulate dives which contain more horizontal than vertical movement (similar to a flee response), even though horizontal movement is not recorded explicitly in the dive data. I simulated both a weak response, where the travel dive is moderately likely (transition probabilities of up to ∼80%) and moderately modified (∼50% reduction in maximum depth and ∼75% reduction in bottom duration), and a strong response where the travel dive is highly likely (transition probabilities of up to ∼90%) and highly modified (∼75% reduction in maximum depth and ∼90% reduction in bottom duration). I also simulated both a short (6 dives) and long (36 dives) response of this type. Tagged grey seals (*Halichoerus grypus*) in the Netherlands were observed to show changes in their dive behaviour during pile driving, in particular decreased descent speeds and shallower dives (Aarts et al., 2018).

To generate each of the response types, I modified the parameters of the multivariate gamma distributions and/or the transition probability matrix in one or more of the dive types, to generate biologically realistic responses.

Simulation setup

Timeline setup

I ran 500 simulations of seal diving data for each type of response. For each simulation, I generated 30 days of baseline (pre-exposure) data, followed by 300 dives (∼1 day). As above, for each simulated response type, dives after the baseline period were modified to emulate responses of different types and durations. To estimate the false positive rate of the method, a control set of 500 simulations were also generated where no response was inserted (i.e. no dive parameters were modified).

Inputs for Mahalanobis distance

In all analyses of the simulated dive data, the metrics used as inputs for the Mahalanobis distance analyses were: maximum dive depth, total dive duration, surface duration, proportion of dive in the bottom phase and descent rate (Table 2.4). As in the horizontal simulations, here I chose a standard set of input variables for all simulations and did not investigate how using different inputs may affect the performance of the method. Instead, it is recommended that individual studies carefully consider the inputs they use are appropriate and informative.

Comparison of different method implementations

For each simulated dataset, a variety of implementations of Mahalanobis distance were

Table 2.4 Mahalanobis distance options for vertical simulations. The window sizes and overlap amounts are presented as the number of dives within each window. All combinations of options were compared, resulting in 40 total options.

Mahalanobis distance specification	**Values and options used**
Inputs (same for all simulations):	Maximum depth (m), Total dive duration (s), Surface duration (s), Proportion of dive in bottom phase, Descent rate (m s^{-1})
Baseline data availability:	5 days; 20 days
Covariance matrix used:	All data; Baseline only
Reference window position:	Fixed; Consecutive
Window size (and overlap) in number of dives:	1 (0); 2 (0, 1); 3 (0, 1, 2); 4 (0, 1, 2, 3)

carried out. I varied the size of the windows used (1–4 dives), the overlap amount of the windows (0–3 dives), the position of the reference window (fixed or consecutive), and the covariance matrix used (all data or baseline data only) (Table 2.4). In total, this resulted in 40 different implementations of Mahalanobis distance that were investigated (see Table 2.4). I also investigated the effect of having different amounts of baseline data available, by running analyses with 5 days and 20 days of baseline data for the response simulations, and 1–30 days for the control simulation.

2.3.3 Detecting simulated responses and quantifying performance

In both the horizontal and vertical movement simulations, Mahalanobis distance (Equation 2.1) was calculated over the entire timeline of each simulated dataset. Different implementations of the method were applied to compare the performance in each scenario, as detailed in Table 2.2 and Table 2.4 for the horizontal and vertical data respectively.

In each case, a randomisation test was used to attempt to identify simulated responses by sampling 1000 mock exposure periods from the baseline data for each simulated animal. All exposure and mock-exposure periods were two hours long. For each individual simulation, the maximum Mahalanobis distance of each of the mock exposure periods was used to determine the 95^{th} percentile of Mahalanobis distances. If the maximum Mahalanobis

distance observed during the exposure period was greater than this value, this was identified as a behavioural response.

For each simulated scenario and Mahalanobis distance option, I calculated the number of simulations (out of 500) in which a response was detected. For simulations involving a simulated response, the response detection rate was therefore:

$$\text{Response detection rate} = \frac{\text{Number of responses detected}}{\text{Total number of response simulations}}$$

For control simulations involving no simulated response, the false positive rate was similarly:

$$\text{False positive rate} = \frac{\text{Number of responses detected}}{\text{Total number of control simulations}}$$

In both cases the denominator was $n = 500$.

2.3.4 Software used

All analyses were conducted in R (version 4.0.4, R Core Team, 2020), with additional packages `CircStats` (Agostinelli and Lund, 2018), `cluster` (Maechler et al., 2019), `dplyr` (Wickham et al., 2020), `fitdistrplus` (Delignette-Muller and Dutang, 2015), `momentuHMM` (McClintock and Michelot, 2018), `lcmix` (Dvorkin, 2012), and `StatMatch` (D'Orazio, 2019). The R packages `ggplot2` (Wickham, 2016), `ggpubr` (Kassambara, 2020), `raster` (Hijmans, 2020), and `viridis` (Garnier, 2018) were also used for data visualisation.

2.4 Results

2.4.1 Horizontal movement: Seal location data

Simulation parameters used

The HMMs fit to the regularised tracks provided a good fit for the two states proposed by the simulation model: one state with longer step lengths and lower variability in turn angles (state 1), and one with shorter step lengths and higher variability in turn angle (state 2)(see Figures S1 and S2 in Appendix). Overall, across all animals, the models estimated a mean step length of 691 metres and mean heading concentration of 0.70 for state 1, and mean step length of 291 metres and heading concentration of 0.58 for state 2.

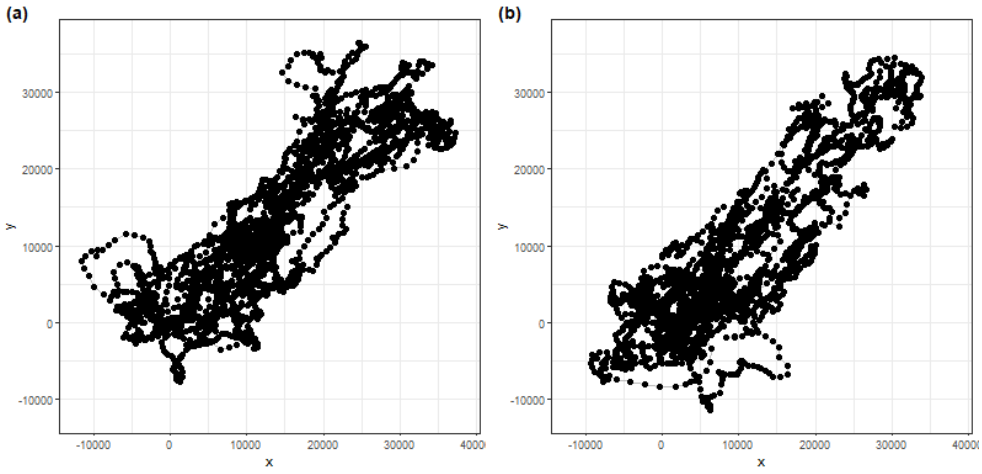


Figure 2.4 Two example simulated seal tracks of baseline behaviour (simulation length = 30 days). The two attraction points are positioned at the x-y coordinates of (0,0) and (30000,30000).

The transition probabilities of switching between the two states during baseline were also obtained for use within the simulation (Table 2.5). Full details of all the parameter values chosen for the simulation are provided in Table 2.5. Using these parameters, simulated tracks were considered to resemble normal central-place foraging behaviour (Figure 2.4), as well as possible behaviour during a behavioural response (Figure 2.5).

Simulation results

The false positive rates of all Mahalanobis distance implementation options were generally low (mean of 6%, maximum of 24%) and were similar across the different options. False positive rates were highest when limited baseline data were available, and decreased as baseline duration became larger (Figure 2.6). In general, the rates settled around the expected error rate (∼5%) with at least 5 days of baseline data. A 95% threshold was used to identify responses and so, by chance, only 5% of baseline behaviours would be more extreme than this threshold. False positives were slightly higher when the covariance matrix was used on baseline data only (Figure 2.6(b)), but only when the baseline data were extremely limited (1 day).

Response detection rates differed by simulated response (Figures S4, 2.7). Detection rates were higher when simulated responses were strong (up to 79%), compared to when simulated responses were weak (up to 64%). Longer simulated responses also had higher detection rates (up to 79%) than shorter simulated responses (up to 54%). When responses

Table 2.5 Summary of horizontal movement simulation setup for baseline data, weak disturbance data, and strong disturbance data. Detailed below are the parameter values used for each simulation, as well as the duration and initial distance of each scenario. It should be noted that the extremely large (and small) values required to parameterise the simulations are likely due to the differences in scale being accommodated in the simulation: simulations at the individual step level are in metres, but the attraction and repel points operate over 10's of kilometres.

	Baseline	**Weak Disturbance**	**Strong Disturbance**
Behavioural states			
G_0 Transition probability	$\begin{pmatrix} 0.93 & 0.07 \\ 0.07 & 0.93 \end{pmatrix}$	$\begin{pmatrix} 0.93 & 0.07 \\ 0.75 & 0.25 \end{pmatrix}$	$\begin{pmatrix} 0.93 & 0.07 \\ 0.95 & 0.05 \end{pmatrix}$
η Switch strength	1,000	no change	no change
Attraction and repulsion			
ω_a Attract strength	3×10^{-5}	no change	no change
ω_r Repel strength	-	5,000	8,000
π Keep attraction point	0.7	no change	no change
Step lengths (speeds)			
v_1 Mean speed ($\pm$ ϵ_1)	691 ($\pm$ 210)	no change	no change
v_2 Mean speed ($\pm$ ϵ_2)	291 ($\pm$ 146)	no change	no change
κ_1 Correlation	0.8	no change	no change
κ_2 Correlation	0.7	no change	no change
$v_{max,1}$ Max mean speed	-	1400	1400
$v_{max,2}$ Max mean speed	-	291	291
Headings (movement direction)			
ρ_1 Mean concentration	0.70	no change	no change
ρ_2 Mean concentration	0.58	no change	no change
γ_a Concentration attract	2×10^{-5}	no change	no change
γ_r Concentration repel	-	5,000	8,000
General setup			
CEE duration	2 hrs	2 hrs	2 hrs
Response duration	-	0.5; 3 hrs	0.5; 3 hrs
$d_{r,0}$ Initial distance	-	5; 15; 25 km	5; 15; 25 km

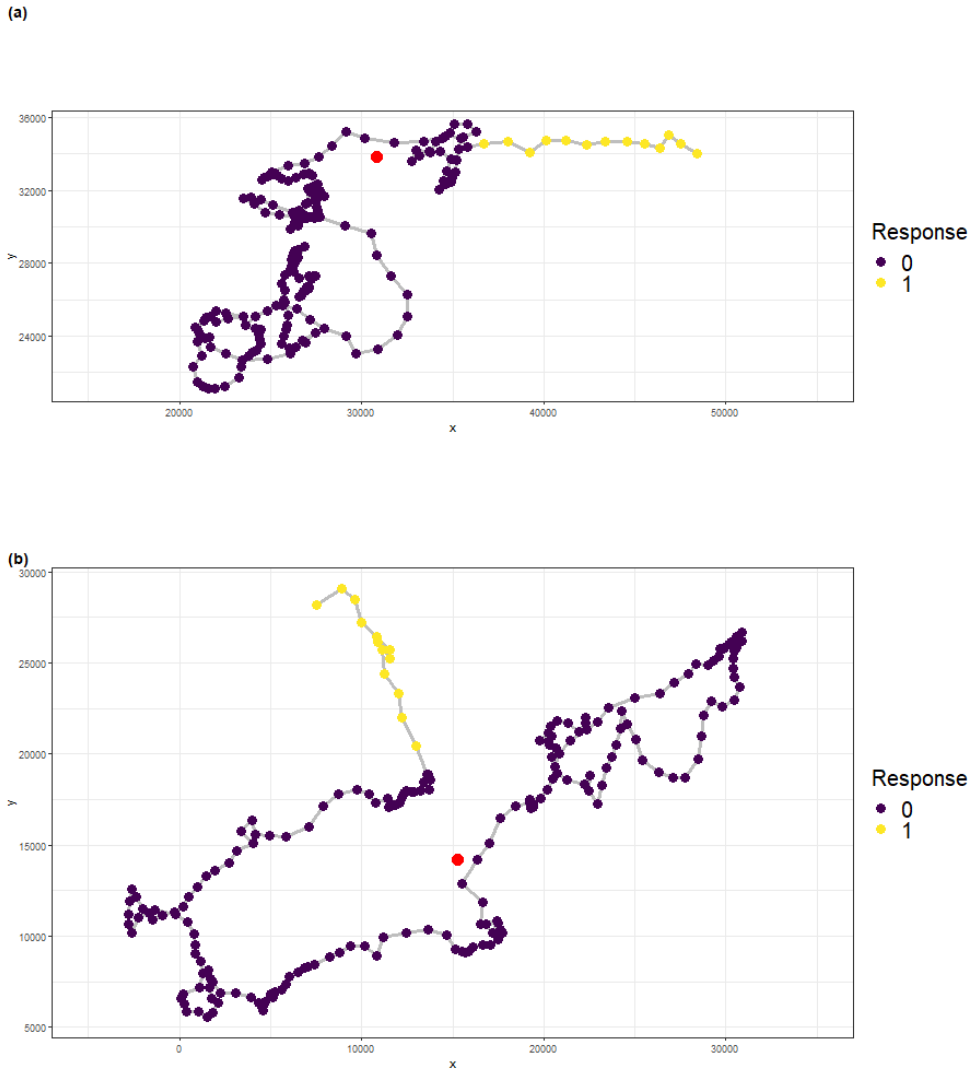


Figure 2.5 Two example simulated tracks of responses. Each seal track shows 200 simulated locations of normal behaviour (purple) followed by a simulated response track (yellow). The red dots denote the location of the simulated repel centre. Both (a) and (b) are strong, long responses with an initial distance of 5 km from the centre of repulsion.

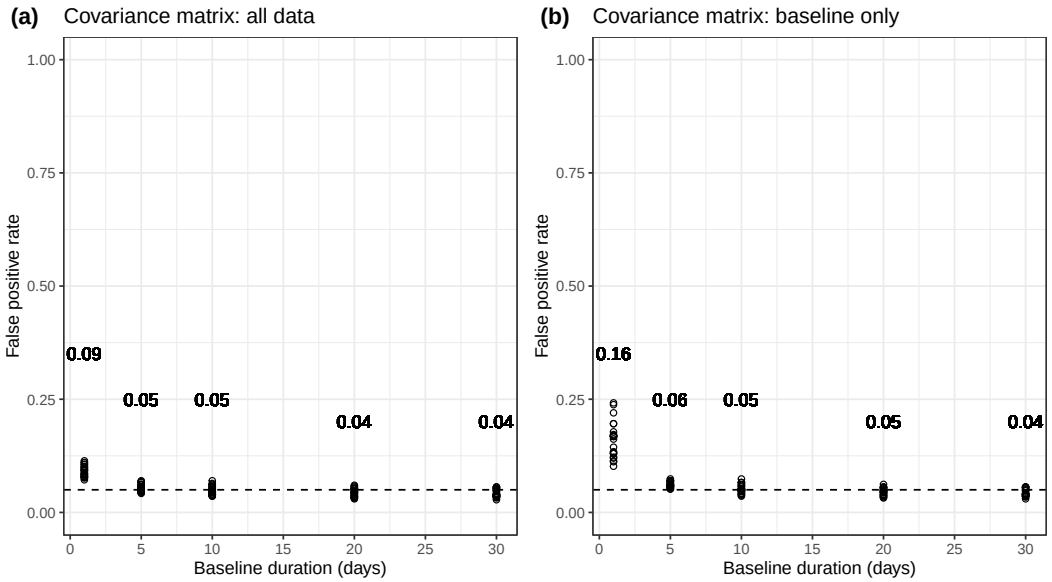


Figure 2.6 The false positive rate of Mahalanobis distance methods with varying baseline data durations for horizontal data. Each dot represents the false positive rate for one of the 24 Mahalanobis distance method implementation options, when the covariance matrix is calculated across all the data (a) and across the baseline data only (b). The annotated numbers give the mean false positive rate across all methods. Each Mahalanobis distance option is detailed in Table 2.6. The dashed line denotes 5%.

Table 2.6 Mahalanobis distance options for horizontal simulations.

Mahalanobis distance option	**Reference window**	**Window size (minutes)**	**Window overlap**
1	Consecutive	15	0.1
2	Fixed	15	0.1
3	Consecutive	20	0.1
4	Fixed	20	0.1
5	Consecutive	25	0.1
6	Fixed	25	0.1
7	Consecutive	30	0.1
8	Fixed	30	0.1
9	Consecutive	15	0.5
10	Fixed	15	0.5
11	Consecutive	20	0.5
12	Fixed	20	0.5
13	Consecutive	25	0.5
14	Fixed	25	0.5
15	Consecutive	30	0.5
16	Fixed	30	0.5
17	Consecutive	15	0.9
18	Fixed	15	0.9
19	Consecutive	20	0.9
20	Fixed	20	0.9
21	Consecutive	25	0.9
22	Fixed	25	0.9
23	Consecutive	30	0.9
24	Fixed	30	0.9

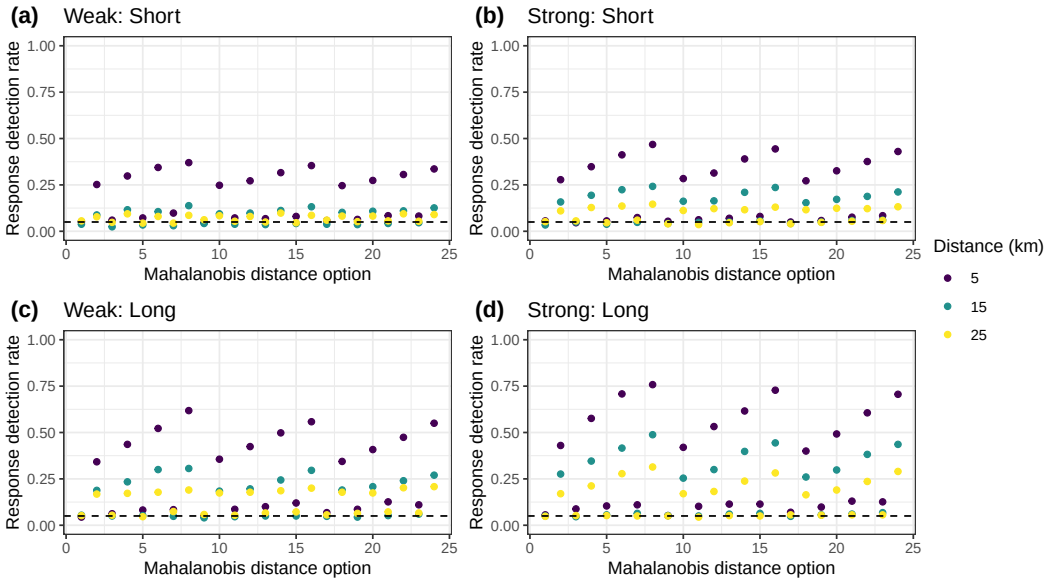


Figure 2.7 Response detection rates for horizontal simulated responses, with the covariance matrix calculated over the baseline data for the simulated animal. Each dot is the response detection rate (out of 500 simulations) for responses to centres of repulsion of different strengths (weak, strong), durations (short, long), and initial distances (5, 15, 25 km; coloured dots). All simulations had 20 days of baseline data. Each Mahalanobis distance option is detailed in Table 2.6. The dashed line denotes 5%.

were simulated to begin at closer initial distances to the centre of repulsion, responses were detected more reliably (up to 79% at 5 km, up to 55% at 10 km, up to 42% at 25 km). In most simulations, detection rates were slightly higher when the covariance matrix was used on the baseline data only (Figure 2.7), compared to when all data were used (Figure S4). Using only the baseline data increased the number of responses detected (out of 500 simulations) by a mean of 12 (min $= -4$, max $= 94$).

Response detection rates also differed by Mahalanobis distance implementation option (Figure 2.8). For all simulation scenarios, increasing the window size increased detection rates, although the benefit of increasing window size was less apparent when a shorter response was simulated ($\sim$28% to $\sim$48%), compared to a longer response ($\sim$43% to $\sim$76%) (Figure 2.8(a)(b)). Using a reference window that was fixed in position prior to the exposure period also had markedly higher detection rates (up to 79%) than using a pair of consecutive sliding windows (up to 15%) (Figure 2.8). The overlap amount of the windows chosen did not appear to have a substantial effect on the overall detection rates (detection rates generally did not differ more than 5% between different window overlap amounts). The detection rates and differences between implementation options were also similar when only

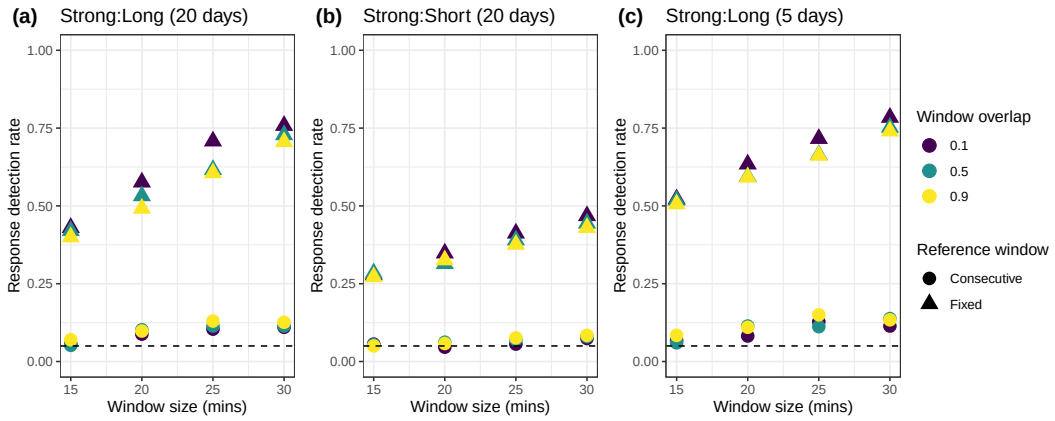


Figure 2.8 Response detection rates by Mahalanobis distance implementation option, for three example scenarios with horizontal data: (a) a strong and long response with 20 days baseline data, (b) a strong and short response with 20 days baseline data, and (c) a strong and long response with 5 days baseline data. Shown are the detection rates by window size, window overlap proportion (colours), and reference window position (symbols). All examples used the covariance matrix on the baseline data only. The dashed line denotes 5%.

5 days of baseline data were available (Figure 2.8(c)). Overall, each of these observed patterns between implementation options were consistent across different simulated response strengths, durations, distances, and baseline durations.

2.4.2 Vertical movement: Seal dive data

Simulation parameters used

Clustering of the dive data identified five dive types as the most optimal grouping of the data (Figure 2.10). The first dive type was characterised by an extremely long post-dive surface interval (Table 2.7; Figure S5). The second and third dive types both consisted of a high proportion of time spent in the bottom phase (Figure S7), as well as a moderate depth and long dive duration. It was most likely that these clusters represented two different types of foraging dive (Table 2.7). Dives in type four were generally very shallow (Figure 2.9) and short (Figure S6), suggesting they may represent some sort of resting behaviour (Table 2.7). Dives in type five were most similar to the foraging dive types (two and three), but generally were shallower (Figure 2.9), shorter duration (Figure S6), and spent less time in the bottom phase (Figure S7; Table 2.7). Dive type five is therefore most likely to represent dives used in travelling. Examining the original locations of these recorded dive types further supports these potential functions (Figure S3). Dive types one

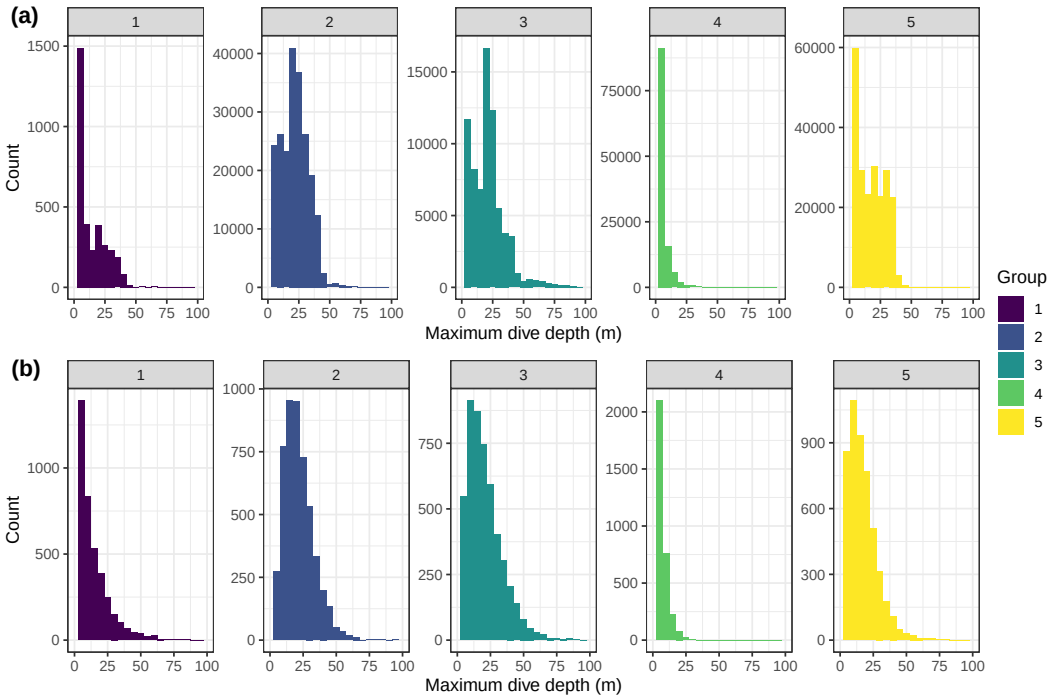


Figure 2.9 Histograms of maximum dive depth for each dive type (1–5) for original (a) and simulated (b) dive data. (a) The original data are all recorded dives from 24 tagged harbour seals in the UK. (b) The simulated data consist of 5000 simulated dives of each dive type.

(surface) and four (rest) were mostly found clustered near land and the known haulout sites, whereas the other types were more widely spread offshore and across the study area. Dive metrics simulated directly (e.g. surface duration, Figure S5; maximum depth, Figure 2.9) and derived from the simulated dive metrics (e.g. total dive duration, Figure S6; proportion bottom time, Figure S7) resembled the distributions of the original data. The final parameters used to simulate baseline dive behaviour as well as response types 1, 2 and 3 are detailed in Tables 2.8, 2.9 and 2.10 respectively.

Simulation results

The false positive rates of all dive Mahalanobis distance implementation options were generally low (mean of 6%, maximum of 15%) and were similar across the different options. As observed with the horizontal movement simulations, false positive rates were highest when limited baseline data were available, and decreased as baseline duration increased (Figure 2.11). In general, the rates settled around the expected error rate (∼5%) with at least 10 days of baseline data. False positives were marginally higher when the covariance

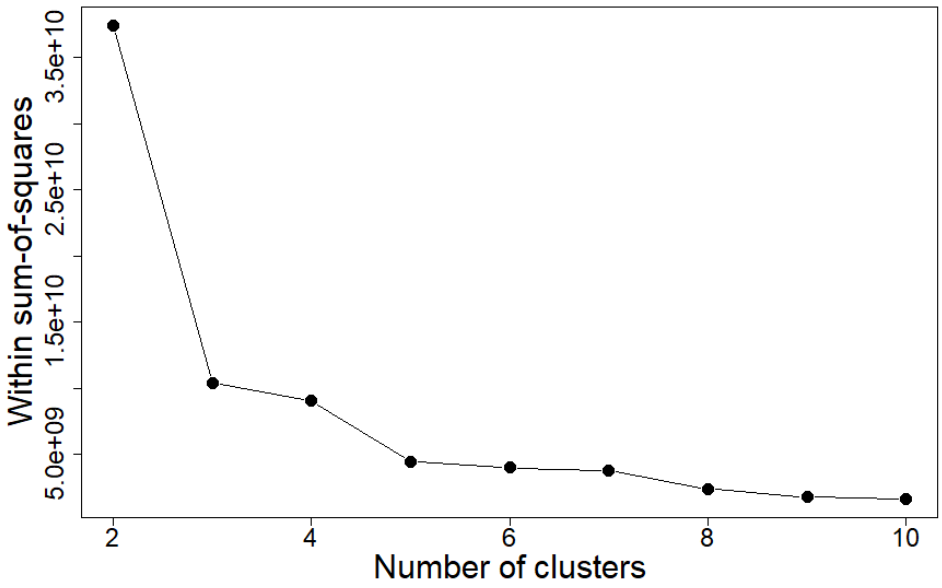


Figure 2.10 Elbow plot used to determine the number of clusters of dive types from K-means clustering. Shown is the total within sum-of-squares for when dives are clustered into different numbers of clusters (2–10). The reduction in within sum-of-squares levels off when 5 or more clusters are used.

Table 2.7 The five dive types identified from the K-means clustering and used in the dive simulator. For each dive type, listed are the means of each metric used to inform the clustering: maximum depth, dive duration, surface duration, descent rate, ascent rate, and proportion of the dive in the bottom phase. Also provided are the possible biological function of each identified dive type, and the percentage of each dive type in the original dataset.

Type	**Maximum depth (m)**	**Total dive duration (s)**	**Surface duration (s)**	**Descent rate (m s^{-1})**	**Ascent rate (m s^{-1})**	**Proportion bottom phase**	**Possible biological function**	**% of dives**
1	11.1	94	2546	0.32	0.36	0.56	**Surface**	0.6 %
2	21.6	238	52	0.64	0.65	0.73	**Foraging 1**	31.4 %
3	21.0	348	57	0.47	0.48	0.77	**Foraging 2**	10.7 %
4	4.8	42	84	0.25	0.24	0.53	**Rest**	24.0 %
5	17.1	149	46	0.58	0.59	0.60	**Travel**	33.2 %

Table 2.8 Summary of vertical (dive) movement simulation setup for baseline data, weak response data, and strong response data for response type 1: surface dive type. Dots denote parameters that are unchanged from the baseline.

	Baseline	**Type 1 Response: Weak**	**Type 1 Response: Strong**
Dive type transition probabilities			
	G_n	G_r	G_r
$\begin{pmatrix} \psi_{1,1} & \cdots & \psi_{1,5} \\ \vdots & \ddots & \vdots \\ \psi_{5,1} & \cdots & \psi_{5,5} \end{pmatrix}$	$\begin{pmatrix} 0.14 & 0.05 & 0.02 & 0.62 & 0.17 \\ < 0.01 & 0.72 & 0.08 & 0.03 & 0.16 \\ < 0.01 & 0.23 & 0.70 & 0.03 & 0.04 \\ 0.01 & 0.04 & 0.01 & 0.73 & 0.20 \\ < 0.01 & 0.16 & 0.01 & 0.14 & 0.69 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$
Dive metrics (mean of each dive type 1–5)			
	1 *2* *3* *4* *5*	*1* *2* *3* *4* *5*	*1* *2* *3* *4* *5*
Max dive depth (m)	11.1 21.6 21.0 4.8 17.1	· · · · ·	· · · · ·
Descent duration (s)	21 32 40 11 30	· · · · ·	· · · · ·
Bottom duration (s)	54 175 270 21 92	· · · · ·	· · · · ·
Ascent duration (s)	19 31 38 10 28	· · · · ·	· · · · ·
Surface duration (s)	2546 52 57 84 46	3600 · · · ·	4800 · · · ·
General setup			
CEE duration	2 hrs	2 hrs	2 hrs
Response duration	-	1 dive	1 dive

Table 2.9 Summary of vertical (dive) movement simulation setup for baseline data, weak response data, and strong response data for response type 2: surfacing after a normal dive. Dots denote parameters that are unchanged from the baseline.

	Baseline	**Type 2 Response: Weak**	**Type 2 Response: Strong**
Dive type transition probabilities			
	G_n	G_r	G_r
$\begin{pmatrix} \psi_{1,1} & \cdots & \psi_{1,5} \\ \vdots & \ddots & \vdots \\ \psi_{5,1} & \cdots & \psi_{5,5} \end{pmatrix}$	$\begin{pmatrix} 0.14 & 0.05 & 0.02 & 0.62 & 0.17 \\ < 0.01 & 0.72 & 0.08 & 0.03 & 0.16 \\ < 0.01 & 0.23 & 0.70 & 0.03 & 0.04 \\ 0.01 & 0.04 & 0.01 & 0.73 & 0.20 \\ < 0.01 & 0.16 & 0.01 & 0.14 & 0.69 \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$	$\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$
Dive metrics (mean of each dive type 1–5)			
	1 *2* *3* *4* *5*	*1* *2* *3* *4* *5*	*1* *2* *3* *4* *5*
Max dive depth (m)	11.1 21.6 21.0 4.8 17.1	· · · · ·	· · · · ·
Descent duration (s)	21 32 40 11 30	· · · · ·	· · · · ·
Bottom duration (s)	54 175 270 21 92	· · · · ·	· · · · ·
Ascent duration (s)	19 31 38 10 28	· · · · ·	· · · · ·
Surface duration (s)	2546 52 57 84 46	3600 3600 3600 3600 3600	4800 4800 4800 4800 4800
General setup			
CEE duration	2 hrs	2 hrs	2 hrs
Response duration	-	1 dive	1 dive

Table 2.10 Summary of vertical (dive) movement simulation setup for baseline data, weak response data, and strong response data for response type 3: travelling. Dots denote parameters that are unchanged from the baseline.

	Baseline	**Type 3 Response: Weak**	**Type 3 Response: Strong**
Dive type transition probabilities			
	G_n	G_r	G_r
$\begin{pmatrix} \psi_{1,1} & \dots & \psi_{1,5} \\ \vdots & \ddots & \vdots \\ \psi_{5,1} & \dots & \psi_{5,5} \end{pmatrix}$	$\begin{pmatrix} 0.14 & 0.05 & 0.02 & 0.62 & 0.17 \\ < 0.01 & 0.72 & 0.08 & 0.03 & 0.16 \\ < 0.01 & 0.23 & 0.70 & 0.03 & 0.04 \\ 0.01 & 0.04 & 0.01 & 0.73 & 0.20 \\ < 0.01 & 0.16 & 0.01 & 0.14 & 0.69 \end{pmatrix}$	$\begin{pmatrix} 0.03 & 0.01 & < 0.01 & 0.15 & 0.80 \\ < 0.01 & 0.17 & 0.02 & 0.01 & 0.80 \\ < 0.01 & 0.06 & 0.18 & 0.01 & 0.75 \\ < 0.01 & 0.01 & < 0.01 & 0.18 & 0.80 \\ < 0.01 & 0.02 & < 0.01 & 0.02 & 0.95 \end{pmatrix}$	$\begin{pmatrix} 0.02 & 0.01 & < 0.01 & 0.07 & 0.90 \\ < 0.01 & 0.09 & 0.01 & < 0.01 & 0.90 \\ < 0.01 & 0.04 & 0.11 & < 0.01 & 0.85 \\ < 0.01 & 0.01 & < 0.01 & 0.09 & 0.90 \\ < 0.01 & 0.01 & < 0.01 & 0.01 & 0.98 \end{pmatrix}$
Dive metrics (mean of each dive type 1–5)			
	1 *2* *3* *4* *5*	*1* *2* *3* *4* *5*	*1* *2* *3* *4* *5*
Max dive depth (m)	11.1 21.6 21.0 4.8 17.1	· · · · 10.0	· · · · 5.0
Descent duration (s)	21 32 40 11 30	· · · · 140	· · · · 200
Bottom duration (s)	54 175 270 21 92	· · · · 20	· · · · 10
Ascent duration (s)	19 31 38 10 28	· · · · 140	· · · · 200
Surface duration (s)	2546 52 57 84 46	· · · · ·	· · · · ·
General setup			
CEE duration	2 hrs	2 hrs	2 hrs
Response duration	-	6; 36 dives	6; 36 dives

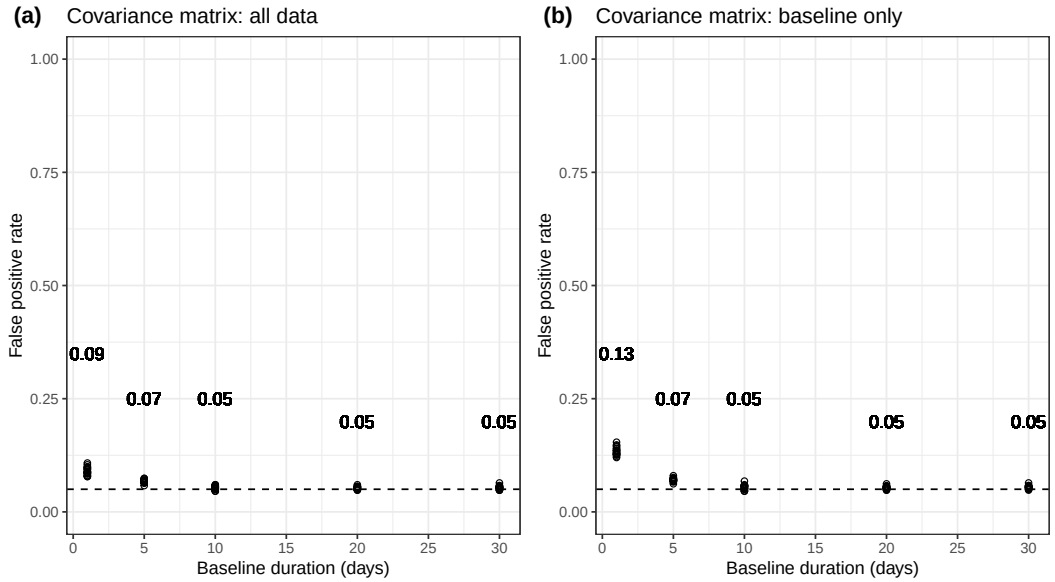


Figure 2.11 The false positive rate of Mahalanobis distance methods with varying baseline data durations for vertical data. Each dot represents the false positive rate for one of the 20 Mahalanobis distance method implementation options, when the covariance matrix is calculated across all the data (a) and across the baseline data only (b). The annotated numbers give the mean false positive rate across all methods. Each Mahalanobis distance option is detailed in Table 2.11. The dashed line denotes 5%.

matrix was used on baseline data only (Figure 2.11(b)), but only when the baseline data were extremely limited (1 day).

In most simulations, detection rates were higher when the covariance matrix was calculated from the baseline data only (Figure 2.12), compared to when all data were used (Figure S8). Using only the baseline data increased the number of responses detected (out of 500 simulations) by a mean of 19 (min = −4, max = 309). This improvement was particularly apparent for simulated response type 3 (travel), with a mean increase of 31 detected responses (min = −3, max = 309) across all scenarios, compared to a mean increase of 6 for response type 1 (min = −2, max = 26) and a mean increase of 7 for response type 2 (min = −4, max = 33). Response detection rates differed by the simulated type of response behaviour (Figure 2.12). For response type 1 (changing to an unusual surface-type dive) and type 2 (a long surface interval after a normal dive), strong responses were detected with high certainty (∼95–100% of simulations). Weak responses of type 1 and 2 had relatively lower detection rates (∼55–98% of simulations). For dive response type 3 (altered travel dives), detection rates were higher for longer responses (up to 100%) than shorter responses (up to 97%), and higher for stronger (up to 100%) than weak (up to 68%) responses.

Table 2.11 Mahalanobis distance options for vertical simulations.

Mahalanobis distance option	Reference window	Window size (dives)	Window overlap (dives)
1	Consecutive	1	0
2	Fixed	1	0
3	Consecutive	2	0
4	Fixed	2	0
5	Consecutive	2	1
6	Fixed	2	1
7	Consecutive	3	0
8	Fixed	3	0
9	Consecutive	3	1
10	Fixed	3	1
11	Consecutive	3	2
12	Fixed	3	2
13	Consecutive	4	0
14	Fixed	4	0
15	Consecutive	4	1
16	Fixed	4	1
17	Consecutive	4	2
18	Fixed	4	2
19	Consecutive	4	3
20	Fixed	4	3

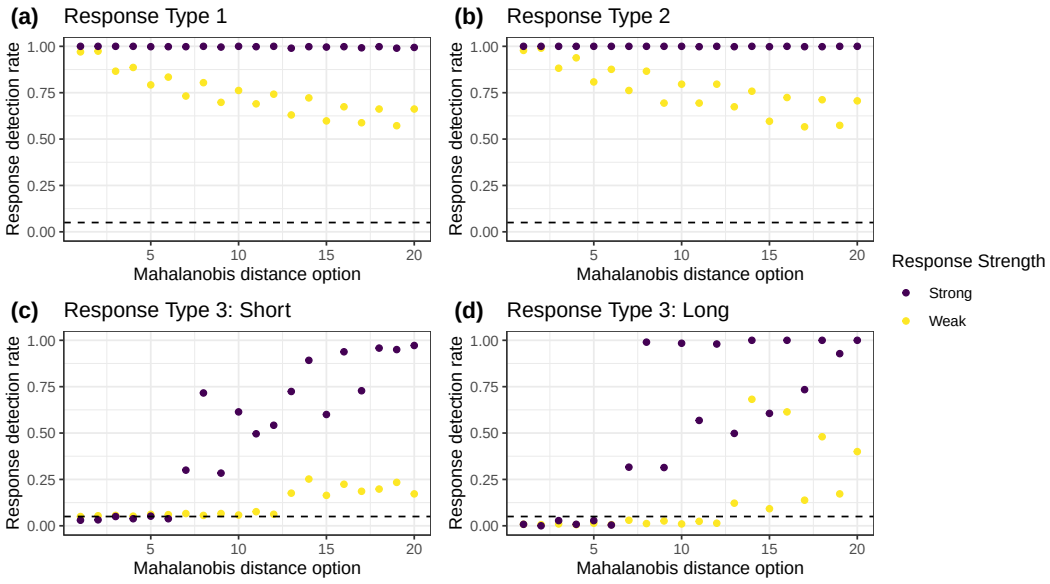


Figure 2.12 Response detection rates for simulated vertical responses, with the covariance matrix calculated over the baseline data for the simulated animal. Each dot is the response detection rate (out of 500 simulations) for responses of different strengths (coloured dots) and types: (a) Type 1: surface dive; (b) Type 2: surface after a normal dive; (c) and (d) Type 3: travel. All simulations had 20 days of baseline data. Each Mahalanobis distance option is detailed in Table 2.11. The dashed line denotes 5%.

Response detection rates also differed by Mahalanobis distance option (Figure 2.13). For response types 1 and 2, detection rates were highest when the time window size was only one dive (up to 100%). Detection decreased with increasing window size (to ∼55–75% for time windows containing four dives), and there was a negligible difference between consecutive or fixed reference windows (Figure 2.13(a)(b)). In contrast, for type 3 responses, detection rates increased as window size increased (Figure 2.13(c)), from ∼5% for time windows of one dive to ∼50–100% for time windows of four dives. Using a fixed reference window also generally improved detection rates for this type of response (up to 100% for fixed windows vs. up to 92% for consecutive windows). In the simulations, the overlap amount of the time windows generally did not have a clear effect on detection rates; however, in response type 3 the detection rate appears to increase with window overlap amount (Figure 2.13(c)(d)). The detection rates and differences between implementation options were also similar when only 5 days of baseline data were available (e.g. Figure 2.13(d)). Overall, each of these observed patterns between implementation options were consistent across the different response strengths, response durations, and baseline durations simulated.

2.5 Discussion

In this chapter, a simulation study was designed to quantify the performance of Mahalanobis distance, a statistical approach used to detect behavioural responses in biologging data. To do this, two data simulators were developed to generate realistic (1) harbour seal GPS locations and (2) harbour seal dive summary records. Different implementations of the Mahalanobis distance approach were applied to the simulated data to investigate the performance of the method in detecting different types of behaviour change. In this discussion, I summarise the main findings of this simulation study, highlight recommendations for future behavioural response studies aiming to use this approach, and discuss the limitations of the seal data simulated in this study.

2.5.1 Performance of Mahalanobis distance

This study used simulated data to quantify both the ability of Mahalanobis distance to detect behavioural responses, and the false positive rate of the method. The false positive rate (proportion of times the method detected a response when none was present) was generally low for both types of data considered (Figures 2.6, 2.11). Over all the control (non-response) simulations, the maximum false positive rate recorded was 24%; however,

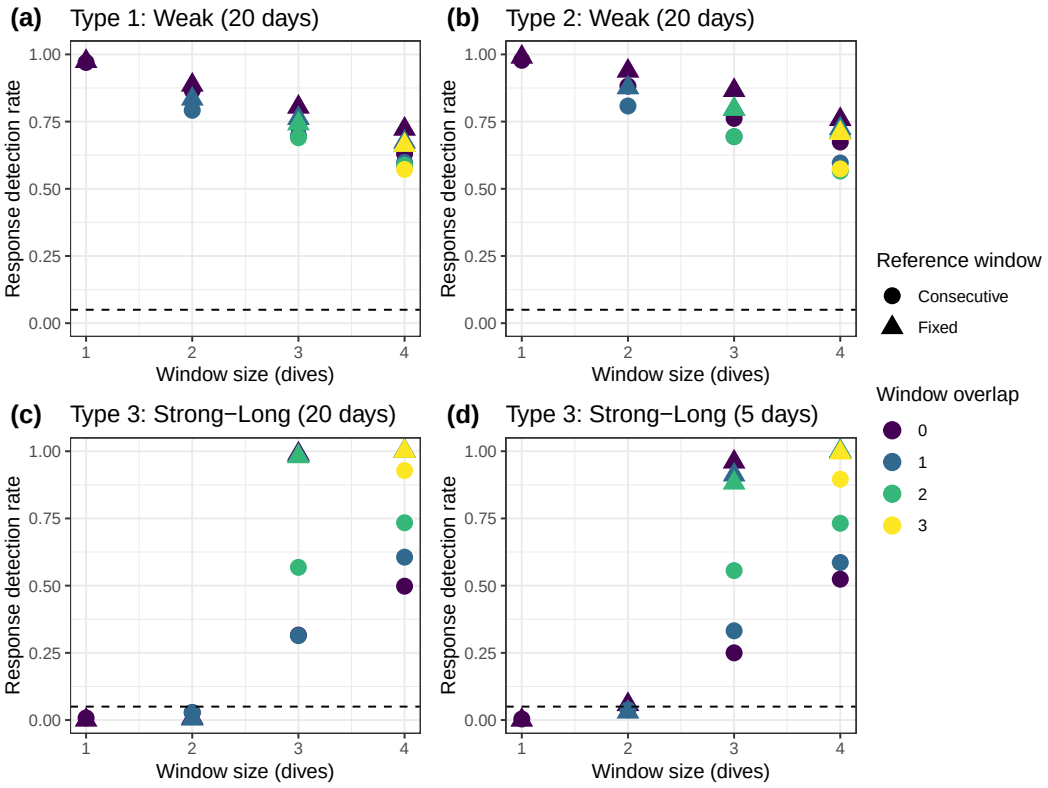


Figure 2.13 Response detection rates by Mahalanobis distance implementation option, for four example scenarios with vertical data: (a) a weak Type 1 response with 20 days baseline data, (b) a weak Type 2 response with 20 days baseline data, (c) a strong and long Type 3 response with 20 days baseline data, and (d) a strong and long Type 3 response with 5 days baseline data. Shown are the detection rates by window size (number of dives), window overlap (number of dives; colours), and reference window position (symbols). All examples used the covariance matrix on the baseline data only. The dashed line denotes 5%.

this was for simulations which had only one day of baseline data. As the duration of baseline data available in the simulations increased, the estimated false positive rate decreased. When the baseline duration is larger, there is more information on what the individual's normal behaviour is and it is more likely to have exhibited its full range of natural behaviours, meaning that the method is less likely to identify normal behaviour changes as unusual. In both the movement and dive simulation studies, when there was at least five days of baseline data the maximum false positive rate across all method implementation approaches was 8%. For most seal tagging studies, five days of baseline data is a feasible target as tags will typically remain attached to the individuals for several months. For cetacean studies with suction-attached tags, the availability of baseline data is likely to be more challenging and in some studies there is only 30 minutes of data recorded before exposure (Miller et al., 2014; Southall et al., 2019a). In order to counteract this limitation, some studies have pooled baseline data from across different individuals in the study population (DeRuiter et al., 2013; Wensveen et al., 2019). Whilst this approach will increase the baseline duration and likely reduce false positives, it is important to be cautious about variation between individuals. With this modification, the method would be attempting to identify behaviour that is unusual across all individuals, rather than behaviour that is unusual for a particular individual. It is recommended that, in order to minimise false positives, future studies using Mahalanobis distance aim to maximise the amount of baseline data collected prior to disturbance.

The choice of how to calculate the covariance matrix affected the performance of Mahalanobis distance. In general, when the method was switched from using all of the data to only using the baseline data for each individual, response detection rates increased while false positives remained relatively stable (e.g. Figure S4 to 2.7). When the covariance matrix is calculated over all the data, this includes the exposure period and any response behaviour within that. The estimated variability in measured behaviours will therefore be larger, making it more challenging for the method to detect a response when one is present. This supports the findings of Brownie et al. (1990), who examined the power of statistical tests used in biological studies attempting to detect the effect of some treatment. They proposed, and demonstrated, that using only control data to estimate variances (and not the pooled variance of the whole data) improved the power of the test. Previous studies with Mahalanobis distance have typically used the covariance matrix over all the data, stating this to be a more conservative approach, as false positives are less likely (Southall et al., 2019a; Stimpert et al., 2014). In contrast, these simulations have highlighted that taking this approach may not always be optimal as many true responses can be missed. It

is recommended that future studies consider using the covariance matrix across the baseline data only, except perhaps in cases where baseline data are extremely limited.

The choice of time window structure also led to different detection rates. In the horizontal simulations, increasing the time window size increased detection rates (Figure 2.8). The window sizes tested were 15–30 minutes, and all simulated responses were at least 30 minutes long. Therefore, as time window size increases, the likelihood of a time window containing more data from the response increases. The mean movement parameters for that window therefore become more extreme, making it more likely to be identified as a response. This pattern was also seen for the dive simulations with a travel response (type 3), where detection increased as window size increased (Figure 2.13(c)). The opposite trend was seen for dive responses type 1 and 2 (Figure 2.13(a)(b)), as the simulated surfacing responses were only 1 dive long and therefore smaller window sizes performed better (as the time windows were not diluted with data points from non-response times). Overall, the choice of overlap between sliding time windows did not appear to have a consistent or substantial effect on detection or false positive rates. This modification affects the resolution with which a response can be detected or reported. There were differences in detection rates between the choice of a consecutive or fixed reference window approach. By some method implementation combinations, the difference between the two approaches was minimal, but the fixed reference window frequently performed better (e.g. Figure 2.8). The fixed approach enables users to identify if the animals behaviour was unusual compared to its average behaviour during baseline, whereas the consecutive approach identifies if the behaviour was unusual compared to its behaviour in the previous time window. The consecutive approach could prove better for some types of behaviour change, depending on the type of response that is trying to be detected (e.g. snap changes in behaviour, see Antunes et al. (2014)). It is recommended that future studies consider carefully the type and duration of potential responses they are interested in detecting, and use this to inform the choice of window size and structure for analysis. The choice of window size should also consider the natural variability in the baseline data, in order to pick a biologically meaningful resolution for the particular species and data type (aiming to minimise variation between baseline time windows). In addition, the choice of overlap amount should consider both computational feasibility and the resolution with which one could identify and report a response (e.g. if the data are at a 15-minute resolution, testing for a response every 15 seconds would probably be inappropriate).

As expected, when simulated responses were stronger or longer, they were more likely to be detected by the method. Increasing the magnitude of an unusual metric or increas-

ing the occurrence of unusual values, increases the likelihood of the time window being unusual overall. In the horizontal simulations, responses which began further away from the repel centre were equivalent theoretically to responses which began closer to the repel centre but were of a short duration (e.g. Figure S4(b)(d)). For responses further away, as the animal moves away from the repel centre, the effect of the repel centre decreases with distance (weighting parameter β_t), meaning there is only a limited amount of time before the response effect becomes undetectable and it decays back to normal behaviour. Similarly, for the simulations of close-range responses, the short duration means the response is only detectable for a limited time. Some simulated response types were easier to detect than others. For example, the dive simulations with response Types 1 and 2 (modified surfacing durations) had consistently high detection rates. Mahalanobis distance is a statistic which is designed to identify extreme values (e.g. unusually long surface duration), and so reductions in values towards the mean (e.g. decreased dive depth) or changes in time budgets (e.g. more time spent travelling than usual) may not be as straightforward to detect. For studies which are interested in broader time allocation between different behaviours, a state-specific approach such as HMMs (e.g. Isojunno et al. (2017)) may be more appropriate. Additionally, it may be more challenging to identify responses in animals (or particular metrics) which naturally exhibit behaviour close to their biological and physiological limitations (cost of transport speeds (Gallon et al., 2007), aerobic dive limits (Kooyman et al., 2021), etc.), as there is limited potential for it to increase during a response.

Although the methods tested here were focussed on the surface locations (GPS) and dive data from harbour seals, the strong trends found across the simulations are likely to be applicable across systems (see Appendix E for additional work on beaked whales). Further simulations could be conducted for different species and/or biologging data types, in order to estimate the corresponding false positive and response detection rates. Further simulations could also investigate the effects of including different input metrics for Mahalanobis distance calculations. However, similar to model selection advice, it is recommended that studies continue to consider carefully which metrics may be biologically important or informative and that studies do not automatically include every metric that is available. If it is expected that a response may comprise a lack of a particular behaviour, studies should consider whether they can include a metric to represent this (e.g. a short dive may not be unusual, but an extended time resting at the surface might be). At present, Mahalanobis distance methods calculate the multivariate distance between the means of each time window (Equation 2.1). Further developments could consider whether another

metric could be used instead (e.g. maximum or median value of the windows) that might be less sensitive to dilution from non-response data points; however, it is possible that this may also increase noise in the baseline data and reduce overall performance. One aspect not considered in this simulation study was the accuracy of the method in identifying the precise timing of the response; this could be explored in further work, as it is of importance when attempting to relate behavioural changes to a known level of stimulus.

2.5.2 Simulating seal biologging data

Simulated data were used in order to quantify the performance of Mahalanobis distance methods in this study. This approach was beneficial as it enabled complete knowledge of the presence and timing of behavioural responses in the data. It also provided the flexibility to modify the strengths and types of responses present, ensure consistency between iterations, and generate a large sample of data on which to perform the tests. Whilst it is not possible for simulations to capture the full complexities of animal behaviour and decision-making, the data generated here were informative for answering the research questions considered in this study.

The horizontal simulator developed in this study was a multi-state, discrete-time, correlated random walk, with bias towards simulated attraction points and away from a repulsion point. Here, the simulation was constructed with two behavioural states: (1) a travel state characterised by high step lengths and low heading variability, and (2) an encamped state characterised by small step lengths and high heading variability. The simulated encamped state encapsulates both resting and foraging (area restricted search) behaviours. Whilst these behaviours serve at least two different biological functions, they can be challenging to differentiate in pinniped horizontal tracking data (Carter et al., 2016; McClintock et al., 2013). Previous studies of seal movement have included additional measurements of dive behaviour to distinguish resting from foraging (Carter et al., 2020a; McClintock et al., 2013; Russell et al., 2015). In this chapter, the aim was to estimate the average detection rates across all behaviours. In future, extending the simulator to contain three (or more) distinct behavioural states could allow investigation into whether response detection rates differ by behavioural context. Switching between behavioural states was assumed to be a first order Markov process, where the probability of being in a behaviour was determined by the behaviour in the previous time step. Incorporating a longer term memory within the simulation, such as the behaviour over the last five time steps, or including a count of the amount of time spent in a particular state, could further improve the quality of the

simulated data. For example, Langrock et al. (2014) modelled different states of beaked whale dives as a hidden semi-Markov model, where the dwell time within a particular state is taken into account for behaviour switching. For central-place foragers such as seals, which invest energy in travelling to and from foraging patches, including a state dwell time once individuals have switched to foraging would be a natural development of the simulator. Here, I simulated movement between two attraction points. These attraction points had a dual function as potential foraging or resting locations, due to the combined simulated encamped state discussed earlier. Future work could explore the inclusion of multiple attraction points, with different attraction strengths and/or functions, to explore if and how the motivation to reach an attraction point may impact the probability of detecting a response.

The dive simulator developed here generated multivariate state-dependent dive measurements, with a dive type state sequence simulated as a first order Markov process. By this formulation, each dive metric is independent of the metrics of the previous dive (conditional on behavioural state). In reality, it is likely that dive metrics (e.g. depth, duration) may exhibit subtle patterns over time, within behavioural state, but this is currently poorly understood. Further work would be required to identify and quantify this in the real data, and then the simulator could be extended to include correlation between dive metrics of successive dives to incorporate this. Similar to the horizontal simulator, another possible extension is to incorporate state dwell times (e.g. Langrock et al. (2014)) for the different dive types. Given the observed variability in seal movement behaviour between regions (Hastings et al., 2004), seasons (Crawford et al., 2019), ages (Carter et al., 2020a), sexes (Wilson et al., 2014), and individuals, expanding and parameterising dive and horizontal simulations to explore these differences are possible extensions of this work. Both of the simulators developed here could also be used to simulate data to address different research questions, for example to test other statistical methods.

The behavioural responses simulated here were chosen to be representative of a range of possible reactions by seals to disturbance. Horizontal flee responses to noise disturbance are often reported in cetaceans, but for pinnipeds there is a lack of individual tracking data on responses. Many studies have reported changes in the distribution of pinnipeds as a result of noise disturbance (e.g. Harris et al., 2001, Hastie et al., 2015, Russell et al., 2016), which could be as a result of an initial fleeing reaction. Dive responses of pinnipeds to sound have been reported in captivity (e.g. Kastelein et al., 2018a, Kvadsheim et al., 2010), but how individuals may change their dive behaviour in the wild is relatively unknown. Collecting and analysing tracking data on how wild pinnipeds move and react to a variety of sound

sources remains a key area for future research.

2.5.3 Conclusions

This chapter developed a simulation study to quantify the performance of a statistical method used for biologging data in behavioural response studies: Mahalanobis distance. It was found that the amount of baseline data available was an important factor in controlling false positives, and so it is recommended that future studies make efforts to maximise the duration of data collected before disturbance events. The simulations showed that using the covariance matrix across only the baseline data improved detection rates, whilst false positives remained relatively stable. Future users of this method should consider using this modified approach. There were also marked differences in detection rates between different window structures, and so it is recommended that future studies choose a window structure that is appropriate for the duration and type of response they are attempting to detect. Overall, there were consistent patterns in method performance across the different simulated scenarios, suggesting that this general advice is applicable across species and data types.

Chapter 3

Horizontal responses of harbour seals to pile driving sounds

The acoustic propagation models (described in Section 3.3.4) were produced by Bas Binnerts and Sander von Benda-Beckmann at TNO (Acoustics and Sonar Expertise Group, TNO, The Hague, Netherlands), in consultation with myself, Gordon Hastie and Carol Sparling. The rest of the analysis described in this chapter is my own work.

3.1 Abstract

Transitioning to renewable energy is essential to mitigate climate change; however, it is important to consider the potential impact of new technologies on wildlife. For animals that spend extended time periods underwater, such as marine mammals, a key environmental concern are the high intensity sounds produced during pile driving for offshore wind farm construction. While it is known that reductions in harbour seal density occur near to active pile driving (Russell et al., 2016; Whyte et al., 2020), how individual seals behave and move during exposure to these sounds is currently unclear. In this chapter, I used GPS tracking data from 24 tagged harbour seals (*Phoca vitulina*) to investigate how individual seals behaved during pile driving sound exposure. To do this, I calculated biologically informative metrics summarising the seals' movements, and used Mahalanobis distance randomisation tests to identify statistically unusual instances of movement behaviour during pile driving. Predictions from an acoustic propagation model were used to understand the relationship between the probability of these behavioural changes and the estimated received sound level from piling, using a Bayesian hierarchical model. Using estimates of seal density across the study area, the total number of seals likely to be affected was also estimated. In total, there were 216 encounters between seals and pile driving activity, and 15 unusual movement responses were detected. The detected changes in horizontal behaviour consisted of either (1) unusually high movement speeds, (2) cessation of horizontal movement, or (3) suddenly initiating travel after a long stationary period. The mean population-level response threshold at which these changes were estimated to occur was 186 (95% CI: 169–199) dB re $1\mu\text{Pa}^2 \cdot \text{s}$, with a standard deviation between individuals of 14 dB re $1\mu\text{Pa}^2 \cdot \text{s}$ and between encounters of 24 dB re $1\mu\text{Pa}^2 \cdot \text{s}$. The estimated sound level at which there was a 50% probability of response, p_{50}, was 175 (95% CI: 166–181) dB re $1\mu\text{Pa}^2 \cdot \text{s}$. Across the study area, this was equivalent to an estimated 311 seals responding per pile driving bout. The dose-response relationships estimated in this chapter provide an important tool for estimating the probability of these behavioural changes for offshore developments using pile driving. Overall, the individual behavioural changes observed in this chapter provide important insights into how seals may react to *in-situ* sound disturbance events, enabling improved biological and quantitative understanding of the differences between individuals and encounters, and ultimately contributing towards important knowledge gaps in predicting the population-level consequences of disturbance.

3.2 Introduction

The rapid expansion of the offshore renewable energy sector brings increased levels of human development into the marine environment (Bailey et al., 2014). Understanding the potential effects of this activity on marine wildlife, and being able robustly to predict impacts at the individual and population level, are critical steps in ensuring that new renewables developments are built and operated in an environmentally-responsible manner. In the last 30 years, 25 GW of offshore wind capacity (5,402 turbines) have been constructed in European waters; a further 29 GW of offshore wind capacity is predicted to be installed over the next five years (WindEurope, 2021). This expansion is predicted to continue as countries convert to renewable energy sources in order to meet global climate change targets.

For animals that spend extended time periods underwater, such as marine mammals, one of the main environmental concerns of offshore wind farm developments are the high levels of sound produced during the construction phase. Pile driving, a common method used in installation of the wind turbine foundations, has been reported to produce source levels of up to 250 dB re 1μPa @ 1 m (peak-peak) (Bailey et al., 2010). Installation typically consists of repeated hammering ($\sim$ every 1–2 seconds) of foundations into the seafloor, and can last for several hours at a time (Bailey et al., 2010; Graham et al., 2019). The potential effects of this activity on marine mammals include hearing damage from exposure to sound from pile driving (Hastie et al., 2015; Schaffeld et al., 2020; Whyte et al., 2020), exclusion of individuals from important areas of their natural habitat, and disruption to natural behaviour and distribution (e.g. Brandt et al., 2011; Graham et al., 2019; Whyte et al., 2020).

The at-sea distribution of seals, and the requirement for individuals to regularly return to haul-out sites, can lead to high spatial and temporal overlap between seals and areas of human activity in the coastal environment (Jones et al., 2017; Russell et al., 2014; Sharples et al., 2012). Visual observations of ringed seals (*Phoca hispida*) present around construction and pipe driving at an oil production site appeared to show little or no reaction to construction activity (Blackwell et al., 2004), with individuals occasionally swimming or sitting on ice within 2 km of the construction site. Culloch et al. (2016) also found no evidence of construction-related activities (seismic surveys, dredging, rock breaking, etc.) for an offshore gas pipeline affecting the occurrence of grey seals (*Halichoerus grypus*) from shore-based marine mammal observations. Visual surveys of seals (ringed, bearded (*Erignathus barbatus*), and spotted (*Phoca largha*)) found similar sightings rates between

different levels of seismic activity; however, the mean radial distance of seal sightings was larger during "full-array" (8–11 airguns) seismic activity compared to no activity (234 m vs. 144 m) (Harris et al., 2001). Both Edrén et al. (2010) and Skeate et al. (2012) investigated patterns in the numbers of grey and harbour seals hauled out near pile driving construction, for offshore wind farms in Denmark and the UK respectively. Aerial surveys, visual observations, and time-lapse photography showed that there was a 20–60% decline in the numbers of seals (both species were considered together) hauled out on the sandbanks during pile driving at the Danish Nysted wind farm; it was highlighted that the sound produced from the pile driving may deter seals from returning to the haulout site (Edrén et al., 2010). Aerial surveys of seals near the Scroby Sands wind farm in the UK detected a significant decline in the number of hauled out harbour seals during the year of construction; in contrast, the number of hauled out grey seals increased (Skeate et al., 2012). In general, studies of the at-sea behaviour and distribution of seals during exposure to sounds from human activity have produced mixed results; however, many studies have relied on visual observations alone, making it difficult to follow how individuals responded (Edrén et al., 2010).

For pile driving sounds in particular, tracking data from tagged harbour seals have been used to estimate seal density during times of piling and non-piling during the construction of an offshore wind farm (Russell et al., 2016; Whyte et al., 2020). It was estimated that seal density significantly reduced during piling up to distances of 25 km, or above single-strike sound exposure levels (SELss) of 145 dB re $1\mu\text{Pa}^2 \cdot \text{s}$ (Whyte et al., 2020). The spatial modelling approach used in these studies necessarily pooled across all piling events, to make a binary comparison between baseline vs. all piling events. Therefore, all predicted changes in seal density were made relative to the centre of the wind farm, and not to the respective piling locations. Additionally, sound propagation models were averaged to produce an single estimated soundscape across all 27 piling locations. Importantly, at this offshore wind farm in particular, the maximum distance between turbines was $\sim$13 km, and so there was considerable uncertainty in the true relationship between seal behaviour and the corresponding sound levels from pile driving activity (Whyte et al., 2020). There is a need to further understand how individuals may respond and change their movement behaviour during pile driving, to facilitate robust estimation of population-level effects and quantify variation in responses between and within individuals.

3.2.1 Aims

To improve understanding of individual behavioural impacts, and the mechanisms by which these may accumulate to population-level effects, in this chapter I use biologging data to examine how individual seals respond to *in-situ* anthropogenic sound exposure. Using the same data as described above (Hastie et al., 2015; Russell et al., 2016; Whyte et al., 2020) on 24 GPS-tagged harbour seals, I aim to investigate the effect of pile driving activity on the horizontal movement behaviour of individuals through four objectives. First, to objectively identify potential responses, unusual events of horizontal movement behaviour will be identified by Mahalanobis distance randomisation tests, which compare behaviour observed during pile driving to that observed during baseline periods. Second, this chapter will quantify the relationship between received pile driving sound level and the probability of response, using a Bayesian hierarchical model to combine the identified behavioural responses with estimated sounds levels from an acoustic propagation model. Third, variation in response thresholds between, and within, individuals will be estimated, and a suite of potential covariates will be included in the model to explore their effect on the dose-response relationship. Fourth, using the best fitting model of the dose-response relationship and estimates of seal density across the study area, the effective response range and total number of seals predicted to respond will be estimated to provide further context to the obtained results.

3.3 Methods

3.3.1 Data collection

Seal tag data

To record the movement behaviour of seals around pile driving activity, tags were deployed on harbour seals in the intertidal sandbanks of The Wash, south-east England, UK. In January 2012, 25 harbour seals were caught and fitted with a SMRU Instrumentation GPS telemetry tag (hereafter GPS/GSM tag; SMRU Instrumentation, University of St Andrews, Fife, UK). Seals were first anesthetized using Zoletil® or Ketaset® in combination with Hypnovel®, and GPS/GSM tags were attached to the fur at the back of the neck using a fast-setting two-part epoxy adhesive or Loctite® 422 Instant Adhesive. All seal handling and procedures were carried out under Home Office Licence 60/4009. The fitted GPS/GSM

tags can only attempt to measure the seals' locations when they are at the surface and the tag is sufficiently above the water. Fastloc GPS allows quick (∼100 ms; Carter et al., 2016) recording of the information required to estimate the location, which is later transmitted via the GSM phone network.

Out of the 25 deployed tags in The Wash, three tags collected data for less than two days and so were excluded from further analyses. Two seals from a concurrent study approximately 200 km to the south (in the Thames) moved into The Wash during pile driving, and so were included in the dataset. This resulted in a total sample size of 24 individuals (11 males, 13 females; Table 3.1). The data were cleaned and erroneous locations removed based on thresholds of residual error and the number of satellites. For more details of the data collection and study site, see Hastie et al. (2015) and Russell et al. (2016). Excluding haulout periods, the median time interval between received GPS locations was 10 minutes. For ∼86% of GPS locations, the time intervals between successive locations were less than 20 minutes; for ∼94% of locations, time intervals were less than 30 minutes.

Pile driving operations

Operational data on pile driving at Lincs offshore wind farm were provided by Centrica plc. Throughout the period of the seal tag deployment, 27 (of a total of 75) monopiles were installed at Lincs by pile driving (Figure 3.2(b)). Between 28^{th} January and 11^{th} May 2012, a total of 77,968 piling strikes occurred, with a mean strike energy of 1,202 kJ (SD= 613) (Figure 3.2(a)).

Pile driving at Lincs occurred intermittently across the study period, with a maximum gap of 19.5 days with no piling activity. To compare time intervals of piling and non-piling, individual records of piling hammer blows were divided into bouts, where piling was considered as a separate bout if successive hammer blows were more than one hour apart. This resulted in 64 piling bouts across the study period and the 27 piling locations. Piling bouts had a mean duration of 1.0 hour (min = 0.2, max = 3.2), containing an average of 1,218 hammer blows each (min = 132, max = 3,772), and a median inter-strike interval (time between successive hammer blows) of two seconds.

Table 3.1 Details of the 24 tagged harbour seals. Deployment is the date that GPS/GSM tags were attached to each individual. The time period covered is the duration of days that data were available for each individual within the study period, excluding the first week of deployment. The baseline duration is the effective duration of that data available for use as baseline, incorporating data removed by chosen spatial constraints, pile driving times, and haulouts. Individuals with the prefix 'pv42-' were tagged in The Wash; individuals with the prefix 'pv40-' were tagged in the Thames.

Seal reference	**Deployment date**	**Time period covered** (days)	**Baseline data duration** (days)	**Sex**	**Age class**
pv42-162-12	23/01/12	110	39	F	Adult
pv42-165-12	21/01/12	57	10	F	Juvenile
pv42-194-12	23/01/12	107	20	M	Adult
pv42-198-12	24/01/12	110	32	M	Adult
pv42-220-12	24/01/12	110	34	M	Adult
pv42-221-12	24/01/12	44	12	M	Adult
pv42-266-12	24/01/12	78	32	F	Adult
pv42-277-12	23/01/12	110	43	F	Adult
pv42-287-12	24/01/12	12	7	M	Adult
pv42-288-12	21/01/12	110	48	F	Adult
pv42-289-12	25/01/12	74	14	M	Adult
pv42-290-12	25/01/12	53	17	F	Adult
pv42-291-12	23/01/12	102	37	F	Adult
pv42-292-12	24/01/12	99	38	M	Adult
pv42-293-12	25/01/12	65	13	F	Adult
pv42-294-12	25/01/12	97	32	M	Adult
pv42-295-12	25/01/12	64	26	F	Adult
pv42-316-12	22/01/12	98	36	M	Juvenile
pv42-317-12	23/01/12	105	37	F	Adult
pv42-318-12	23/01/12	110	38	F	Adult
pv42-319-12	22/01/12	106	43	M	Juvenile
pv42-320-12	21/01/12	96	38	F	Adult
pv40-268-12	18/01/12	114	33	F	Adult
pv40-270-12	18/01/12	86	11	M	Adult

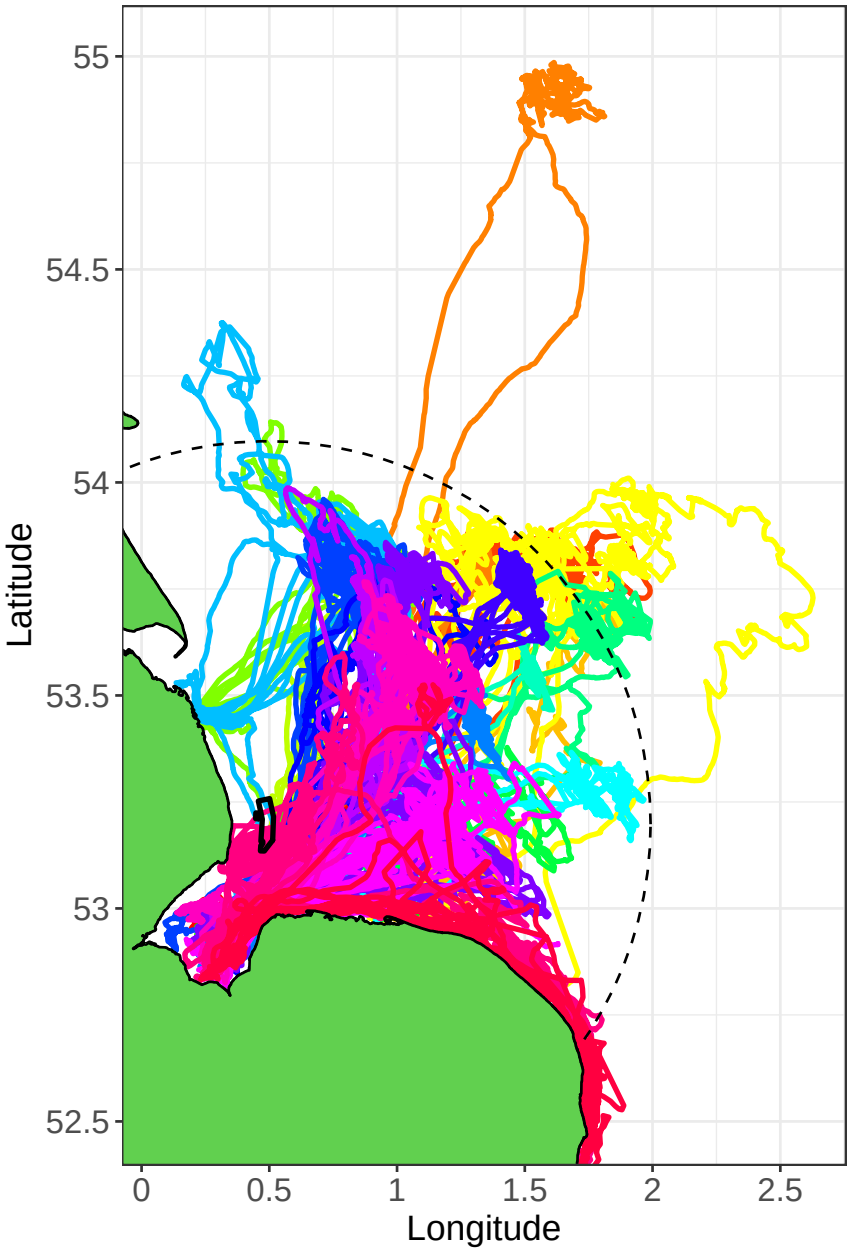


Figure 3.1 Overview of movement tracks collected around The Wash (UK) in 2012 from 24 harbour seals fitted with GPS/GSM tags (lines coloured by individual). Also shown are the outline of Lincs offshore wind farm (solid black line), and a line representing 100 km from the centre of the wind farm (dashed black line).

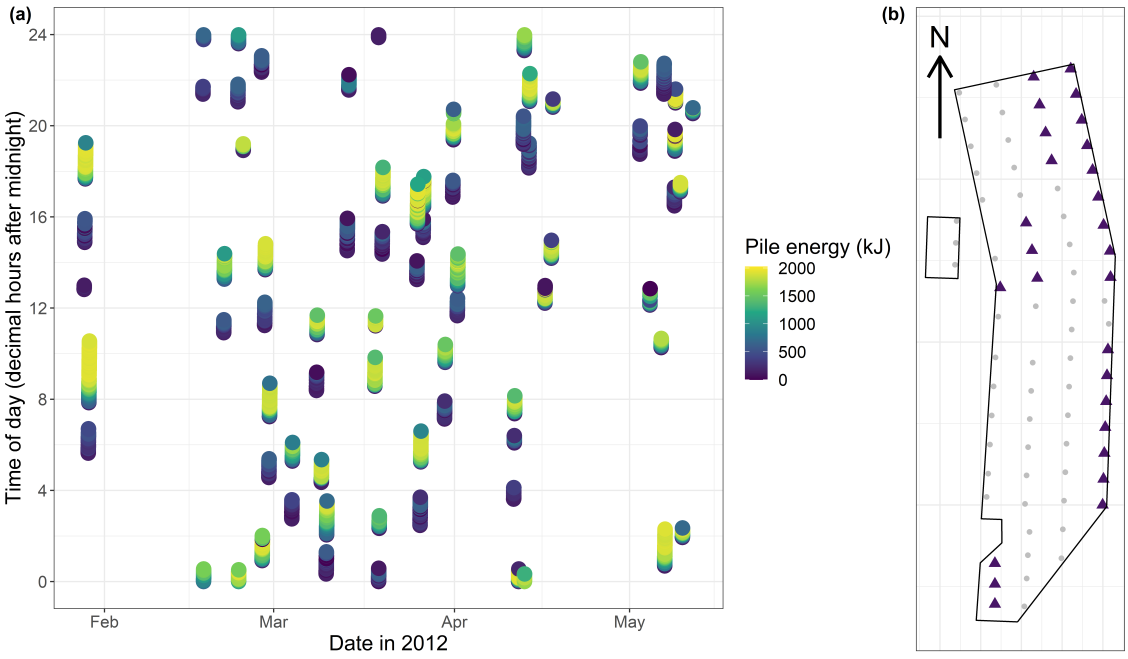


Figure 3.2 Pile driving activity at Lincs offshore wind farm in 2012. (a) The date, time of day, and piling hammer energy (kJ; colourscale) of each of the 77,968 piling strikes. (b) Map of Lincs wind farm with active piling locations (purple triangles) and locations of the previously installed wind turbine foundations on the site (grey circles). The wind farm site is ∼13 km by ∼3 km.

3.3.2 Quantifying movement behaviour

For each tagged individual, GPS locations recorded in the first week of deployment were removed to ensure that any potentially altered behaviour following the tagging procedure (sedation, disturbance) were not included in the analyses (McKnight, 2011). GPS locations recorded more than one week after the last pile driving bout were also removed. This approach enabled comparisons of seal behaviour during piling and non-piling periods to be limited to the same time of year (January–May), and in particular excluding data collected during the harbour seal breeding period (June–July) when movement behaviour is likely to be different (Thompson et al., 1994; Van Parijs et al., 1997).

Seal locations from the GPS/GSM tags were used to calculate several informative movement metrics, quantifying movement behaviour between successive GPS locations (e.g. between locations a_{xy} and b_{xy}):

- **Speed:** The average horizontal speed (displacement) since the previous recorded location (m s^{-1}):

$$v = \frac{D(a_{xy}, b_{xy})}{t_b - t_a}$$

  where v is the estimated speed, $D(a_{xy}, b_{xy})$ is the horizontal distance between the pair of locations, and t_a and t_b are the times of locations a_{xy} and b_{xy} respectively. It should be noted that this is an estimate of the minimum speed travelled at by the animal, as it assumes a straight line distance between GPS locations and the true path is unobserved.

- **Acceleration:** The absolute value of acceleration, i.e. the difference in speed since the previous recorded location (m s^{-2}):

$$a = \frac{|v_t - v_{t-1}|}{\Delta t}$$

  where a is the estimated absolute acceleration, v_t is the horizontal speed at time t, v_{t-1} is the horizontal speed at time $t-1$, and Δt is the difference in time between the two locations.

- **Heading variability:** The circular variance in heading (Mardia and Jupp, 1999), i.e. a measure of the difference in heading between successive movement steps:

$$c = 1 - \frac{\sqrt{(\cos(\psi_t) + \cos(\psi_{t-1}))^2 + (\sin(\psi_t) + \sin(\psi_{t-1}))^2}}{2}$$

where $\psi_t = \arctan(b_{xy}, a_{xy})$ is the heading angle of a GPS location at time t, and ψ_{t-1} is the heading angle at the previous GPS location $t-1$. The circular variance in heading c ranges between 0 and 1, where similar angles results in values near 0, and pairs of angles which are highly different results in values near 1.

- **Radial avoidance rate:** A measure of the rate at which the seal is moving away from (or towards) the centre of the wind farm:

$$\text{RAR} = \frac{\Delta r}{\Delta t \; v \; r}$$

where RAR is the estimated radial avoidance rate, r is the current radial distance between the centre of the wind farm and the seal's location (km), Δr is the change in the radial distance between the two GPS locations (m), Δt is the difference in time between the two locations (s), and v is the current speed (m s^{-1}). This metric measures the amount of the seal's movement which is directed directly away from the wind farm, accounting for the current speed of the animal. Positive values signify movement away from the wind farm, negative values signify movement towards the wind farm.

If the time between two observed GPS locations was greater than two hours, movement metrics were not calculated or used in further analysis. Similarly, metrics were not calculated for GPS locations which overlapped with haulout times recorded on the GPS/GSM tag.

3.3.3 Identifying a behavioural response

Specifying baseline data

To make comparisons of seal behaviour during piling and non-piling periods, it was necessary to specify a subset of the tracking data which could be considered as baseline. As pile driving activity took place intermittently over (and prior to) the study period, baseline data was also pooled from intermittent periods across the study period. For all individuals, movement data collected during the 6 hours prior to, and 24 hours following, any pile driving bout was excluded from the baseline. This was a conservative approach to ensure that any potential disturbance during the setup and operation of pile driving was not considered as baseline, as well as allowing time for recovery to normal behaviour after potential disturbance. From Russell et al. (2016), it was estimated that seal distribution

returned to normal within two hours of the cessation of pile driving (i.e. there was no detectable difference in seal density). For each individual, any sections of track which were more than 100 km away from the wind farm site (Figure 3.1) were also excluded from the baseline. Comparisons were therefore carried out for movement behaviour within the same spatial area and types of habitat.

Summarising movement metrics by Mahalanobis distance

Mahalanobis distance was used to summarise the chosen movement metrics into a single variable (DeRuiter et al., 2013) representing the overall difference in movement behaviour (see Chapter 2; Equation 2.1), using a custom-written R function `mdist_xy`. Following the findings of the simulation study for horizontal movement data described in Chapter 2, Mahalanobis distance implementation options were chosen that would best detect relatively short-term movement responses in the available data. For each individual, time windows of 20 minutes in duration were constructed across the timeline of available data, with windows sliding along the data every 10 minutes. Using Equation 2.1 (Chapter 2), the Mahalanobis distance was calculated between the mean value of movement metric observations in the baseline period (a fixed reference window approach), and the mean value of movement metric observations in each 20 minute time window. For each individual, the covariance matrix used in the Mahalanobis distance calculation was from the baseline period only, following the findings of Chapter 2. The movement metrics used were speed, absolute acceleration, heading variability, and radial avoidance rate, and all comparisons were made within individual only.

As seal GPS locations occurred irregularly in time, and due to the data removals mentioned above (e.g. haulouts), not all of the 20 minute windows contained data. When no data observations were present within a time window, no Mahalanobis distance calculation was possible and so the window slid on 10 minutes to the next position.

Identifying unusual behaviour by randomisation tests

To identify movement behaviour changes that were unusual, Mahalanobis distance values observed during pile driving were compared to those observed during baseline periods by randomisation tests. For each pile driving bout, each individual that was estimated to have a received level of piling sound above the ambient sound level (see next Section) was considered to have had an "encounter" with pile driving. For each encounter, responses

were investigated within the period of pile driving plus a 30 minute buffer afterwards. This buffer allowed for delayed responses to be identified; this was important as the irregular timing of the GPS locations and the time required for an individual to change its behaviour in a significant way meant that it was possible that an individual responding towards the end of the pile driving bout would only be detectable in the data after the bout had ended. For each encounter, 1000 randomly sampled "mock exposure periods" were taken from the baseline data for that individual, where each mock exposure is of the same duration as the piling bout. The maximum Mahalanobis distance $D_{\max}$ observed in each of the mock exposures was then compared to the Mahalanobis distance values observed in the true exposure encounter. The 95^{th} percentile of the $D_{\max}$ samples across the mock exposures was used as a threshold D_{th}, above which it is considered unlikely that behaviour changes of this magnitude will occur by chance. Mahalanobis distance values exceeding this threshold during a piling encounter are considered to be a response to the pile driving activity itself, as they occur with low probability (<5%) in the baseline data.

3.3.4 Estimating the received sound level

Acoustic modelling of pile driving sounds

To estimate the sound levels resulting from piling across the study area, the Aquarius pile driving model (for detailed description of the model and its validation see de Jong et al. (2019)) was used to model source characteristics and acoustic propagation loss. The Aquarius model uses information on the properties of the hammer and the pile to determine a source excitation spectrum, which is integrated into a range dependent propagation model to predict acoustic propagation loss across the study area, incorporating information on seabed characteristics and water depth. Here, the bathymetry was set to Mean Sea Level (MSL) and the seabed was assumed to be homogeneous, with properties corresponding to medium sand (Table 4.18 in Ainslie, 2010), the most common value in the considered modelling area (using data from the EMODnet Bathymetry Data Portal). Water depths relative to Mean Sea Level (MSL) were derived by applying the United Kingdom Hydrographic Office Vertical Offshore Reference Frame (VORF) correction for the study area (Iliffe et al., 2013), as the original bathymetry data were relative to chart datum at Lowest Astronomical Tide (LAT). Full details of the parameters used in the acoustic modelling are described in Whyte et al. (2020).

Depth specific model predictions were output as estimated single strike sound exposure

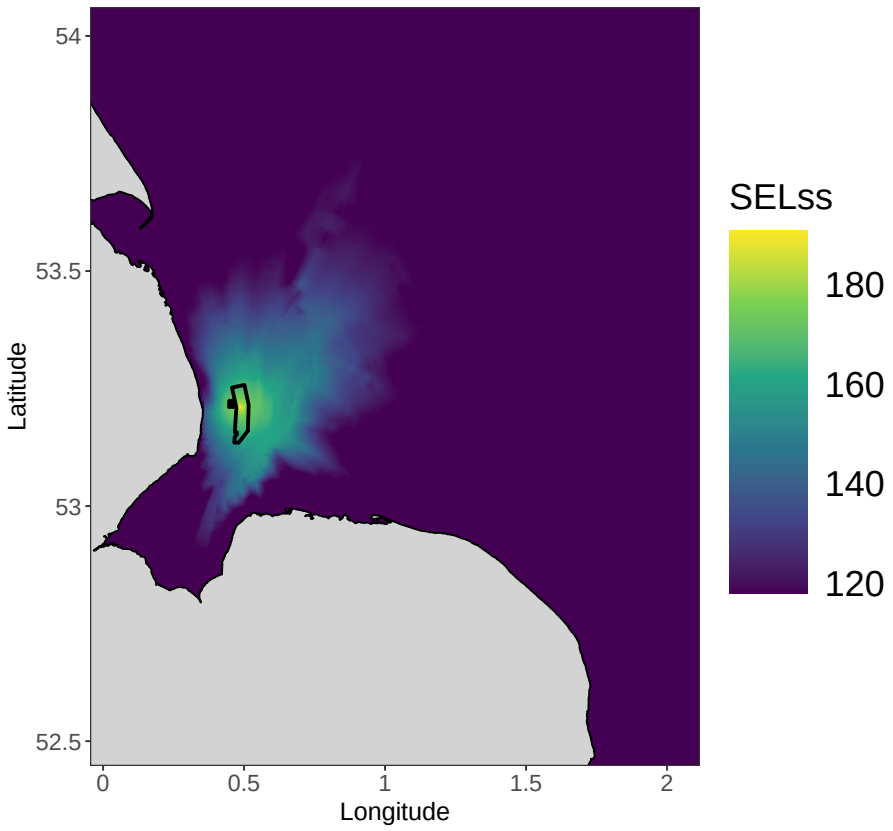


Figure 3.3 Predicted piling sound levels (single-strike sound exposure levels, SELss, dB re $1\mu Pa^2 \cdot s$) across the study area for one of the piling locations (pile identification LS22). Shown is the median SELss across all modelled depth bins. Predicted SELss below the estimated ambient noise level of 118 dB re $1\mu Pa^2 \cdot s$ were set to 118 dB re $1\mu Pa^2 \cdot s$.

levels ($SEL_{ss,\,ref}$, dB re $1\mu Pa^2 \cdot s$) at a reference piling strike energy of 1000 kJ. These were calculated across a series of spatial grids within the study area at ~279 m resolution (Longitude: from $-1°$ to $3°$ with a 15 s resolution, Latitude: from $52°$ to $55°$ with a 9 s resolution) for each of the 27 piling locations. Individual grids were produced for each 2.5 m depth bin (from 2.5 to 107.5 m depth); sound levels below the seabed were indicated by a "NaN" value. Frequencies from 16 Hz to 20 kHz were modelled, using third octave centre frequency bands.

Acoustic exposure of the tagged seals

The acoustic "dose" of received piling sound was estimated for all seal encounters, including both responses and non-responses (as identified by the randomisation test). For identified responses, the response interval (the time window within which a detected response first occurred) was considered to be from the start to the end of the first 20-minute window exceeding D_{th} (the Mahalanobis distance threshold used to determine a response). For encounters with a data gap prior to the time window exceeding D_{th}, the start of the response interval was extended back to the end of the previous time window (the last recorded window not exceeding D_{th}), to account for uncertainty in the time at which the response began. This accounted for sound levels that may have been received in between available GPS locations and contributed to a behavioural response.

For each piling strike that occurred within the response interval, seal tracks were linearly interpolated to estimate the location of the seal at the time of each piling strike. Each of these seal locations was matched to the corresponding spatial grid cell in the acoustic model (for the corresponding piling location), and the median received $\text{SEL}_{\text{ss, ref}}$ across all depth layers was identified. Information on the blow energy of each strike was then used to scale the modelled reference $\text{SEL}_{\text{ss, ref}}$ (at 1000 kJ strike energy) to obtain final estimates of the received SEL_{ss} at each seal location. This was carried out through energetic (broadband) scaling of the SEL_{ss} spectrum:

$$\text{SEL}_{\text{ss}} = \text{SEL}_{\text{ss, ref}} + 10 \log_{10} \frac{\text{E}}{\text{E}_{\text{ref}}} \tag{3.1}$$

where E is the energy (kJ) of the pile driving strike, E_{ref} is the reference strike energy (1000 kJ), $\text{SEL}_{\text{ss, ref}}$ is the modelled single strike exposure level at the reference strike energy, and SEL_{ss} is the resulting scaled single strike sound exposure level (dB re $1\mu\text{Pa}^2 \cdot \text{s}$). The estimated distance between the piling location and the seal's location for each piling strike was also calculated, as well as the received levels in the quietest and loudest depth bins. Measurements by Nedwell et al. (2011) of ambient noise in The Wash during construction of Lincs wind farm gave a median ambient sound level of 118 dB re $1\mu\text{Pa}^2 \cdot \text{s}$, and so any estimated SELss below this value were assigned to 118 dB re $1\mu\text{Pa}^2 \cdot \text{s}$.

For encounters in which seals were not identified to respond, seal tracks were linearly interpolated to estimate the location of the seal at the time of all piling strikes within the piling bout. Using the same approach as above, these piling strikes were matched to the acoustic models and scaled by Equation 3.1 to estimate the received acoustic dose for

encounters in which seals did not respond.

3.3.5 Modelling the dose-response relationship

A Bayesian hierarchical model of response thresholds

A Bayesian hierarchical model (based on that proposed by Miller et al., 2014) was used to estimate the relationship between the piling sound level (acoustic dose) and the likelihood of a behavioural response, using information from all piling encounters (both responses and non-responses). The model is structured around the concept that all seals will have a sound level at which they will respond in an encounter with pile driving. Seals that respond have reached their response threshold, whereas seals that do not respond have simply not exceeded their response threshold for that encounter. The response threshold can vary between and within individuals, and depend on individual-level and encounter-level covariates.

Let μ be the mean response threshold (in dB re $1\mu\text{Pa}^2\cdot\text{s}$) for the population, where response thresholds across individuals in the population are assumed to be normally distributed. Each individual i is assumed to have a mean response threshold μ_i,

$$\mu_i \sim \text{N}(\mu + \boldsymbol{z_i}^\intercal \boldsymbol{\alpha}, \phi^2) \tag{3.2}$$

where $\boldsymbol{z}_i$ is a vector of individual-level covariates (which can be continuous covariates or factor levels), $\boldsymbol{\alpha}$ is a vector of model parameters, and ϕ^2 is the variance between individuals. Individual-level covariates (e.g. sex) can be included to account for factors that may alter the response threshold between different groups of individuals. Within each individual, response thresholds can vary between each encounter j. The true (but unobserved) response threshold $t_{i,j}$ for a given individual i and encounter j, is assumed to be normally distributed,

$$t_{i,j} \sim \text{N}(\mu_i + \boldsymbol{z_{i,j}}^\intercal \boldsymbol{\beta}, \sigma^2) \tag{3.3}$$

where $\boldsymbol{z_{i,j}}$ is a vector of encounter-level covariates, $\boldsymbol{\beta}$ is a vector of model parameters, and σ^2 is the variance between encounters. Encounter-level covariates (e.g. time of day) can be included to account for factors that may alter the response threshold between encounters. When no covariates are included in the model, Equations 3.2 and 3.3 reduce to $\mu_i \sim \text{N}(\mu, \phi^2)$ and $t_{i,j} \sim \text{N}(\mu_i, \sigma^2)$ respectively.

To account for uncertainty in the estimated acoustic dose eliciting the response, it was assumed that the true response threshold $t_{i,j}$ is observed with error:

$$y_{i,j} \sim \mathrm{N}(t_{i,j}, \epsilon^2) \tag{3.4}$$

For encounters in which the individual was detected to respond, $y_{i,j}$ and ϵ are the estimated median and standard deviation respectively of all received SELss from piling blows within the initial response interval. It should be noted that this measure of uncertainty in the observed dose only accounts for uncertainty in which piling blow contributed to the response, and not uncertainty in estimated SELss from the acoustic modelling stage (Whyte et al., 2020). For encounters in which the individual was not detected to respond, there is no observation of the response threshold and so these encounters are considered to be right-censored. The maximum of the received SELss across the piling bout is taken to be the observed dose $y^c_{i,j}$, and the true response threshold $t_{i,j}$, which has not yet been reached, is assumed to be above this value, such that $t_{i,j} > y^c_{i,j}$.

Prior distributions for model parameters were set so as to be generally uninformative, but aimed to ensure that the model remained within biologically reasonable limits. A uniform prior with a lower limit of 115 dB re $1\mu\mathrm{Pa}^2 \cdot \mathrm{s}$ and an upper limit of 200 dB re $1\mu\mathrm{Pa}^2 \cdot \mathrm{s}$ was used for the mean population response threshold μ. It was assumed that no individuals would respond to sound below the lower limit (as it is below the median level of ambient noise), and that all individuals would respond by the upper limit of this interval (the approximate maximum received level at 0 m away from pile driving). To accommodate this within the model, $t_{i,j}$ was assumed to follow a truncated normal distribution (Antunes et al., 2014; Miller et al., 2014), with a lower limit of 115 dB re $1\mu\mathrm{Pa}^2 \cdot \mathrm{s}$ and an upper limit of 200 dB re $1\mu\mathrm{Pa}^2 \cdot \mathrm{s}$. Priors for the standard deviation between individuals ϕ and between encounters σ were set to be uniform between 0 and 30 dB re $1\mu\mathrm{Pa}^2 \cdot \mathrm{s}$. Standard deviations of 30 dB would be equivalent to a mean response threshold $\pm$60 dB; response thresholds are unlikely to differ more than this between or within individuals.

Models were fit in JAGS (version 4.3.0; Plummer (2003)) via the R package `rjags` (Plummer, 2019). JAGS uses Markov Chain Monte Carlo (MCMC) to estimate values for the parameters in the model and produce a posterior distribution for each parameter. Here, model parameters were estimated from 400,000 iterations, across 4 chains of 100,000 iterations each. A conservative burn-in period of 10,000 iterations was used and discarded for each chain. Starting values were set at the midpoint of each prior range ($\mu = 150$, $\sigma = 15$, $\phi = 15$). A range of starting values across the spread of the priors were tested to

Table 3.2 Covariates considered in the dose-response model of harbour seal responses to pile driving. Covariates were either of type factor (F, with number of levels in brackets) or continuous (C), and applied at either the individual (Ind) or encounter (Enc) level of the model.

Name	Type	Level	Description
Sex	F (2)	Ind	Sex of seal (male; female)
Age class	F (2)	Ind	Age class of seal (adult; juvenile)
Daylight	F (2)	Enc	Time of day (day; night)
Exposure history	F (3)	Enc	Number of previous encounters ($\leq$4; 5–14; $\geq$15)
Behavioural state	F (3)	Enc	Estimated state pre-piling (travel; forage; rest)
Bout length	C	Enc	Total length of piling bout (hours)
Initial distance	C	Enc	Initial distance at start of piling encounter (km)

ensure that the final posterior distributions were not sensitive to the starting values chosen. Chains were assessed for convergence by visual inspection of the chains and checking that the Gelman and Rubin convergence diagnostic was <1.1 (Gelman and Rubin, 1992).

Including covariates on estimated response threshold

A suite of potential covariates were considered in the dose-response model (Table 3.2). Individual-level covariates (sex, age class) were obtained during the seal tagging procedure, whereas exposure-level covariates (e.g. exposure history, bout length) were estimated or measured for each piling encounter. The R package `suncalc` (Thieurmel and Elmarhraoui, 2019) was used to compare piling times to local sunrise and sunset times. Piling bouts that began before sunrise or after sunset were categorised as night; all other piling bouts were considered to take place during the day. For each individual, the order and number of piling encounters were calculated. For each encounter, the number of previous known encounters an individual had with pile driving (within the study period) was used a measure of exposure history; this was divided into a low ($\leq$4), medium (5–14), and high ($\geq$15) number of previous encounters. Russell (2016) used movement and activity data within a state-space modelling framework to infer behavioural states (interpreted as travelling, foraging, or resting), on a 2 hour resolution, of tagged harbour seals (including the individuals in this study). Estimated state in the previous 2-hour interval before each pile driving bout was used as an estimate of the behavioural state of the individual before each encounter. The total duration of the piling bout and the initial distance of the individual from pile driving were also included as potential covariates that may influence response threshold.

A maximal model with all covariates and a simple model with no covariates were both fitted. In the maximal model, covariates were included at both the individual (Equation 3.5) and encounter level (Equation 3.6), by expansion of Equations 3.2 and 3.3 respectively:

$$\mu_i \sim \mathrm{N}(\mu + \alpha_{\mathrm{sex}}\mathrm{S}_i + \alpha_{\mathrm{age}}\mathrm{A}_i, \phi^2) \tag{3.5}$$

$$t_{i,j} \sim \mathrm{N}(\mu_i + \beta_{\mathrm{day}}\mathrm{D}_{i,j} + \boldsymbol{\beta}_{\mathrm{exp}}\mathrm{E}_{i,j} + \boldsymbol{\beta}_{\mathrm{behav}}\mathrm{B}_{i,j} + \beta_{\mathrm{lengthbout}}\mathrm{L}_{i,j} + \beta_{\mathrm{initdist}}\mathrm{R}_{i,j}, \sigma^2) \tag{3.6}$$

Here, each α term is a model parameter for each individual-level covariate, and S_i and A_i are, respectively, the sex and age class of each individual i. Each β term is a model parameter for each encounter-level covariate, and $\mathrm{D}_{i,j}$, $\mathrm{E}_{i,j}$, $\mathrm{B}_{i,j}$, $\mathrm{L}_{i,j}$ and $\mathrm{R}_{i,j}$ are, respectively, the daylight, exposure history, behavioural state, piling bout length and initial distance for each encounter j (Table 3.2). The parameters $\boldsymbol{\beta}_{\mathrm{exp}}$ and $\boldsymbol{\beta}_{\mathrm{behav}}$ are vectors as both exposure history and behavioural state were factor covariates with three levels, requiring two β parameters each to estimate.

Gibbs variable selection (O'Hara and Sillanpää, 2009) was used to assess the level of support for including each of these covariates within the final dose-response model. This was implemented by using an indicator parameter I for each covariate, e.g. the sex covariate was $I_{\mathrm{sex}}\alpha_{\mathrm{sex}}S_i$. The indicator parameter I was binary, and took value 1 when the parameter was included in the model, and 0 when it was not included. This enabled the model to include and exclude different covariates throughout the iterations. The proportion of MCMC samples in which the term was included is an estimate of the posterior probability that this term is present in the true model. Models without indicator variables were also fit to examine what effect each α and β term might have, assuming all covariates were included. Prior distributions for each of the indicator term I were Bernoulli with a mean of 0.5, indicating the initial uncertainty on whether the covariate should be included or excluded. Prior distributions for each of the α and β terms were normal, with a mean of zero and a standard deviation of 10 dB re $1\mu\mathrm{Pa}^2 \cdot \mathrm{s}$.

Estimating the dose-response function

The posterior samples from the Bayesian hierarchical model were used to estimate a dose-response function. The estimates of μ, ϕ and σ for each posterior sample were used to construct a cumulative distribution function in 1 dB increments. The mean dose-response function was estimated by taking the mean values of these cumulative distribution functions

across all posterior samples, and 95% credible intervals were estimated by taking the 2.5% and 97.5% quantiles across all posterior samples for each dB increment. Due to truncation of the probability distribution, the dose-response function may not be symmetric and so the dose at which the probability of response is 0.5, p_{50}, was also calculated. For the maximal model containing covariates, the posterior samples were used to estimate a series of potential dose-response functions to illustrate the effect of each covariate. To do this, posterior samples of the α and β term(s) for each covariate, as well as μ, ϕ and σ were used to calculate the mean and 95% credible interval of the dose-response function, as above. This approach assumes that all other covariates are present in the maximal model.

3.3.6 Estimating an effective response range (ERR) and the total number of seals affected

Due to the spreading of sound away from a source, the area that has moderate or low levels of piling sound is much larger (and therefore contains more individuals) than the area with the highest levels of sound. Using the p_{50} estimate alone can lead to underestimation of the numbers of animals affected (Tyack and Thomas, 2019), and in particular can ignore the most sensitive individuals which will respond at the lower thresholds further from the source. To increase the applicability of these results to management where a single number is preferred, and to estimate the total number of seals likely to respond, an effective response range (ERR) was estimated using the dose-response curve. The effective response range is the distance at which the number of animals responding beyond this distance is equal to the number not responding within this distance (Tyack and Thomas, 2019). Thus, it can be a relatively straightforward approach to estimate the total number of animals affected across the study area. Due to the spread of piling locations (the wind farm is ∼13 km in length and ∼3 km in width) and variation in bathymetry of the habitat surrounding each piling location, each piling location was considered separately.

For each of the 27 piling locations, a spatial grid with 250 m by 250 m resolution was constructed across the study area. At each grid cell, the predicted received SELss from piling (median across all depths) was extracted from the acoustic model, at the maximum strike energy of 2,000 kJ (Equation 3.1). Grid cells that were on land were removed. By comparing the SELss at each each grid cell g to the mean dose-response function, the probability of response p_g at each grid cell was calculated. Theoretically, if each grid cell contained n individuals then, on average, the number of individuals in that grid cell that would respond would be np_g. The effective response range (ERR) is the distance at which

the number of individuals that are predicted to respond in grid cells beyond the ERR is equal to those that failed to respond in grid cells within the ERR distance:

$$\sum_{r_g<\text{ERR}} (1 - p_g) = \sum_{r_g>\text{ERR}} p_g \tag{3.7}$$

where r_g is the distance between each grid cell g and the piling location. If a uniform animal density is assumed, then animal density is not required to estimate the ERR as it would cancel out on both sides of the equation. Here, the ERR was determined by increasing the radius around the piling location in increments of 250 m until the sum of the probabilities of non-response for grid cells within the radius were equal to the sum of the probabilities of response for grid cells outside the radius. The median estimated ERR across the 27 piling locations was used to determine the overall ERR.

The overall ERR was compared to estimates of harbour seal density to illustrate the total number of seals potentially affected in the study area. Carter et al. (2020b) produced predictions of the proportion of the harbour seal at-sea population in 5 x 5 km grid cells across the UK and Ireland. Here, as an example, the mean predicted proportional density of the population across the four nearest 5 x 5 km grid cells around Lincs wind farm was calculated using the predictions from Carter et al. (2020b). This was multiplied by the estimated total at-sea population (42,800; Carter et al. (2020b)) to estimate the number of seals in this grid. This was then converted to the average density D per km^2. The total number of seals affected N was then estimated by $N = D\ \pi\ \text{ERR}^2$.

3.3.7 Software used

All analyses were conducted in R (R Core Team, 2020), with additional packages `dplyr` (Wickham et al., 2020), `lubridate` (Grolemund and Wickham, 2011), `maptools` (Bivand and Lewin-Koh, 2021), `raster` (Hijmans, 2020), `rgdal` (Bivand et al., 2021), `rjags` (Plummer, 2019), `sp` (Pebesma and Bivand, 2005), `StatMatch` (D'Orazio, 2019), and `suncalc` (Thieurmel and Elmarhraoui, 2019). The R packages `ggplot2` (Wickham, 2016), `ggpubr` (Kassambara, 2020) and `viridis` (Garnier, 2018) were also used for data visualisation.

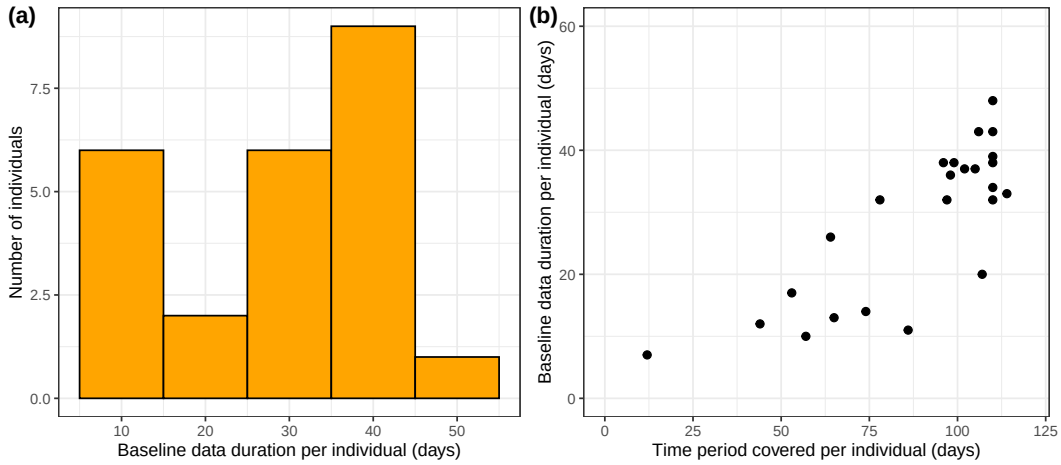


Figure 3.4 (a) The baseline data available for each tagged individual, and (b) baseline data duration against the full time period covered by each tag. Note that both measurements exclude the first week of deployment and any data collected after the end of the study period (18^{th} May).

3.4 Results

3.4.1 Seal responses to pile driving sounds

Across all 24 tagged seals, the median baseline duration for each individual was 33 days (min = 7, max = 48; Figure 3.4(a); Table 3.1). In general, $\sim$ 35% of the full data duration was at-sea behaviour which could be used as baseline (Figure 3.4(b)). In total, 20 out of 24 (83%) of the tagged seals had encounters with pile driving, i.e. they were estimated to have received SELss from piling above a median ambient sound level during a piling bout. The mean number of encounters per seal was 11 (min=1, max=36), with the initial distance between seals and pile driving during these encounters ranging from 4.2–45.1 km.

Out of a total of 216 encounters between seals and pile driving activity, 15 responses were identified by the Mahalanobis distance randomisation test (Table 3.3). Nine seals responded to pile driving at least once during the study (Table 3.3). The initial distance of seals to pile driving in these identified responses ranged from 5.0 to 28.6 km, and the maximum values of Mahalanobis distance calculated appeared to decline with distance (Figure 3.5). Estimated median received SELss (median across depths) at the identified time of response ranged from 119 to 169 dB re $1\mu\text{Pa}^2 \cdot \text{s}$; however, had the individuals always been present in the quietest or loudest parts of the water column, the range of maximum received SELss across all response intervals may have been 118–162, or 121–172 dB re $1\mu\text{Pa}^2 \cdot \text{s}$ respectively (Table 3.3).

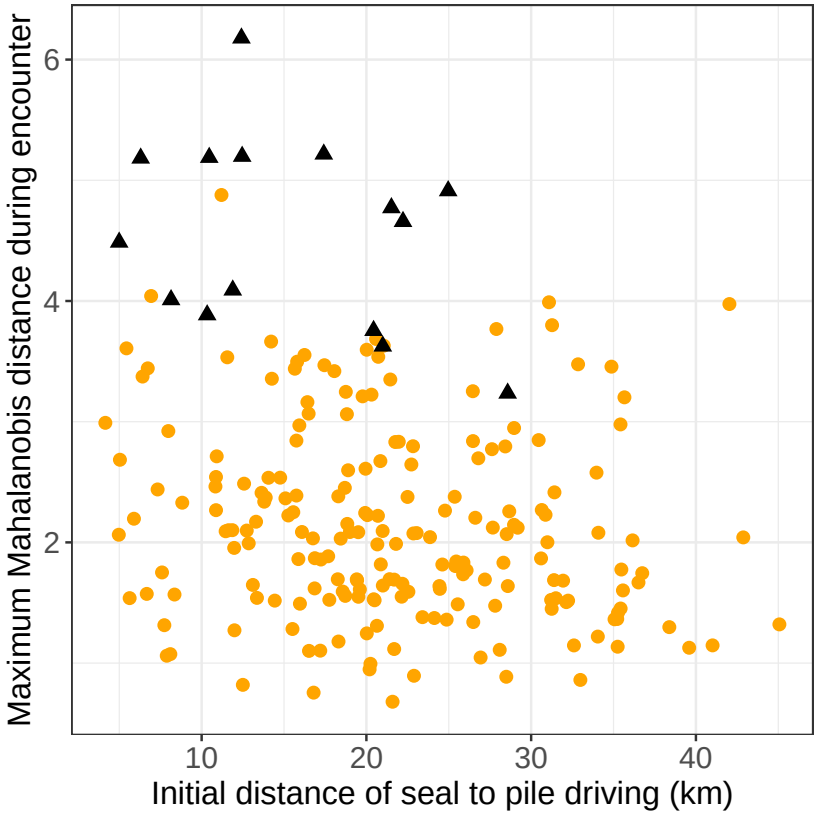


Figure 3.5 Maximum Mahalanobis distance during all encounters between tagged seals and pile driving. By randomisation test, each encounter was identified as either a detected response (black triangles) or a non-response (orange circles).

Table 3.3 Summary of the identified horizontal responses by Mahalanobis distance. For each identified response time interval, the median SELss across all depths was calculated for each received piling hammer blow, and the median $y_{i,j}$ and standard deviation ϵ of these piling blows was used in the dose-response model (Equation 3.4). Shown here also are the initial distance between the seal and pile driving, the estimated maximum SELss the individuals may have received had they always been present in the quietest or loudest parts of the water column, and the percentile of the Mahalanobis distance values corresponding to each response. All measurements of SELss are in dB re $1\mu\text{Pa}^2 \cdot \text{s}$.

Seal ID	**Pile bout ID**	**Median (±SD) SELss across depth & time** $y_{i,j}$ $(\pm\epsilon)$	**Max SELss at quietest depth**	**Max SELss at loudest depth**	**Initial distance (km)**	**Percent of MD values**
pv42-162-12	LS75.3	146 (±1.6)	144	148	10.5	0.018
pv42-162-12	LS65.5	136 (±0.1)	129	138	20.4	0.043
pv42-266-12	LS54.2	135 (±1.4)	127	137	25.0	0.011
pv42-277-12	LS75.2	156 (±3.7)	155	162	10.3	0.034
pv42-277-12	LS65.4	168 (±1.0)	162	170	8.1	0.016
pv42-290-12	LS74.1	139 (±1.5)	134	142	21.0	0.035
pv42-293-12	LS74.2	119 (±0.7)	118	121	28.6	0.050
pv42-295-12	LS56.1	144 (±1.8)	137	146	21.5	0.019
pv42-295-12	LS54.1	137 (±2.9)	133	144	17.4	0.014
pv42-316-12	LS66.1	160 (±1.3)	152	163	6.3	0.001
pv42-316-12	LS20.1	153 (±0.2)	143	155	11.9	0.040
pv42-320-12	LS75.3	169 (±2.6)	162	172	5.0	0.025
pv40-268-12	LS69.2	132 (±2.2)	132	136	12.4	0.002
pv40-268-12	LS56.1	135 (±1.7)	127	138	22.2	0.047
pv40-268-12	LS34.3	156 (±1.6)	145	157	12.5	0.033

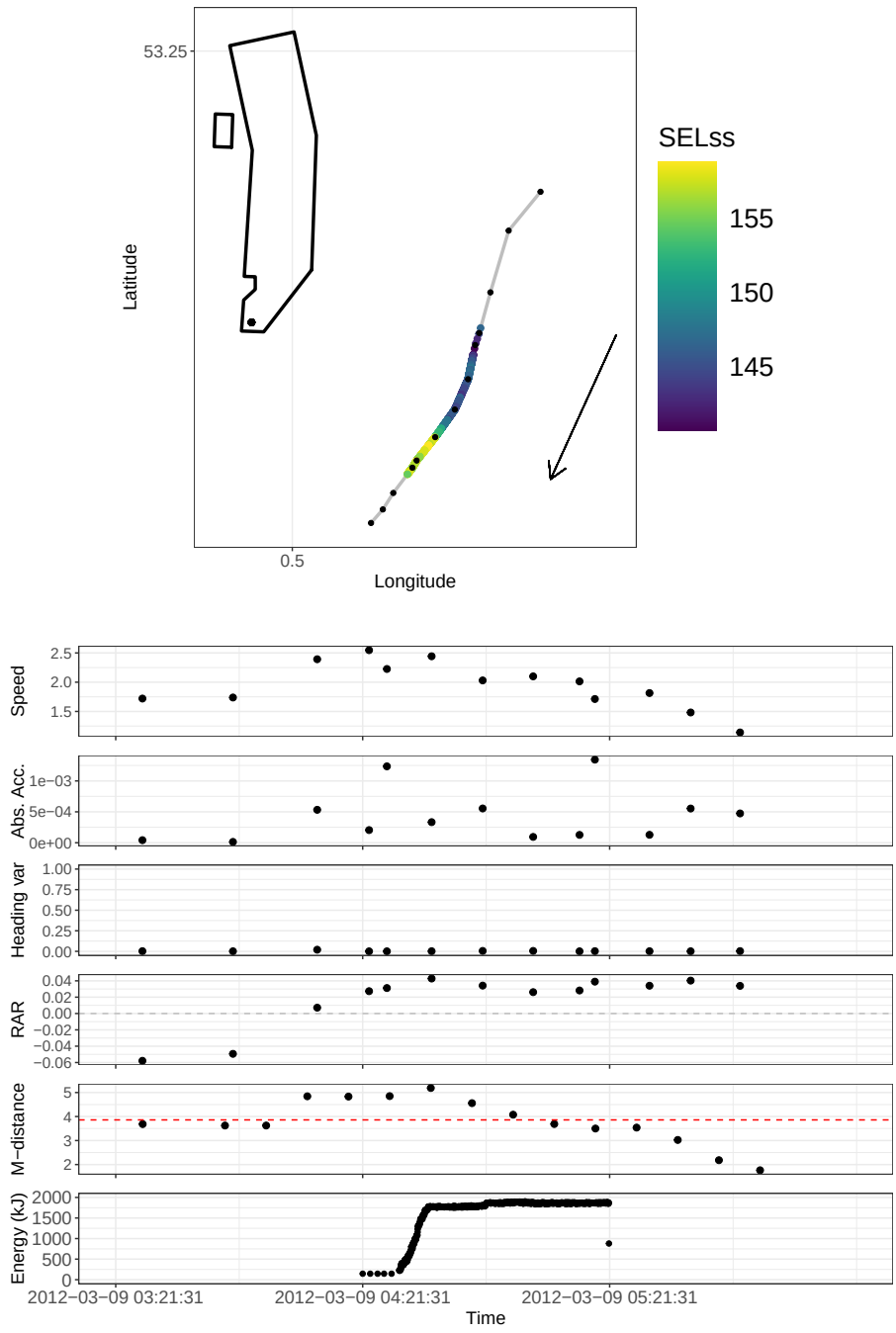


Figure 3.6 Example track and recorded metrics from one of the identified behavioural responses (seal pv42-162, initial distance of 10.5 km). Top panel: GPS locations (black dots) of a tagged individual moving south-west past Lincs offshore wind farm (black outline) during piling at pile LS75 (black dot within wind farm outline). Estimated received SELss (dB re $1\mu\text{Pa}^2 \cdot \text{s}$; median across depth) for each piling blow are also shown (coloured dots). Tracks are shown for one hour before and one hour after piling. Bottom panel: horizontal speed (m s^{-1}), absolute acceleration (i.e. change in speed, m s^{-2}), heading variation, radial avoidance rate (RAR), Mahalanobis distance, and piling blow energy from 1 hour pre-piling to 1 hour post-piling. The red dashed line denotes the Mahalanobis distance threshold (identified from randomisation tests) used to identify responses.

During one of the detected responses (Figure 3.6), the seal is travelling south-west at $\sim$1.7 m s^{-1} past the wind farm at a distance of approximately 10 km. As pile driving begins, the seal accelerates to a speed of $\sim$2.5 m s^{-1}, and continues travelling at a relatively high speed (>2 m s^{-1}) for approximately an hour as it moves away from the wind farm. Further plots and a brief description of each identified horizontal response are provided in the supplementary material for this chapter (Appendix C). Overall, the behavioural responses identified could be divided into three broad categories. First, encounters in which seals moved at unusually high speeds ($n = 6$). Seals either increased their speed while already travelling away from the wind farm, or turned around and then moved at a higher speed away from the wind farm. Second, encounters in which seals appeared to decelerate suddenly and reduce movement during pile driving ($n = 6$). Third, encounters in which seals suddenly initiated travel behaviour ($n = 3$), after a substantially long (3–12 hours) period of remaining relatively stationary. Responses were identified across the length of the study period in both sexes, both age classes, at different times of day, and during different estimated seal behavioural states.

3.4.2 The dose-response relationship

Both the full maximal model and the reduced simple model showed rapid convergence of the MCMC chains, with a Gelman-Rubin of 1.0. Gibbs variable selection indicated a low level of support for including covariates (sex, age class, daylight, exposure history, behavioural state, bout length, initial distance) within the dose-response model, estimating probabilities of 6–48% that each of these covariates are included in the true model (Table 3.4). Illustrative dose-response relationships of all considered covariates are shown in Figure 3.7.

As overall there was not enough evidence to support inclusion of covariates in the model, the reduced model (no covariates) was chosen to produce the final dose-response function (Figure 3.8; Table S1). Using this model, the mean estimated population response threshold μ was 186 (95% CI: 169–199) dB re 1μPa$^2\cdot$s, with a standard deviation between individuals ϕ of 14 dB re 1μPa$^2\cdot$s and between encounters σ of 24 dB re 1μPa$^2\cdot$s (Table 3.4). The estimated p_{50} for the final model was 175 (95% CI: 166–181) dB re 1μPa$^2\cdot$s. Posterior distributions from both of the considered models are presented in Appendix C (Figures S24; S25).

There was evidence of a sex-related difference in the encounters of harbour seals with pile driving (Figure 3.9). Across all 216 encounters, only 20 of these encounters occurred in

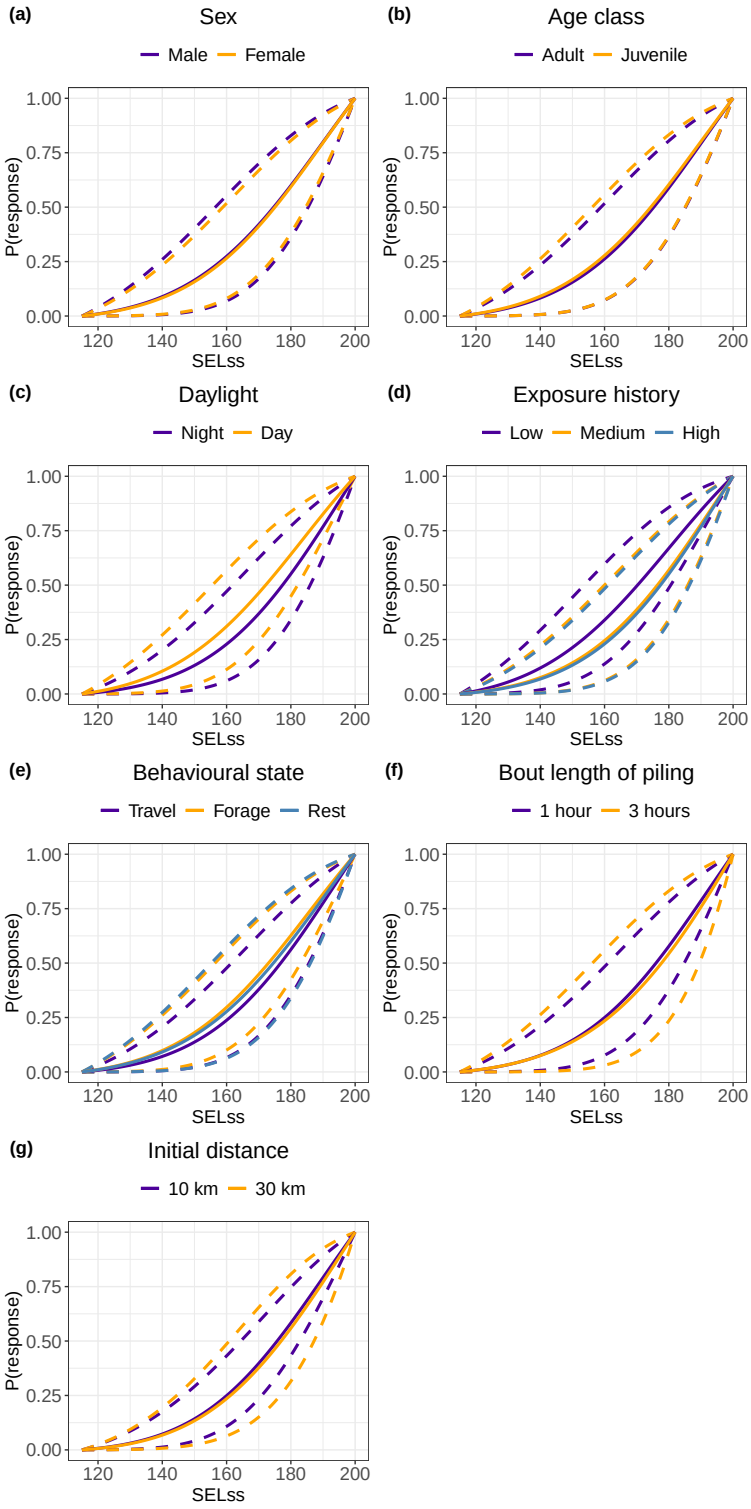


Figure 3.7 Estimated dose-respone relationships for horizontal responses in harbour seals as a function of single-strike sound exposure level (SELss in dB re $1\mu\text{Pa}^2\cdot\text{s}$; median across depth) of pile driving. Shown are results from the maximal model, with the predicted dose-response relationship from each covariate, assuming all were included in the model. For each covariate (panel), the mean (solid line) and 95% credible intervals (dashed lines) are illustrated.

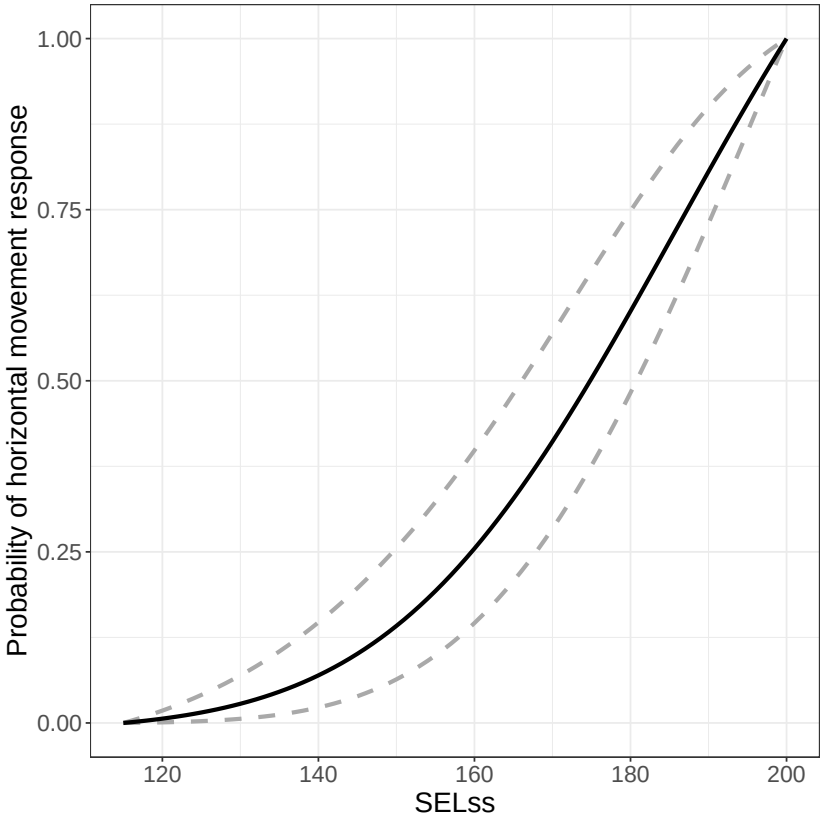


Figure 3.8 Estimated dose-response relationship for horizontal movement responses in harbour seals as a function of single-strike sound exposure level (SELss in dB re $1\mu\text{Pa}^2 \cdot \text{s}$; median across depth) of pile driving. The mean estimated function (black) and the 95% credible intervals (grey dashed) are shown.

Table 3.4 Estimated model parameters from posterior samples of the maximal model, and the final model with no covariates. Also shown are the results of Gibbs variable selection (GVS p), where the posterior probability of inclusion is the proportion of MCMC samples in which the candidate covariate (and associated α and β terms) is included. Factor covariates with three levels required two β parameters to estimate; Gibbs support was estimated for inclusion of the factor variable as a whole.

		Maximal model			Final model	
		Mean	**SD**	**GVS p**	**Mean**	**SD**
	μ	187	10.4	-	186	8.3
	ϕ	14	7.8	-	14	7.1
	σ	26	3.1	-	24	3.7
Sex:	α_{sex}	-0.5	7.7	0.42	-	-
Age class:	α_{age}	1.0	8.0	0.42	-	-
Daylight:	β_{day}	-5.6	5.2	0.45	-	-
Exposure history:	β_{exp1}	-9.0	6.1	0.48	-	-
	β_{exp2}	3.3	6.0		-	-
Behavioural state:	β_{behav1}	4.3	6.7	0.36	-	-
	β_{behav2}	-3.7	6.6		-	-
Bout length:	$\beta_{\text{lengthbout}}$	2.4	6.3	0.37	-	-
Initial distance:	β_{initdist}	0.3	0.7	0.06	-	-

males, and the number of encounters per individual was higher in females (median = 7, SD = 13) than males (median = 1, SD = 2). When encounters did occur in males, they tended to be at further initial distances from piling (median = 28.4 km) than in females (median = 20.2 km). Both sexes had similar temporal coverage of tag data (median = 98 days in males, 102 days in females; Table 3.1) and a near equal sex ratio was present in the tagged individuals used in analysis (11 males, 13 females). Therefore, it is possible this pattern may have been driven by sex-related differences in the distribution of seals across the study area (Figure S26). According to the haulout records on the GPS tags, the proportion of time spent hauled out was also slightly larger in males (median = 0.23, SD = 0.04) than in females (median = 0.18, SD = 0.02).

3.4.3 Effective response range and estimated numbers of seals affected

Using the final mean dose-response function (Figure 3.8), the predicted probability of response across the study area was estimated for each piling location (e.g. Figure 3.10(a)). The effective response range (ERR) estimated for each piling location ranged from 7.75 to 9.00 km, with a median ERR of 8.75 km (SD = 0.35 km; Figure 3.10(b)). It should be

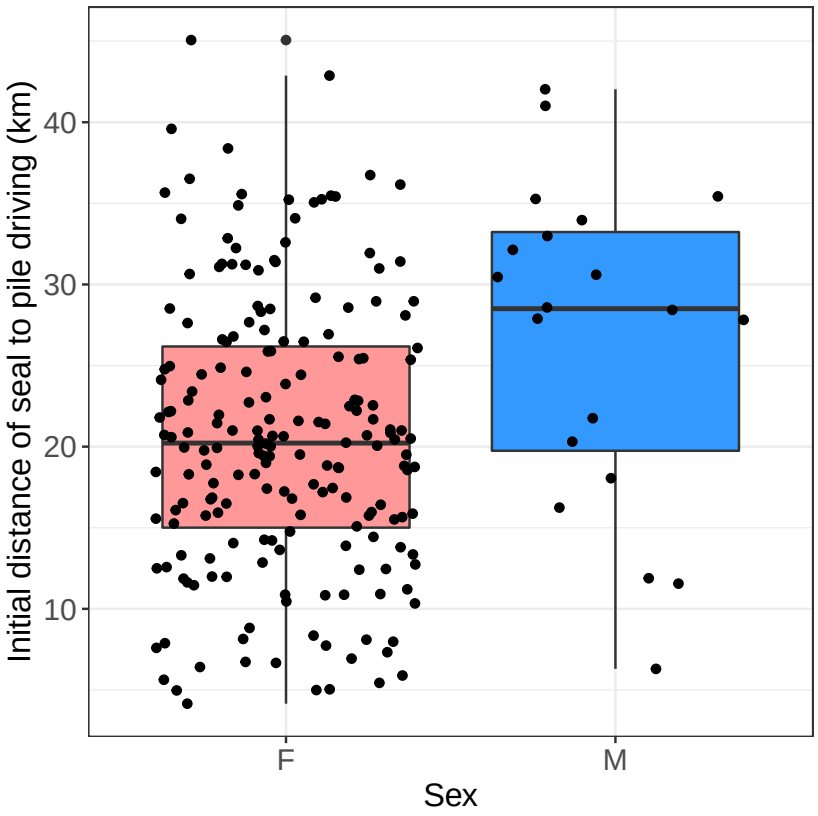


Figure 3.9 Initial distance of all harbour seal pile driving encounters by sex: female (n=196), male (n=20).

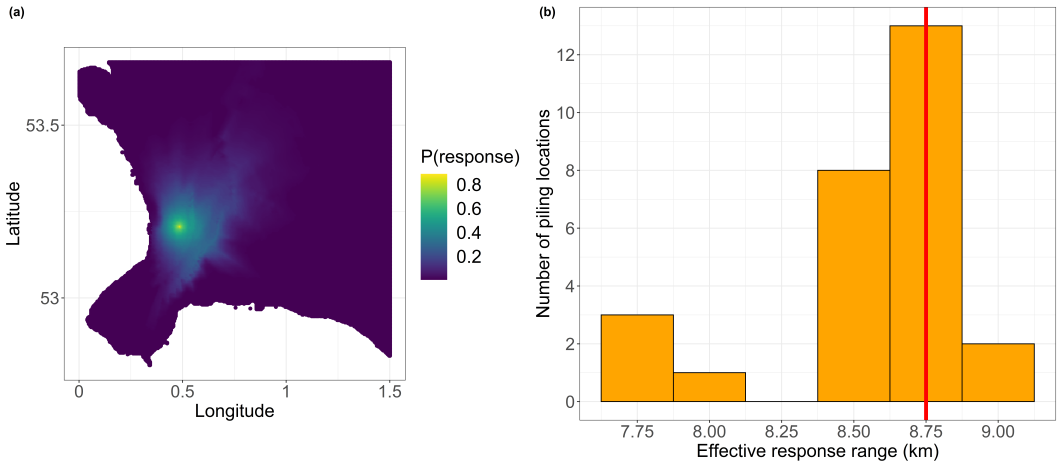


Figure 3.10 (a) Predicted response probability across the study area for one piling location (pile LS22) using the final mean dose-response relationship. (b) The estimated effective response range for each of the 27 piling locations. The median effective response range across all piling locations is shown by the red line.

noted that this uncertainty only accounts for variation in the probability of response among piling locations, and does not account for uncertainty in the dose-response relationship. The average density of harbour seals in The Wash near Lincs farm was estimated to be 1.30 seals per 1 km squared (Carter et al., 2020b). If a median ERR of 8.75 km is assumed, then this would correspond to, on average, 311 harbour seals exhibiting movement responses to a single pile driving bout.

3.5 Discussion

This study used GPS tracking data from 24 tagged harbour seals to examine the individual movement responses of seals to pile driving sounds during offshore wind farm construction. Twenty (83%) of the tagged seals were estimated to be exposed to pile driving sounds at levels above the median ambient sound level, and in nine seals (38%) a horizontal movement response was detected at least once during the study. Overall, the mean estimated population response threshold was at a single-strike sound exposure level (SELss) of 186 dB re $1\mu\text{Pa}^2\cdot\text{s}$, with high variability between (SD = 14) and within (SD = 24) individuals. The estimated p_{50} sound level at which there is a 50% chance of a horizontal movement response was 175 dB re $1\mu\text{Pa}^2\cdot\text{s}$ (95% CI: 166–181). There was limited evidence to support inclusion of individual-level or encounter-level covariates in the dose-response relationship; however, there was a difference in the number of times male (median per individual = 1)

and female (median per individual = 7) individuals were exposed to pile driving sounds. The median effective response radius (ERR; the distance at which the number of seals predicted to respond beyond this radius is equal to the number of seals failing to respond within this radius) for this study was estimated to be 8.75 km, equivalent to approximately 311 harbour seals responding per pile driving bout.

Metrics of seal horizontal movement behaviour (speed, acceleration, heading variability, radial avoidance rate) were used to identify times when individuals exhibited significantly unusual movement behaviour. The approach used here was multivariate, i.e. all variables were considered together, and so the identified responses are unusual with respect to the combination of all of the movement metrics. In the detected responses in this study, three broad types of unusual horizontal movement behaviour were observed. During pile driving, seals were either observed to (1) exhibit unusually high travel speeds, (2) suddenly decelerate and reduce horizontal movement, or (3) initiate travel away after a long time period of limited horizontal movement. There are a variety of possible reasons why seals may respond in these ways. Seals which respond by approach 1 or 3 are increasing their movement away from the pile driving source, either by increasing their speed when they are already intending to pass the wind farm, or by initiating movement when they were previously relatively stationary. The initial burst of speed observed in some of the detected responses could be as a result of an initial startle response to the sound itself. Captive playback studies of sounds with a sudden onset have been found to elicit a startle reflex in grey seals (Götz and Janik, 2011). Repeated playbacks of the sound led to sensitisation (increased responsiveness over time), with responses consisting of rapid fleeing and spatial avoidance of the sound source (Götz and Janik, 2011). Seals that increase the rate at which they move away from the sound source are also reducing their potential noise exposure, and theoretically escaping from a perceived threat or nuisance sound. In contrast, seals which respond by approach 2 appear to stop horizontal movement altogether. Seals which behave in this way may be ceasing their current behaviour to assess the potential threat situation, or waiting to resume their intended behaviour once the disturbance is over. Whilst these seals are not clearly reducing their potential noise exposure by moving away, if they are spending more time near the surface they would have a reduced noise exposure. The dive behaviour of harbour seals during pile driving is analysed separately in Chapter 4.

Overall, data on the individual movement responses of pinnipeds to underwater sounds are relatively limited. An unpublished report by Aarts et al. (2018) examined tagged grey seal movement behaviour during pile driving in the North Sea. On at least one occasion, grey seals were observed with a persistently high swim speed (1.8 m s^{-1}) during piling; sudden

changes in heading direction were observed on several occasions, as well as changes in dive behaviour. In Aarts et al. (2018), each movement metric (e.g. dive depth, speed) was considered separately (in contrast to all metrics considered together in this Chapter), and only the behaviour 4 hours pre-piling and 2 hours post-piling was considered (in contrast to the ∼33 days of combined baseline data per individual here). Gordon et al. (2019) also observed some increases in swim speed and changes in heading in free-ranging tagged harbour seals exposed to playbacks of different acoustic deterrent devices. During sound exposure, seals that were assumed to be previously foraging showed significant increases in swim speed; however, the increase in speed was only slight (∼7%) for seals travelling prior to exposure (Gordon et al., 2019). In Gordon et al. (2019), each movement metric was also considered separately, and these results were part of a targeted exposure experiment, rather than the *in-situ* disturbance events examined in this Chapter. Considered together, these initial findings (this Chapter; Aarts et al., 2018; Gordon et al., 2019) of individual seal movement responses to sound disturbance are broadly consistent with studies of harbour porpoise (*Phocoena phocoena*), where changes in horizontal movement behaviour and swim speed (Kastelein et al., 2018b; Mikkelsen et al., 2017; van Beest et al., 2018) have been observed during noise exposure. Further data collection and further understanding of individual responses by both pinnipeds and small cetaceans is required. Understanding behaviour at the individual level is a critical step in understanding population-level effects, and in informing simulation-based approaches (Chudzinska et al., 2021; Nabe-Nielsen et al., 2018), which are increasingly being used in environmental impact assessments.

As multiple Mahalanobis distance randomisation tests were carried out, it is possible that some of the identified responses were false positive detections (Chapter 2). The approach used in this chapter provided a mechanism to identify the most unusual instances of movement behaviour, and it was assumed that these changes were related to the sound level of pile driving activity. The relatively few detected responses in this chapter is likely because, from the available data observations in this study, the measurements of movement behaviour during behavioural responses are not highly extreme, when compared to a long duration (weeks–months) of baseline data. This does not, however, mean that these behavioural changes are not biologically important. For example, seals which increase their movement speed may incur increased energetic expenditure, seals which are prompted to move away may be leaving important resting or foraging areas, and seals which reduce movement may be delaying their intended behaviour (e.g. travelling to a foraging area). The observational approach used in data collection in this study also meant that there were relatively few close-range encounters (<10 km) between seals and piling, and that

the baseline data may have included other potential sources of disturbance (e.g. shipping, predator encounters). The benefits and challenges of using observational data in this way are discussed further in the General Discussion. From the simulations conducted in Chapter 2, it is clear that weaker horizontal responses were not detected as well (up to 64%), and even strong responses were not detected all of the time (up to 79%). It is therefore also likely that there are other undetected horizontal responses within the data. As the data considered in this chapter were from GPS locations only, there are also possible response behaviours that would never be detected by this approach (e.g. porpoising out of the water during fast swimming). Furthermore, whilst the variable frequency of GPS locations over time makes this challenging, estimation of the true duration of behavioural responses is an important area of further work.

The detected horizontal responses in this chapter provide likely mechanisms for the seal density changes described in Russell et al. (2016) and Whyte et al. (2020). Although the approach used here was focussed on identifying sudden individual change-points in behaviour, the resulting changes to movement could contribute to different patterns in seal density over space. Seals which increase their speed or decide to move away from the sound source during piling will lead to lower densities of seals closer to the wind farm. Seals which pause in their current movement trajectory may reduce the number of seals entering areas closer to the wind farm. Additionally, the observed responses highlight that seals can respond to pile driving sounds by increasing swim speeds (up to 2.5 m s^{-1}). In particular, Figure 3.6 shows an individual that responded by sustaining swim speeds of over 2 m s^{-1} for approximately one hour. This highlights a potential concern for data cleaning practices, which sometimes assume a maximum swim speed of seals of 2 m s^{-1} when removing erroneous locations (Carter et al., 2017; McConnell et al., 1999). Typically this approach is only used for lower-resolution ARGOS satellite locations; however, careful consideration of the maximum swim speed and time interval over which this speed is assumed is required in order to avoid bias and the removal of potential disturbance behaviours from biologging datasets.

In estimating the acoustic dose received by seals, there were several sources of potential uncertainty. Due to the resolution of the tagging data, and the duration of behavioural responses expected, the Mahalanobis distance analysis used to identify responses was carried out using 20-minute time windows. The received level of sound at the seal will vary during this 20-minute interval as a result of the seal's changing location and the hammer energy of the pile driving itself (Figure 3.2). Here, the median and standard deviation of the received levels for all blows within the response interval was used to account for this;

however, in practice, the received dose did not vary highly within this interval (maximum SD of 2.18; Table 3.3) and was more strongly determined by the seal's initial location and whether the response was detected in the ramp-up phase or full energy pile driving. Variation in potential received level across depth was accounted for by using the median sound level across all modelled depth bins for each location. This dose therefore represents the acoustic dose available at a given location (and therefore reflects the information typically available in impact assessments), but not necessarily the exact sound level received by the seal at that time. Incorporating information on dive behaviour or using biologging devices which record acoustic information (Mikkelsen et al., 2019) would be important steps to examine this in further detail; however, the possible sample size and study longevity with high-resolution tags such as these is typically more limited. Uncertainty in the predictions from the acoustic model will also contribute to uncertainty in the estimated doses. A comparison of the acoustic predictions to opportunistic recordings of pile driving in the study area gave an estimated mean absolute error of 4 dB re $1\mu\text{Pa}^2 \cdot \text{s}$ across all measured piling blows (Figure S27; Whyte et al., 2020). These acoustic comparisons were not part of a formal acoustic validation exercise, and were relatively limited in spatial and temporal coverage. Future studies should aim to conduct more comprehensive acoustic measurements across the study area to validate predicted estimates. Expanding the hierarchical model to account for this uncertainty in modelled sound level across the study area would be a key extension to this work, and may reduce the estimated variability between encounters, which appears to be constrained by the upper limit of the prior distribution (Figure S24).

A Bayesian hierarchical model was used to estimate the distribution of response thresholds across the population, and quantify the dose-response relationship between piling sound level and the probability of horizontal response. This approach gave an estimated mean population response threshold μ in SELss of 186 dB re $1\mu\text{Pa}^2 \cdot \text{s}$, with an estimated p_{50} of 175 dB re $1\mu\text{Pa}^2 \cdot \text{s}$. The estimated standard deviation between individuals ϕ and between encounters σ was high at 14 and 24 dB re $1\mu\text{Pa}^2 \cdot \text{s}$ respectively. Other studies using this approach have estimated similar levels of variation for pilot whale (ϕ=18; σ=21; Antunes et al., 2014) and killer whale (ϕ=17; σ=26; Miller et al., 2014) responses to sonar, although it should be noted that different units of acoustic measurement (SEL_{cum}; dB re $1\mu\text{Pa}^2 \cdot \text{s}$) were used in their analysis. Overall, the results highlight that the levels of sound that can initiate behavioural responses in seals can be highly variable. It is likely that behavioural responses are dependent on both individual variability and the ecological, behavioural, and exposure context (Ellison et al., 2012; Isojunno et al., 2017) of the encounter with sound; considering the relative role of both of these factors are important areas of future research.

A suite of candidate covariates were included in the dose-response model to assess whether response thresholds differed by individual and encounter-level conditions; however there was insufficient evidence to support inclusion of them within the final model (Table 3.4). Illustrative dose-response curves (Figure 3.7) of the estimated effect of each of these covariates suggested that the difference in response probability between individuals of different sexes and age classes (Figure 3.7(a)(b)) was very limited. There was weak evidence that animals were more likely to respond during daylight hours, which may have suggested they tolerate higher sound levels during pile driving at night. When individuals had a lower previous exposure history, there was weak evidence that they were more likely to respond (Figure 3.7(d)). Responses of harbour porpoises to pile driving sounds have been found to decline over time (Graham et al., 2019), highlighting the value of long-term studies of disturbance and responses to sound. These changes could be a result of habituation to the sounds, or potentially sound-related hearing damage as a result of repeated exposures. There was also weak evidence that individuals were more likely to respond when they were previously in an estimated resting state. It is possible that individuals which have higher behavioural motivation to continue foraging or continue travelling would tolerate higher levels of noise before responding. Gibbs variable selection showed very low support (0.06 probability of inclusion in the model) for inclusion of distance as a covariate. Distance is highly correlated with predicted sound exposure level. In this analysis, the inclusion of distance in the dose-response relationship does not appear to provide additional information, compared to using sound level alone. All of these observed patterns should be interpreted with caution, as in practice the support for all of these variables was very low (GVS p scores of 0.09–0.48). Similar studies that have included covariates in dose-response models have not identified GVS support above 0.54 (Antunes et al., 2014; Miller et al., 2014). The relatively small number of detected responses likely provides low statistical power to detect the effect of additional covariates on the dose-response relationship. It is also likely that additional contextual factors are present in reality which are not captured by the covariates available in this study.

In this study, tagged female seals were exposed to pile driving bouts more often than males. The observed distribution of GPS locations across the study area (Figure S26) suggests that females, in general, tend to be spend more time close to the coastal wind farm site, compared to males which appear to spend more time further offshore. Sex-related differences in activity budgets of grey and harbour seals have been described in Russell et al. (2015); in particular, the proportion of time spent in different behavioural states (travelling, foraging, resting) depended on the combination of sex, age, and time of

year. In contrast, Sharples et al. (2012) found that sex was a relatively poor predictor of trip duration and distance travelled from the haulout site in harbour seals, compared to factors such as geographic region or time of year. Further investigation into sex-related differences in habitat use of proposed offshore development sites is required to understand how any potential differences may influence population-level predictions of noise exposure and potential impact. Sex-related differences may also differ by region, and so it is unclear if this observation is more widespread to other harbour seal populations. It is important to note that this study was limited to a single four-month period of the year (January–May), and therefore it is likely that differences between the sexes may be more extreme during the breeding season (June–July).

The dose-response curve produced in this chapter (Figure 3.8) can be used by offshore developments to predict the probability of horizontal movement responses for a given sound exposure level of pile driving. In doing so, the provided credible intervals (Table S1) should be used to account for uncertainty in this relationship. It should be noted that this dose-response curve was produced based on harbour seals from a single population in the UK and at a particular time of year (January–May). The wind farm construction took place near the only entrance to the main haulout site, and so individuals had to pass within ∼20 km of the wind farm to return or leave the haulout site. Further particulars of this study population, and potential implications for the observed findings, are discussed further in the General Discussion.

3.5.1 Conclusion

This study identified unusual changes in the movement behaviour of individual harbour seals during exposure to pile driving sounds, and presented a dose-response function that can be used to predict these behavioural responses for different levels of sound exposure. The detected horizontal responses improve our understanding of how individual seals may react to the presence of *in-situ* anthropogenic sounds, and, in particular, provide insights into the mechanisms which contribute to population-level changes in density over space (Russell et al., 2016; Whyte et al., 2020). By investigating movement and behaviour at the individual level, variation between individuals and between piling events was accounted for, and it was possible to observe sex-related differences in predicted sound exposure. Overall, this chapter provides an important contribution towards estimating a direct link between predicted sound exposure, behavioural changes, energetic consequences, and ultimately population-level effects of sound disturbance.

Chapter 4

Dive behaviour of harbour seals during pile driving for offshore wind farm construction

The acoustic propagation models (in Section 4.3.3) were produced by Bas Binnerts and Sander von Benda-Beckmann at TNO (Acoustics and Sonar Expertise Group, TNO, The Hague, Netherlands), in consultation with myself, Gordon Hastie and Carol Sparling. The rest of the analysis described in this chapter is my own work.

4.1 Abstract

For many aquatic mammals, diving is an essential behaviour to obtain food; however, in pinnipeds, the potential effects of human disturbance on dive behaviour are poorly understood. In particular, pile driving activities for offshore wind farms produce high intensity sounds in areas which overlap, spatially and temporally, with seals. Understanding how the dive behaviour of seals may be affected during exposure to these sounds is critical in predicting how pile driving may affect both individual fitness and the population-level consequences of disturbance. In this chapter, I used dive records from biologging tags deployed on 24 harbour seals (*Phoca vitulina*) to examine changes in dive behaviour during pile driving at an offshore wind farm. Metrics summarising dive behaviour were calculated. These metrics were used to identify times of unusual dive behaviour (by Mahalanobis distance randomisation tests) and quantify the relationship of these dive behaviours with piling sound level (by a Bayesian hierarchical model). Hidden Markov models (HMMs) were also used to investigate the effect of pile driving on the probability of seals switching between different dive types. There were 232 encounters between diving harbour seals and pile driving, and in 16 of these encounters unusual dive behaviour was detected. Twenty of the tagged seals (83%) were predicted to have been exposed to pile driving sounds during the study, and nine seals (38%) were identified as exhibiting unusual dive behaviour at least once. The detected unusual dive behaviour consisted of dives that had a shorter dive duration, a longer post-dive surface duration, used a lower proportion of the available water column, spent a lower proportion of the dive in the bottom phase, and had a slower descent rate. The mean population-level response threshold at which these changes were estimated to occur was 188 (95% CI: 172–199) dB re $1\mu\text{Pa}^2 \cdot \text{s}$, with a standard deviation between individuals of 16 dB re $1\mu\text{Pa}^2 \cdot \text{s}$ and between encounters of 25 dB re $1\mu\text{Pa}^2 \cdot \text{s}$. The estimated sound level at which there was a 50% probability of response, p_{50}, was 175 (95% CI: 167–180) dB re $1\mu\text{Pa}^2 \cdot \text{s}$. Pile driving was also found to have a significant effect on dive behaviour-switching in ten (42%) of the tagged seals. In general, seals were more likely to switch between dive types during pile driving than during baseline (non-piling) periods. Overall, this chapter contributes the first in-depth study of seal dive behaviour during anthropogenic disturbance. The observations that seals may dive less fully (e.g. shorter, shallower) and more irregularly (e.g. switch dive types more often) during disturbance provide key insights into the potential impacts of anthropogenic sound on diving animals, information which is critical for the future management of human activities.

4.2 Introduction

When air-breathing animals dive to forage, they must balance the importance of accessing potential prey with the need to access oxygen at the surface (Doniol-Valcroze et al., 2011; Thompson and Fedak, 2001). For marine mammals, diving is often the main mechanism by which they can search for and obtain food; however, diving can also serve other important biological functions. Diving below the water surface may aid in predator (or threat) avoidance (Heithaus and Frid, 2003), enable animals to engage in social behaviour with conspecifics (Hanggi and Schusterman, 1994), and reduce the effects of surface currents or adverse weather when animals are travelling or resting (Lyamin, 1993). Understanding changes in dives can therefore provide important insights into both the natural drivers of this behaviour, and into how disturbances may alter animal activity.

For phocid seals in particular, diving constitutes a large proportion of their time. Whilst at sea, phocids can spend ∼80–90% of their time submerged in dives (Bekkby and Bjørge, 2000; Carter et al., 2020a; DeLong and Stewart, 1991; Ries et al., 1997) and are highly adapted for prolonged periods of diving (Blix, 2018). Harbour seals (*Phoca vitulina*), a common species in the northern hemisphere, have a primarily coastal distribution (Sharples et al., 2012) and are considered to be generalist predators, feeding on a variety of fish and marine invertebrates including sandeel, gadids and flatfish (Tollit et al., 1998; Wilson and Hammond, 2019). Foraging for these prey is typically thought to involve seals diving close to the seafloor, but in some cases mid-water ("pelagic") and opportunistic foraging, for example whilst travelling to a foraging site, may also occur (Blanchet et al., 2015; Planque et al., 2020; Tollit et al., 1998). Harbour seals typically dive for ∼1–10 minutes per dive, with a wide variety of different shapes of dive profile (Baechler et al., 2002; Blanchet et al., 2015).

The coastal distribution of seals, driven by the requirement to regularly haul out, places seals in close proximity to a range of human activities, including fishing (Oksanen et al., 2014), shipping (Jones et al., 2017) and industrial activity. In the last few decades, the rapid expansion of the renewable energy sector has led to the construction of a variety of new structures in the marine environment. In particular, offshore wind developments are expanding across the globe, with 14 GW of offshore wind capacity in 2016 predicted to increase to 41 GW by 2022 (Díaz and Guedes Soares, 2020). In the UK, wind farms have so far been built at a mean water depth of ∼16 m and at a mean distance to shore of ∼12 km (Díaz and Guedes Soares, 2020), resulting in a high spatial overlap with harbour seal distributions (Carter et al., 2020b; Sharples et al., 2012). Impact pile driving, used

to construct many offshore wind farms, produces high intensity sounds for several hours at a time during the installation of each turbine (Bailey et al., 2010). For animals present near to pile driving, these sounds have the potential to cause hearing damage (Hastie et al., 2015; Whyte et al., 2020), and disrupt movement and behaviour (Benhemma-Le Gall et al., 2021; Brandt et al., 2011; Whyte et al., 2020).

Disruptions or changes to dive behaviour during human activity have been observed in a number of marine mammal species. Cuvier's beaked whales (*Ziphius cavirostris*), a species in which strandings have been associated with the timing of naval sonar activities (Simonis et al., 2020), have been shown to undertake longer dives and longer surface intervals during exposure to active sonar (DeRuiter et al., 2013; Falcone et al., 2017). Killer whales (*Orcinus orca*) have been shown to respond to sonar by switching from deep dives to shallow dives at the onset of the sound, and long-finned pilot (*Globicephala melas*) and sperm whales (*Physeter macrocephalus*) may also reduce dive depth and duration (Sivle et al., 2012). Studies of diving behaviour in wild pinnipeds during human disturbance have been relatively more limited. Anecdotal observations of biologging data from grey seals (*Halichoerus grypus*) exposed to shipping noise indicate that dive patterns may be disrupted during these encounters (Mikkelsen et al., 2019). For pile driving sounds in particular, there are no published studies on the potential effect of these sounds on marine mammal diving behaviour; however, an unpublished report by Aarts et al. (2018) observed significant decreases in the descent rate of dives in grey seals during pile driving at an offshore wind farm in the North Sea. Overall, a more comprehensive understanding of how seals may alter their dive behaviour during sound disturbance is required, as it is an important step towards understanding how populations may be affected by different sources of human activity.

The challenge of detecting alterations in dive behaviour as a result of disturbance is twofold. First, dive behaviour can be highly variable, even within the same individual. Often, each individual can exhibit a wide repertoire of dive shapes and durations, with different intervals between them (Baechler et al., 2002; Lesage et al., 1999). During undisturbed periods, the dive behaviour of marine animals will change over varying timescales, as animals pass through different environments and need to balance different biological needs (e.g. oxygen, feeding, resting). Approaches to detect unusual dive behaviour must therefore incorporate some assessment of what dive behaviours and/or patterns are normal, before attempting to identify potential disturbances. Second, dives recorded on animal-borne tags typically consist of long, continuous time series of data on a few individuals (e.g. 1,000 to 50,000 dive records per individual). Methods to analyse and quantify these data

must therefore be efficient. Methods that take account of the temporal nature of this data (repeated measures on individuals over time) may also provide added insight into the drivers and patterns of dive behaviour. There are currently two main statistical methods used to detect disruptions in dive behaviour of marine mammals. One approach, aimed at detecting unusual or extreme behaviours, uses the multivariate statistic Mahalanobis distance to summarise dive behaviour metrics (DeRuiter et al., 2013; Kvadsheim et al., 2017). Values are compared between dives observed during disturbance to those observed during baseline (non-disturbance) to identify particular dives (or groups of dives) that are statistically unusual. An alternative approach uses hidden Markov models (HMMs) to first classify dive behaviour into a given number of behaviour types, called 'states'. These can then be used to model the transitions between these different dive states over time, and quantify whether disturbance affects the switching between states (DeRuiter et al., 2017) or the overall time budget within each state (Isojunno et al., 2017). Here, both methods were used to assess whether pile driving disrupted different aspects of seal dive behaviour.

4.2.1 Aims

This chapter aims to investigate the effects of pile driving activity on the dive behaviour of individual harbour seals. Dive records from 24 animal-borne tags are used to calculate metrics summarising dive behaviour and use these metrics to (1) identify times of unusual dive behaviour, (2) quantify the relationship of this unusual behaviour with pile driving sound level, and (3) determine the effect of pile driving on individuals switching between different dive behaviours.

4.3 Methods

4.3.1 Data collection

Seal tag data

To record the dive behaviour of seals present near wind farm construction activity, GPS/GSM tags (SMRU Instrumentation, University of St Andrews, Fife, UK) were deployed on harbour seals in The Wash, south-east England, UK. Seals were first anesthetized using Zoletil® or Ketaset® in combination with Hypnovel®, and GPS/GSM tags were attached to the fur at the back of the neck using a fast-setting two-part epoxy adhesive or Loctite®

422 Instant Adhesive. All seal handling and procedures were carried out under Home Office Licence 60/4009. A total of 24 tagged harbour seals (11 males, 13 females) were present in The Wash during this study (Hastie et al., 2015; Russell et al., 2016; Whyte et al., 2020). This included 22 seals tagged on intertidal sandbanks in The Wash, and two seals from a concurrent study tagged in The Thames, UK.

The deployed GPS/GSM tags recorded GPS location data and summary information on dive behaviour. Dive recording on the tag was derived from on-board pressure sensors and was initiated when seals were 1.5 m below the water surface for at least 8 seconds, and raw depths ('diving depth points') were recorded at 4-second intervals with 0.1 m resolution within each dive. To accommodate transmission limitations of telemetry, the dive data were then summarised on-board the tag. Once each recorded dive ended (i.e. the seal was shallower than 1.5 m), the dive profile was interpolated between all the recorded diving depth points. Using the interpolated dive profile, the depth at nine given time proportions of the dive (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%) was recorded, resulting in nine pre-defined depth points evenly spaced in time throughout each recorded dive. Each depth point was then assigned to a depth bin for transmission; depth bins were 0.4 m wide for depths recorded from 1.5 – 50.7 m, and 0.8 m wide for depths recorded from 50.8–153.0 m. The resulting dive data reported from the tag consisted of the depth bin entries for each of the nine time points during the dive, the total dive duration, and the duration of the post-dive surface interval (Figure 4.1).

The location data were cleaned and erroneous locations removed based on thresholds of residual error and the number of satellites ($n \geq 5$). Dive locations were estimated from linear interpolation between available GPS locations at the surface. Dives which had no recorded GPS locations within one hour of the dive time were removed to ensure that dives were not included when the location of the dive was potentially uncertain. The median dive duration was 3 minutes, with a median time interval between GPS locations of 10 minutes.

Pile driving operations

Operational data on pile driving at Lincs offshore wind farm were provided by Centrica plc. Throughout the period of the seal tag deployment, 27 (of a total of 75) monopiles were installed at Lincs by pile driving. Between 28^{th} January and 11^{th} May 2012, a total of 77,968 piling strikes occurred, with a mean strike energy of 1,202 kJ (SD= 613).

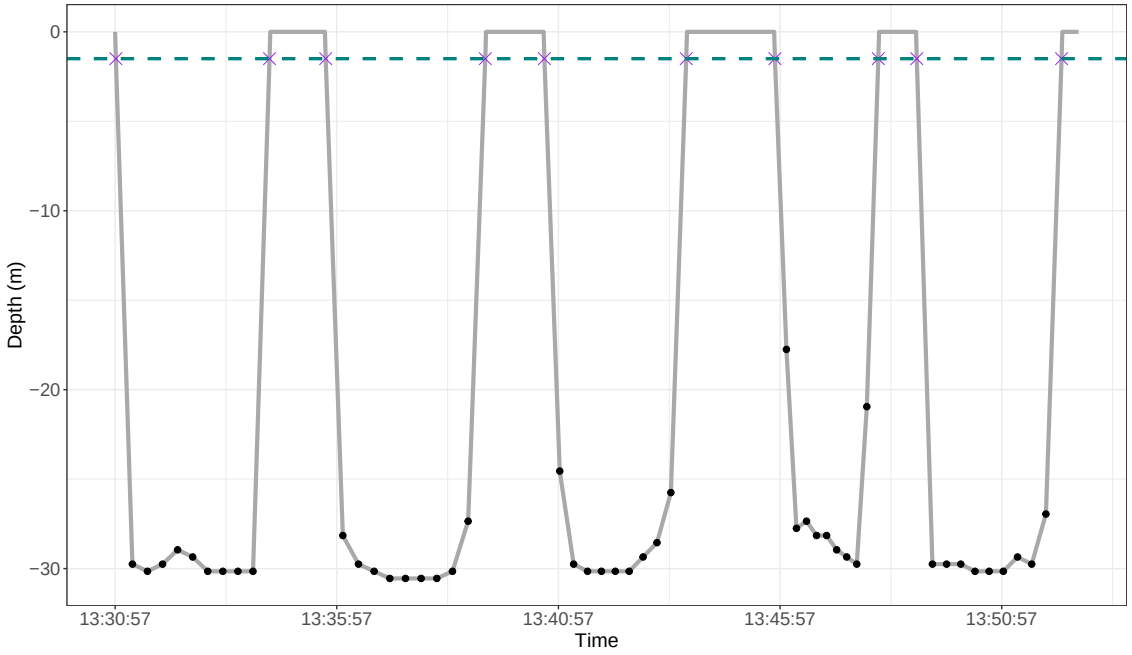


Figure 4.1 An example sequence of five dives (from individual pv42-194-12 on 30$^{\text{th}}$ January 2012). Shown are depth bins recorded at nine equally-space time points during each dive (black dots), and the start and end times of each dive (purple crosses). The tag begins and ends recording the dive at 1.5 m (green dashed line); depths shallower than 1.5 m are recorded as a post-dive surface interval.

Pile driving at Lincs occurred intermittently across the study period, with a maximum gap of 19.5 days with no piling activity. To compare time intervals of piling and non-piling, individual records of piling hammer blows were divided into bouts, where piling was considered as a separate bout if successive hammer blows were more than one hour apart. This resulted in 64 piling bouts across the study period and the 27 piling locations. Piling bouts had a mean duration of 1.0 hour (min = 0.2, max = 3.2), containing an average of 1,218 hammer blows each (min = 132, max = 3,772) and a median inter-strike interval (time between successive hammer blows) of two seconds.

4.3.2 Quantifying dive behaviour

For each tagged seal, dives recorded in the first week of deployment were removed to ensure that any potentially unusual dive behaviour following the tagging procedure (due to sedation or disturbance) was not included in the analysis (McKnight, 2011). Dives recorded more than one week after the last pile driving bout were also removed. This ensured that comparisons of dive behaviour were limited to the same time of year (January–May), and excluded data collected during the harbour seal breeding season (June–July) when movement behaviour and dive behaviour is likely to be different for breeding individuals (Thompson et al., 1994; Van Parijs et al., 1997).

As seals often dive close to the seafloor (Tollit et al., 1998), observed dive depths may be more representative of seafloor topography than behavioural choices to adjust dive depth. To examine meaningful metrics of dive behaviour, dives were matched to environmental data on water depth across space and time. The location of each dive was matched to bathymetry data from the EMODnet Data Portal (∼115 m resolution) to obtain the bathymetric water depth b_{xy} at Lowest Astronomical Tide (LAT) for each dive. The UK Hydrographic Office Vertical Offshore Reference Frame (VORF) correction factor v_{xy} was obtained for each dive location (Iliffe et al., 2013), to estimate the difference in water depth between LAT and Mean Sea Level (MSL). To account for changes in water depth across the tidal cycle, tidal height data from POLPRED (for one central location in the study area; National Oceanography Centre, 2020) was used to calculate the difference in tidal height h_{xyt} at the time of each dive within the study area. The resulting estimated available water depth at the time and location of each dive A_{xyt} was calculated by $A_{xyt} = b_{xy} + v_{xy} + h_{xyt}$.

Dives were divided into phases using the recorded depth points. As each of the nine depth points per dive were recorded in depth bins, the midpoint of each depth bin was used as the estimated depth for use in analyses. Interpolation between the recorded depth points

was used to estimate the overall time-depth dive profile at a time resolution of 0.1 seconds. Seals were considered to start the bottom phase of a dive when their dive was $\geq 85\%$ of the maximum depth of that dive (Baechler et al., 2002; Wilson et al., 2014). It was assumed that the bottom phase ended when the seal swam shallower than this 85% threshold for the last time in that dive, i.e. if a seal ascended and then descended again, this was all considered to be part of the bottom phase. The descent and ascent phases were considered to be, respectively, the time periods before and after this identified bottom phase (Figure 2.3).

For each dive, the following metrics were calculated. These metrics were chosen as they represented several biologically meaningful aspects of dive behaviour, following previous studies (Blanchet et al., 2015; Photopoulou et al., 2014; Wilson et al., 2014):

- **Dive duration:** The total duration of each dive (seconds). This duration refers to the time spent more than 1.5 m below the water surface.
- **Surface duration:** The duration of the post-dive surface interval following each dive (seconds). This duration includes any time spent shallower than 1.5 m, and so includes both very shallow diving and time at the surface.
- **Proportion of water column used:** The proportion of the available water column used by the dive, calculated by $\frac{D_{xyt}}{A_{xyt}}$ where D_{xyt} is the maximum depth reached within the dive at location x, y and time t. The available water depth A_{xyt} for each dive accounted for spatial changes in bathymetry and temporal changes during the tidal cycle, as described above. Dive depths recorded that were greater than the estimated seabed depth were assigned to a value of 1.
- **Bottom phase proportion:** The proportion of the dive duration spent within the estimated bottom phase of the dive (see Figure 2.3).
- **Descent rate:** The rate of change in depth (m s^{-1}) during the descent phase of the dive. The descent phase occurs between the start of the dive (at 1.5 m) and the start of the estimated bottom phase of the dive (Figure 2.3).

4.3.3 Acoustic modelling of pile driving sounds

To estimate the sound levels resulting from piling across the study area, the Aquarius pile driving model was used. The model and its specifications are described in Chapter 3 (Section 3.3.4).

4.3.4 Using Mahalanobis distance randomisation tests to identify unusual dive behaviour and Bayesian hierarchical models to estimate dose-response

Specifying baseline data

To evaluate whether seals exhibited unusual dive behaviour during pile driving, it was necessary to first specify baseline periods where it was assumed that the individual was exhibiting normal dive behaviours. For each individual, only dives that took place within 100 km from the central location of active pile driving were included in the baseline. This approach ensured comparisons of dives observed during baseline and pile driving took place within the same spatial area. Dives that occurred in the 6 hours prior to any pile driving bout were excluded from baseline to ensure that any potential disturbance from the setup of pile driving activity was not included as baseline behaviour. Dives that took place in the 24 hours following any pile driving bout were also excluded from baseline to allow time for seals to return to normal behaviour after any potential disturbance.

Summarising dive metrics by Mahalanobis distance

To quantify how dive behaviour during pile driving compared to pile driving during baseline periods, the multivariate statistic Mahalanobis distance was used to summarise dive behaviour (DeRuiter et al., 2013) into a single number. Five metrics summarising dive behaviour (described above) were used: dive duration, surface duration, proportion of the water column used, bottom phase proportion and descent rate. For each individual, the baseline period was used to calculate the average of each metric across the entire baseline period (a fixed reference window approach), as well the covariance within and between each of the dive metrics. Using Equation 2.1 (Chapter 2), the Mahalanobis distance was calculated between the mean values of dive metrics in the baseline period and the mean values of dive metrics in a moving time window. Each time window contained three dives and moved along the dataset one dive at a time. Comparisons were made within each individual only. This selected analysis approach was based on the dive simulation findings of Chapter 2; in particular, a time window of three dives was chosen to maximise the performance of detecting both singular unusual dives, and sequences of several unusual dives. To ensure that within each time window dives that occurred either side of a data gap (e.g. due to haulouts, dives which occurred >100 km from the wind farm) were not grouped together, if there was more than 15 minutes between the end of the first or sec-

ond dive (including the post-dive surface interval) and the start of the next recorded dive, the Mahalanobis distance for that three-dive window was not calculated. It would not be biologically meaningful to calculate the average dive behaviour over windows containing data gaps (potentially consisting of several hours or days between recorded dives), and it would not be comparable with the remainder of the time windows (the majority of which were <1 hour in duration). These calculations were implemented using a custom-written R function `mdist_dive`.

Identifying unusual dive behaviour by randomisation tests

Mahalanobis distance provided a single number representing how much each three-dive window differed from the average baseline behaviour. Higher values of Mahalanobis distance represented dives that were unusual, and so this metric could be used to examine when individuals exhibited the extremes of their dive behaviour. To identify periods of unusual behaviour during pile driving, Mahalanobis distance values observed during piling were compared to those observed during baseline periods by a randomisation test. Each seal was considered to have an "encounter" with pile driving when they had an estimated received level of pile driving sound above ambient sound level (see next Section) during a pile driving bout. For each encounter, 1000 "mock exposure" encounters were sampled randomly from the baseline period for that seal, each of the same length as the pile driving bout. The maximum Mahalanobis distance $D_{\max}$ observed in each mock exposure period was then compared to the Mahalanobis distance values observed in the true exposure. The 95^{th} quantile of all the $D_{\max}$ values observed in the mock exposures was used as a threshold D_{th}, above which the seal's behaviour was considered to be significantly unusual. If the $D_{\max}$ observed in the true exposure exceeded D_{th}, it was considered to be statistically unusual and therefore a possible response to the pile driving activity.

Estimating received piling sound doses

For each encounter between seals and pile driving, the acoustic propagation model was used to estimate the received acoustic "dose" by each seal. For encounters in which the randomisation test identified a significantly unusual dive behaviour (a response), the first three-dive time window with a Mahalanobis distance exceeding D_{th} was considered to be the time interval in which the seal changed its behaviour (the "initial response interval"). For each piling strike that occurred within the initial response interval, seal tracks were

linearly interpolated between GPS locations to estimate the horizontal location of the seal at the time of each piling strike. For encounters in which seals were not identified to respond, seal tracks were linearly interpolated to estimate the horizontal location of the seal at the time of all piling strikes within the piling bout.

Seal horizontal locations were matched to the predictions from the acoustic propagation model, and scaled by the blow energy of each piling strike to obtain final estimates of the received SEL_{ss} at each seal location. This process is described in detail in Chapter 3 (Section 3.3.4).

For all encounters, the estimated acoustic dose was therefore an estimate of the average dose available (median across all depths) at the horizontal location of each dive. In general, estimated piling sound levels were lower nearer the surface of the water and higher at depth. If a seal responded to piling sound by diving to more shallow depths or staying near to the surface, this could be in response to the high levels of piling sound available across the water column, not only to the low levels experienced in shallow depths. The acoustic doses used and modelled here were therefore estimates of the median sound level available at the seal's location; this approach was also chosen to improve applicability of the results to the spatial predictions of piling noise used in environmental impact assessments (Thompson et al., 2013).

A Bayesian hierarchical model of response thresholds

A Bayesian hierarchical model was used to estimate the relationship between the piling sound level and the probability of a change in dive behaviour, using information from all piling encounters (both responses and non-responses). The model used was the same as that described in detail in Chapter 3 (Section 3.3.5). To recap the relevant notation and pertinent aspects of the model, a brief summary is provided here.

The Bayesian hierarchical model assumes that each individual i has a mean response threshold μ_i (in dB re $1\mu\text{Pa}^2 \cdot \text{s}$),

$$\mu_i \sim \text{N}(\mu + \boldsymbol{z_i}^\intercal \boldsymbol{\alpha}, \phi^2) \tag{4.1}$$

where μ is the mean population response threshold, $\boldsymbol{z}_i$ is a vector of individual-level covariates, $\boldsymbol{\alpha}$ is a vector of model parameters, and ϕ^2 is the variance between individuals. The true (but unobserved) response threshold $t_{i,j}$ for individual i and encounter j is specified

as

$$t_{i,j} \sim \mathrm{N}(\mu_i + \boldsymbol{z}_{\boldsymbol{i,j}}^{\intercal}\boldsymbol{\beta}, \sigma^2) \tag{4.2}$$

where $\boldsymbol{z_{i,j}}$ is a vector of encounter-level covariates, $\boldsymbol{\beta}$ is a vector of model parameters, and σ^2 is the variance between encounters. It is also assumed that the true response threshold $t_{i,j}$ is observed with error,

$$y_{i,j} \sim \mathrm{N}(t_{i,j}, \epsilon^2) \tag{4.3}$$

where $y_{i,j}$ is the estimated median received SELss from piling and ϵ is the standard deviation of all received piling blows. Non-response encounters are considered to be right-censored, whereby the model assumes that the true response threshold $t_{i,j}$ is greater than the maximum observed dose of piling energy $y_{i,j}^{c}$.

Prior distributions were the same as those specified in Chapter 3, with a uniform prior with a lower limit of 115 dB re $1\mu\mathrm{Pa}^2 \cdot \mathrm{s}$ and an upper limit of 200 dB re $1\mu\mathrm{Pa}^2 \cdot \mathrm{s}$ for the population mean threshold μ, and uniform priors of 0–30 dB re $1\mu\mathrm{Pa}^2 \cdot \mathrm{s}$ for the standard deviations ϕ and σ. Model fitting and checks are described in detail in Chapter 3; in this Chapter models were also fit in JAGS and `rjags` (version 4.3.0; Plummer, 2003; Plummer, 2019) using the same starting values and by running 4 chains of 100,000 iterations each.

Including covariates on estimated response threshold

As described in detail in Chapter 3 (Section 3.3.5), a suite of potential covariates were considered in the dose-response model: sex, age class, daylight, exposure history, behavioural state, piling bout length, and initial distance from piling (Table 3.2). A maximal model with all covariates and a simple model with no covariates were both fitted. In the maximal model, covariates were included at both the individual (Equation 4.4) and encounter level (Equation 4.5):

$$\mu_i \sim \mathrm{N}(\mu + \alpha_{\mathrm{sex}}\mathrm{S}_i + \alpha_{\mathrm{age}}\mathrm{A}_i, \phi^2) \tag{4.4}$$

$$t_{i,j} \sim \mathrm{N}(\mu_i + \beta_{\mathrm{day}}\mathrm{D}_{i,j} + \boldsymbol{\beta}_{\mathrm{exp}}\mathrm{E}_{i,j} + \boldsymbol{\beta}_{\mathrm{behav}}\mathrm{B}_{i,j} + \beta_{\mathrm{lengthbout}}\mathrm{L}_{i,j} + \beta_{\mathrm{initdist}}\mathrm{R}_{i,j}, \sigma^2) \tag{4.5}$$

Here, each α term is a model parameter for each individual-level covariate, and S_i and A_i

are, respectively, the sex and age class of each individual i. Each β term is a model parameter for each encounter-level covariate, and $D_{i,j}$, $E_{i,j}$, $B_{i,j}$, $L_{i,j}$ and $R_{i,j}$ are, respectively, the daylight, exposure history, behavioural state, piling bout length and initial distance for each encounter j (Table 3.2). Gibbs variable selection (O'Hara and Sillanpää, 2009) was used to assess the level of support for including each of these covariates within the final dose-response model, as described in Chapter 3 (Section 3.3.5).

Estimating the dose-response function

As described in detail in Chapter 3 (Section 3.3.5), posterior samples from the Bayesian hierarchical model were used to estimate a mean dose-response function and associated 95% credible intervals. The dose at which the probability of response is 0.5, p_{50}, was also calculated. For the maximal model containing covariates, the posterior samples were used to estimate a series of potential dose-response functions to illustrate the effect of each covariate, assuming that all other covariates were present in the maximal model.

4.3.5 Behaviour-switching during pile driving: using hidden Markov models (HMMs) to model different types of dive behaviour

A HMM of dive types

To investigate the occurrence of different dive types for harbour seals in the vicinity of the wind farm, and to evaluate the effect of pile driving on the probability of seals switching between these behaviours, hidden Markov models (HMMs; Zucchini et al. (2016)) were also fit to the seal dive data. HMMs can be used to classify different types of animal movement behaviour and model the transitions between them (Langrock et al., 2012), allowing inference on the underlying behaviours driving the data observations. As HMMs model data in discrete time, they are often fit to regularly spaced observations in time. For air-breathing marine animals, each dive can be considered to be a biological time unit of behaviour, with animals required to return to the surface at the end of each dive; HMMs can therefore be used to model behaviour switching between individual dives (DeRuiter et al., 2017; Hart et al., 2010; Quick et al., 2017).

Here, three types of seal dive behaviour were considered following initial data exploration. Preliminary models were fit with three, four, or five states using some or all of the five dive metrics calculated (see Section 4.3.2); however, models fit with greater numbers of states

did not converge for some individuals, and did not consistently identify the same states, limiting comparability of behaviour and behavioural effects across individuals. In order to fit models that were stable, biologically interpretable and comparable across individuals (see Discussion and Pohle et al., 2017), the three-state model emerged as the most practical option to model broad changes in dive behaviour among seals. These were modelled as three latent behavioural states $Z_t \in \{1, 2, 3\}$ where Z_t is the behavioural state for dive t in a sequence of dives $t = 1, \ldots, T$. Across all dives, the probability of a seal switching behaviour can be described by a transition probability matrix $\mathbf{\Gamma}$,

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} \\ \gamma_{2,1} & \gamma_{2,2} & \gamma_{2,3} \\ \gamma_{3,1} & \gamma_{3,2} & \gamma_{3,3} \end{pmatrix}$$

where $\gamma_{i,j} = \Pr(Z_{t+1} = j \mid Z_t = i)$ and $\gamma_{i,j}$ is the probability of a seal switching to state j at dive $t + 1$, given they are currently in state i at dive t.

Three metrics were chosen as informative observations of the underlying dive state for each dive t: dive duration D_t, surface duration S_t, and proportion of the dive in the bottom phase B_t (as described in Section 4.3.2). Measurements of dive depth were not used as dive depth is strongly correlated with available water depth, and would not be an informative metric in identifying the same behaviour in different locations, particularly within the shallow study area (<100 m water depth). Dive duration D_t and surface duration S_t were assumed to be Gamma distributed as they were continuous and positive values,

$$D_t \mid Z_t = i \sim \text{Gamma}(\mu_{d,i}, \sigma_{d,i})$$

$$S_t \mid Z_t = i \sim \text{Gamma}(\mu_{s,i}, \sigma_{s,i})$$

where the state-dependent distributions were defined by the mean dive duration $\mu_{d,i}$, standard deviation of dive duration $\sigma_{d,i}$, mean surface duration $\mu_{s,i}$, and standard deviation of surface duration $\sigma_{s,i}$ for each dive state i. The proportion of the dive in the bottom phase B_t was assumed to follow a Beta distribution as it was constrained between zero and one,

$$B_t \mid Z_t = i \sim \text{Beta}(\alpha_i, \beta_i, \epsilon_i)$$

where α_i and β_i are the first and second shape parameters for each state i. A point mass

parameter ϵ_i was also specified for each state i as the data contained some values of 1 which are not normally allowed under a Beta distribution. Values of 1 generally occurred during very shallow dives where all measured depth points are equal (or close to) the maximum depth of the dive.

Dives beyond 100 km from the centre location of pile driving were excluded from analyses. This ensured that only dive behaviour within the vicinity of the wind farm was considered, reducing any potentially different dives (e.g. very deep dives only seen far offshore) which could obscure comparisons of dive behaviour between baseline and times of piling sound exposure. Time gaps were present in the data due to dives observed beyond 100 km from the wind farm and dives observed before and after haulouts. Any time gaps of more than 15 minutes between dives (following the end of the post-dive surface interval) were input as missing observations, to ensure that transition probabilities were not directly calculated for dives that did not immediately succeed the previous dive. Dive behaviour was observed to have high variability between individuals, and individuals may respond differently to pile driving sounds. All HMMs were therefore fit to each individual separately.

Including covariates on the dive transition probabilities

Different factors may affect the probability of a seal changing its dive behaviour. Dive behaviour is likely to be strongly related to the habitat the seal is in, either because the animal chose to travel to (or through) a particular habitat to exhibit that behaviour, or because they opportunistically are in in a habitat that is suitable for a particular type of dive behaviour. Thus, to control for these natural changes in dive behaviour as the seal moves horizontally through its environment, bathymetry was included as a covariate affecting the transition probabilities between the different dive states (as a proxy for habitat). Bathymetry was calculated as the available water depth (in metres) at the starting time and location of each dive, accounting for changes in water depth due to the tidal cycle (Section 4.3.2).

To test whether pile driving affected the probability of switching between dive types, the model described above was also fit with pile driving included as a factor covariate (1 or 0) on the transition probabilities. Preliminary models containing piling sound level as a continuous covariate did not contain enough information to estimate a relationship between each of the transition probabilities and received sound level. This is because, as HMMs were fitted to each individual and piling was included as a covariate on all six transition rates between the states, there was a high number of parameters to estimate and relatively

limited variation in some of the received sound levels. The acoustic propagation model was used to estimate the average sound level available in the water column at the starting time and location of each dive. For dives which occurred during pile driving, dive locations were matched to the corresponding spatial grid cell in the acoustic model (for the current piling location), and the median received $SEL_{ss,\,ref}$ across all depth layers was identified. This reference sound level was then scaled by the median blow energy of pile driving strikes during each dive, using Equation 3.1, to estimate the median sound level during each dive. Dives in which the median sound level available during the dive was greater than 118 dB re $1\mu Pa^2 \cdot s$ (estimated median ambient sound level; Nedwell et al., 2011) were considered to be during exposure to pile driving sounds and so were set to 1. Dives which occurred outwith piling times, or in which the estimated median sound level received from pile driving was less than 118 dB re $1\mu Pa^2 \cdot s$, were set to 0.

Covariates were included by a multinomial logit link on the between-state transition probabilities,

$$\gamma_{i,j,t} = \frac{\exp(\beta_{ij0} + \beta_{ijb} b_t + \beta_{ijp} p_t)}{1 + \sum_{l \neq i} \exp(\beta_{il0} + \beta_{ilb} b_t + \beta_{ilp} p_t)}$$

Here, $\gamma_{i,j,t}$ is the probability of switching from dive type i to j at dive t (for $i \neq j$), where b_t is the value of bathymetry at dive t, and $p_t = 1$ if the seal was exposed to pile driving during dive t and $p_t = 0$ otherwise. The β parameters are model parameters for the intercept β_{ij0}, effect of bathymetry β_{ijb}, and effect of pile driving β_{ijp}.

Implementation and model checking

To test the hypothesis that transitions between seal dive behaviours are affected by pile driving, two HMMs, each with three behavioural states, were constructed and fitted for each seal. In the first model (reduced model), bathymetry was included as a covariate influencing all of the transition probabilities between the states. In the second model (full model), both bathymetry and pile driving activity (a factor variable) were included as covariates influencing the transition probabilities. A range of starting values for the three states were tested to ensure that the resulting state distributions were not influenced by these values. During preliminary model fitting, using the nlm optimiser sometimes led to the model becoming stuck in local optima. For each model, an iterative approach was therefore taken to ensure that models converged on the global maximum likelihood estimate. Each model was first fit using the Nelder-Mead optimisation method for the first 500 iterations to get broad estimates for what the parameter space should be. The working

estimates from this model were then used to refit the model using the nlm optimiser to achieve fast convergence. All models were fit using the R package `momentuHMM` (McClintock and Michelot, 2018).

Models were assessed by visual inspection of the dives classified within each dive type, as well as checking of Q-Q plots and pseudo-residuals. For each seal, models with and without pile driving were compared by AIC to determine whether pile driving had an effect on switching between different dive types. If the piling model AIC was lower, and the model without pile driving had $\Delta\text{AIC} > 2$ (Burnham and Anderson, 2002), this was considered to be evidence that pile driving should be included in the model for that seal. The Viterbi algorithm was used to estimate the most likely sequence of dive types from the best-fitting model.

4.3.6 Software used

All analyses were conducted in R (R Core Team, 2020), with additional packages `data.table` (Dowle and Srinivasan, 2021), `dplyr` (Wickham et al., 2020), `lubridate` (Grolemund and Wickham, 2011), `maptools` (Bivand and Lewin-Koh, 2021), `momentuHMM` (McClintock and Michelot, 2018), `raster` (Hijmans, 2020), `rgdal` (Bivand et al., 2021), `rjags` (Plummer, 2019), `sp` (Pebesma and Bivand, 2005), `StatMatch` (D'Orazio, 2019), and `suncalc` (Thieurmel and Elmarhraoui, 2019). The R packages `ggplot2` (Wickham, 2016), `ggpubr` (Kassambara, 2020), `plot.matrix` (Klinke, 2021), `RColorBrewer` (Neuwirth, 2014) and `viridis` (Garnier, 2018) were also used for data visualisation.

4.4 Results

4.4.1 Unusual dive behaviour during pile driving

In total, across the 24 tagged harbour seals, 543,011 seal dives were recorded during the study period. Of these, 325,636 were categorised as occurring within baseline periods (Figure 4.2). Across all seals, the median duration of baseline data available was approximately 14,400 dives (min = 5,000; max = 24,600; Figure 4.2(a)). In general, $\sim 60\%$ of recorded dives were included in the baseline dataset for each seal (Figure 4.2(b)).

There were 232 encounters between diving harbour seals and pile driving, i.e. when it was estimated that they had received levels of piling noise above ambient sound level. Out of

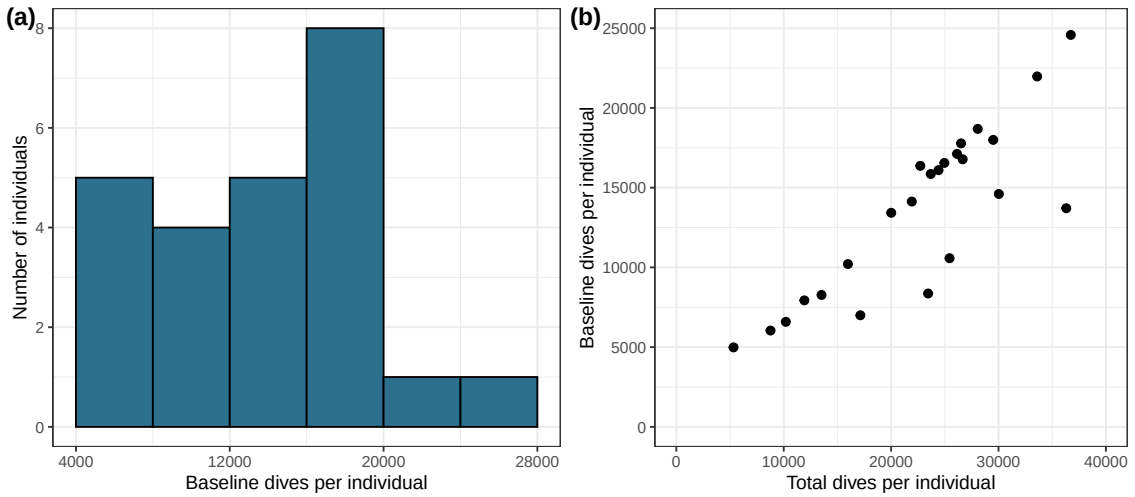


Figure 4.2 Duration of dive data available for each tagged seal. (a) Number of dives per individual during identified baseline periods. (b) Number of dives per individual during baseline periods vs. the total number of dives recorded for each individual.

the 24 seals tagged, 20 seals experienced encounters with pile driving and 18 seals had more than one encounter each (median = 6). From the Mahalanobis distance randomisation test, 16 encounters were identified as containing unusual dive behaviour (Table 4.1). The initial distance of seals from pile driving during these 16 encounters ranged from 10.9 to 32.8 km, and the estimated received sound exposure level (SELss) of pile driving at the time when the unusual dive behaviour was identified ranged from 124 to 160 dB re $1\mu\text{Pa}^2 \cdot \text{s}$ (Table 4.1). Nine individuals were identified as exhibiting unusual dive behaviour during pile driving; four individuals were identified more than once. In 12 of the pile driving encounters, the seal dive behaviour observed was more unusual than the 99^{th} quantile of the baseline distribution ($p<0.01$; Table 4.1); the significance level used to define unusual dives was the 95^{th} quantile.

Two examples of detected dive responses, selected at random, are presented here in more detail. During one of the identified encounters, seal pv42-162-12 was making regular U-shaped dives to ∼20 metres depth prior to pile driving (Figure 4.3). Pile driving at site LS57 began at 14:11 when the seal was approximately 20.5 km away. Over the following 20–30 minutes, the seal made one more U-shaped dive, followed by one shallower V-shaped dive and several dives less than 5 metres depth. During this sequence of dives, there is a gradual decrease in dive duration, proportion of the available water column used and descent rate. On the 7^{th} dive after pile driving began, the seal exhibited an extended surface duration. The seal did not dive again until 15:50, 52 minutes after pile driving

Table 4.1 Summary of the identified dive responses by Mahalanobis distance. For each identified response time interval, the median SELss across all depths was calculated for each received piling hammer blow, and the median $y_{i,j}$ and standard deviation ϵ of these piling blows was used in the dose-response model (Equation 4.3). Shown here also are the initial distance between the seal and pile driving, the estimated maximum SELss the individuals may have received had they always been present in the quietest or loudest parts of the water column, and the percentile of the Mahalanobis distance values corresponding to each response. All measurements of SELss are in dB re $1\mu\text{Pa}^2 \cdot \text{s}$.

Seal ID	**Pile bout ID**	**Median (±SD) SELss across depth & time** $y_{i,j}$ $(\pm\epsilon)$	**Max SELss at quietest depth**	**Max SELss at loudest depth**	**Initial distance (km)**	**Percent of MD values**
pv42-162-12	LS21.1	135 (±2.3)	133	138	17.1	0.000
pv42-162-12	LS55.1	134 (±1.7)	128	138	22.1	0.007
pv42-162-12	LS57.2	139 (±1.8)	134	141	20.9	0.012
pv42-277-12	LS73.2	137 (±0.9)	134	140	24.3	0.001
pv42-290-12	LS21.2	126 (±1.6)	118	128	24.6	0.003
pv42-290-12	LS75.2	153 (±0.5)	144	154	17.1	0.020
pv42-291-12	LS75.2	153 (±0.2)	145	155	17.4	0.001
pv42-293-12	LS21.2	126 (±1.2)	122	128	32.8	0.001
pv42-295-12	LS55.2	158 (±8.3)	153	161	16.3	0.017
pv42-317-12	LS58.1	125 (±2.8)	120	127	32.2	0.000
pv42-318-12	LS21.2	124 (±1.4)	118	127	22.7	0.005
pv42-318-12	LS75.2	147 (±0.7)	136	150	15.1	0.041
pv42-318-12	LS73.1	135 (±1.4)	124	138	13.8	0.006
pv40-268-12	LS59.2	160 (±0.9)	154	162	11.4	0.007
pv40-268-12	LS57.2	151 (±0.1)	140	152	10.9	0.000
pv40-268-12	LS34.2	155 (±1.4)	145	157	11.9	0.005

ceased. Throughout the pile driving encounter, the seal moved away from the pile driving location.

During another encounter, seal pv40-268-12 was alternating between short (∼1 minute) and shallow (∼7 metre) dives, and U-shaped dives of ∼4 minutes and ∼30 metres depth (Figure 4.4). Pile driving at site LS34 began at 18:53, when the seal was approximately 12 km away and was in a shallow dive. After this dive, the seal exhibited a slightly longer surface interval of ∼5 minutes. It then made a V-shaped dive to 31 metres depth, followed by a surface interval of 45 minutes. Two short and shallow dives were made at 19:45, before pile driving had ended. Five V-shaped dives with moderate surface post-dive surface durations (median=9 minutes) were recorded in the hour after pile driving ceased. Throughout the encounter, the seal was travelling away from the pile driving location.

Across all identified dive responses (Figure 4.5; Table 4.1), the dives observed during responses to pile driving were generally shorter than those observed across baseline periods (median of 0.7 minutes vs 3 minutes). Seal dives recorded during responses typically used less of the available water column (median of 0.32 vs. 0.95), and had longer post-dive surface intervals (median of 7.2 vs. 0.7 minutes). Dives during responses also had lower descent rates (median of 0.25 vs. 0.45 m s^{-1}) and a lower proportion of the dive was spent in the bottom phase (median of 0.33 vs. 0.74). Plots of dive metrics recorded for each individual are included in Appendix D (Figures S29, S30, S31).

4.4.2 The dose-response relationship

Both the full maximal model and the reduced model showed rapid convergence of the MCMC chains, with a Gelman-Rubin of 1.0. Gibbs variable selection indicated a low level of support for including the majority of the covariates (sex, age class, exposure history, behavioural state, bout length, initial distance) within the dose-response model, estimating probabilities of 22–45% that each of these covariates should be included in the true model (Table 4.2). The only covariate that was estimated to have moderate support was daylight (Gibbs p of 0.73), with estimates indicating a higher probability of response during the night (Figure S34(c)). Illustrative dose-response relationships of all considered covariates are included in Appendix D (Figure S34).

As overall there was not enough evidence to support inclusion of covariates in the model, the reduced model (no covariates) was chosen to produce the final dose-response function (Figure 4.6, Table S2). Using this model, the mean estimated population response threshold

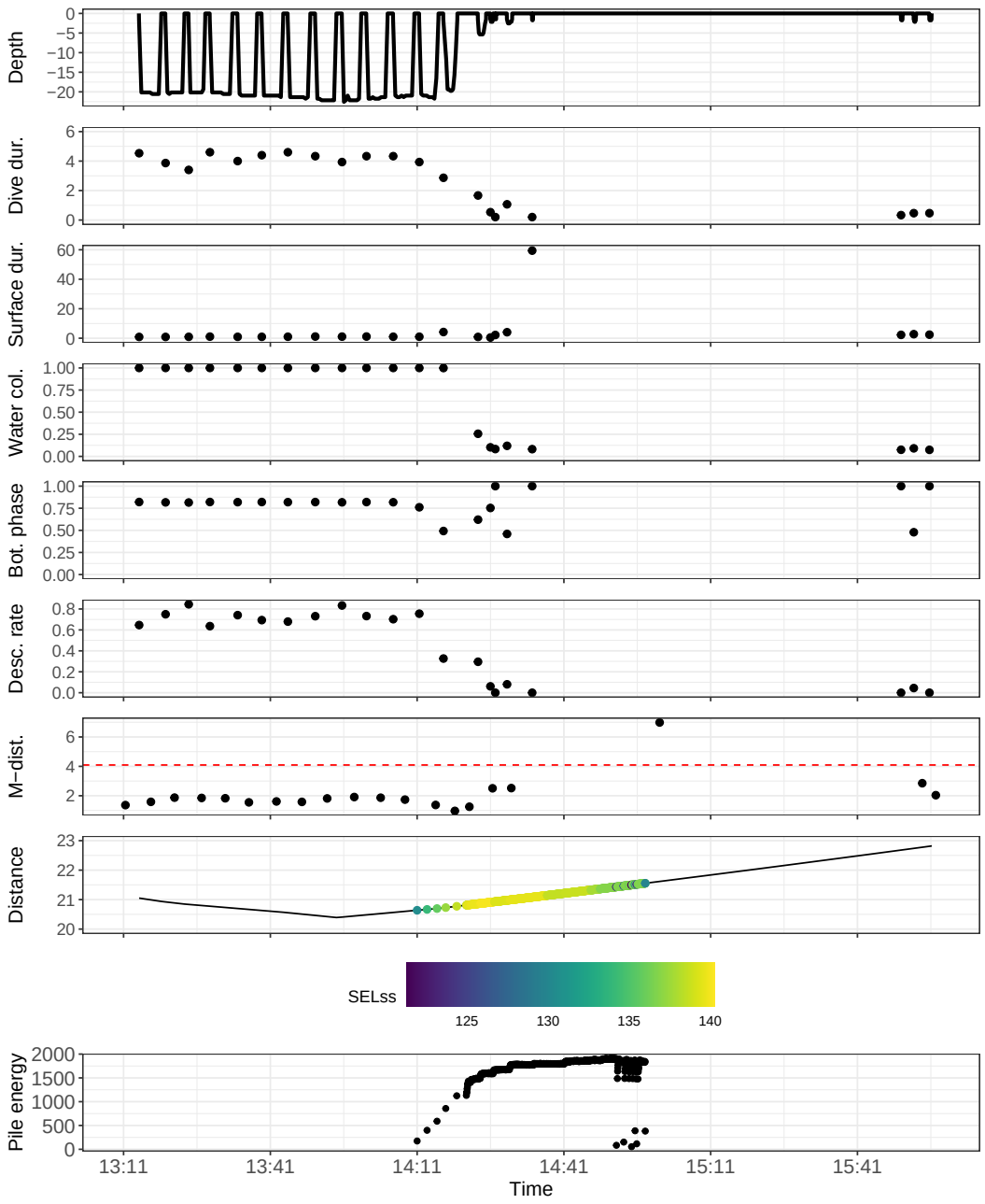


Figure 4.3 Example of an identified behavioural response from seal number pv42-162-12 at pile location LS57 (on 17$^{\text{th}}$ April 2012). The top six panels show metrics of dives observed from one hour pre-piling to one hour post-piling: depth profile (m), duration of dive (mins), duration of post-dive surface interval (mins), proportion of water column used, proportion of dive within estimated bottom phase, and descent rate (m/s). Dive metrics are plotted at the start time of the dive they correspond to. The "M-dist" panel shows the calculated Mahalanobis distance for each group of 3 dives; the red dashed line corresponds to the threshold used to identify a response. Horizontal distance (km) between the seal's location and pile driving, and the corresponding estimated received levels of piling sound (colourscale, SELss, dB re $1\mu\text{Pa}^2 \cdot \text{s}$) for each piling stike (energy in kJ) are also shown in the bottom two panels.

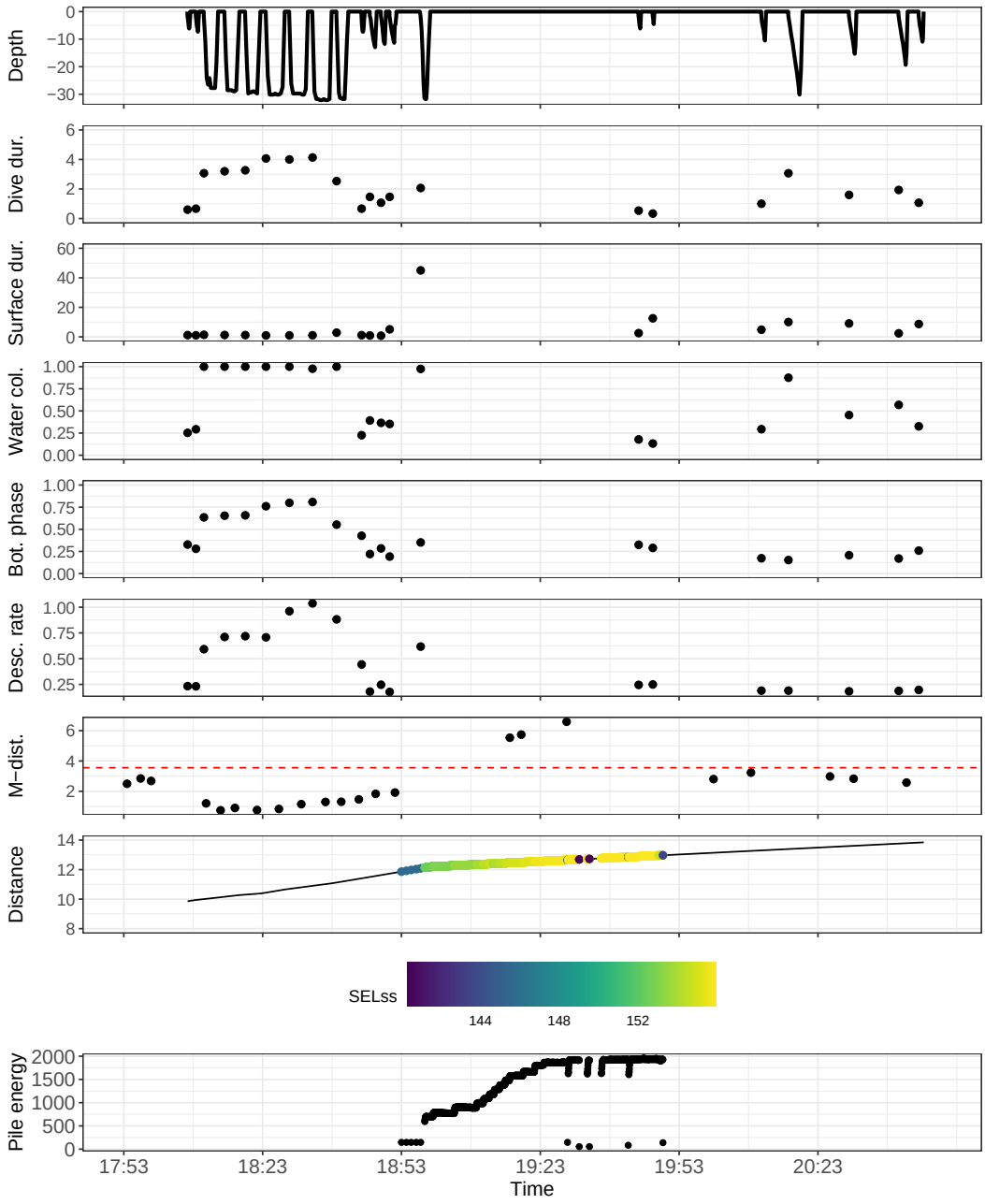


Figure 4.4 Example of an identified behavioural response from seal number pv40-268-12 at pile location LS34 (on 8th May 2012). The top six panels show metrics of dives observed from one hour pre-piling to one hour post-piling: depth profile (m), duration of dive (mins), duration of post-dive surface interval (mins), proportion of water column used, proportion of dive within estimated bottom phase, and descent rate (m/s). Dive metrics are plotted at the start time of the dive they correspond to. The "M-dist" panel shows the calculated Mahalanobis distance for each group of 3 dives; the red dashed line corresponds to the threshold used to identify a response. Horizontal distance (km) between the seal's location and pile driving, and the corresponding estimated received levels of piling sound (colourscale, SELss, dB re $1\mu\text{Pa}^2 \cdot \text{s}$) for each piling stike (energy in kJ) are also shown in the bottom two panels.

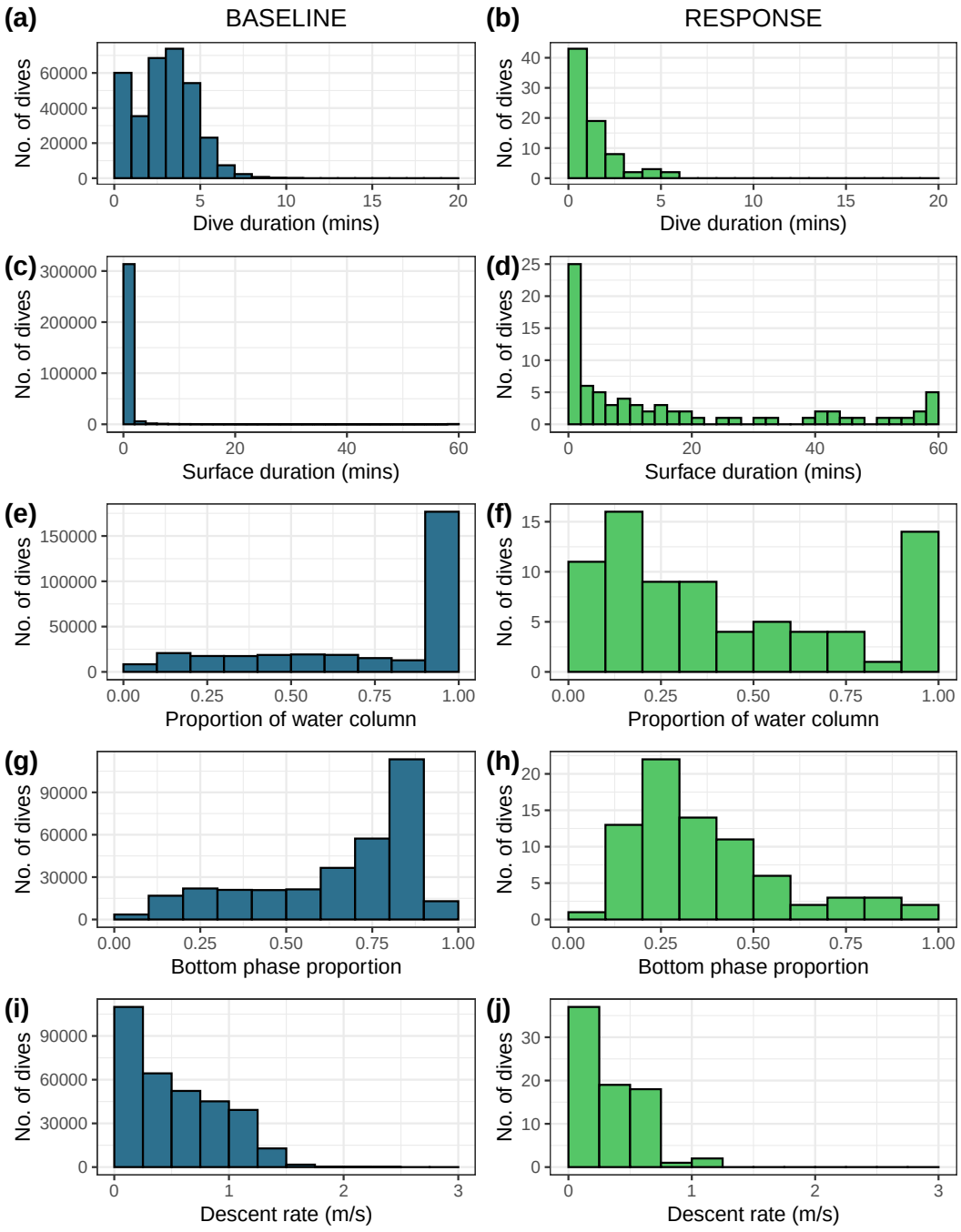


Figure 4.5 Dive metrics for all dives within baseline periods (blue; left column) and dives during identified behavioural responses to pile driving (green; right column). Shown are histograms of the duration of each dive, the duration of the post-dive surface interval, the proportion of the available water column used, the proportion of the dive duration identified as the bottom phase, and the descent rate. Values are pooled across all tagged seals.

Table 4.2 Estimated model parameters from posterior samples of the maximal model, and the final model with no covariates. Also shown are the results of Gibbs variable selection (GVS), where the level of support for the covariate is the proportion of MCMC samples in which the candidate covariate (and associated α and β terms) is included. Factor covariates with three levels required two β parameters to estimate; Gibbs support was estimated for inclusion of the factor variable as a whole.

		Maximal model			Final model	
		Mean	**SD**	**GVS p**	**Mean**	**SD**
	μ	192	7.0	-	188	7.5
	ϕ	12	7.0	-	16	6.5
	σ	24	3.8	-	25	3.2
Sex:	α_{sex}	3.4	7.2	0.45	-	-
Age class:	α_{age}	1.7	7.9	0.44	-	-
Daylight:	β_{day}	6.7	4.6	0.73	-	-
Exposure history:	β_{exp1}	-1.8	5.5	0.24	-	-
	β_{exp2}	0.4	5.3		-	-
Behavioural state:	β_{behav1}	3.2	5.4	0.42	-	-
	β_{behav2}	5.7	5.7		-	-
Bout length:	$\beta_{\text{lengthbout}}$	5.3	6.4	0.41	-	-
Initial distance:	β_{initdist}	-0.7	0.5	0.22	-	-

μ was 188 (95% CI: 172–199) dB re $1\mu\text{Pa}^2\cdot\text{s}$, with a standard deviation between individuals ϕ of 16 dB re $1\mu\text{Pa}^2\cdot\text{s}$ and between encounters σ of 25 dB re $1\mu\text{Pa}^2\cdot\text{s}$ (Table 4.2). The estimated p_{50} for the final model was 175 (95% CI: 167–180) dB re $1\mu\text{Pa}^2\cdot\text{s}$. Posterior distributions from both of the considered models are presented in the Appendix D (Figures S32; S33).

4.4.3 Behaviour-switching during pile driving

Model selection and behavioural states

Twenty of the tagged seals were recorded diving during exposure to pile driving sounds, and so the full model (pile driving and bathymetry) and reduced model (bathymetry only) HMMs were fit to each of these individuals. Pile driving was retained as an effect within the model for ten of the twenty seals, following comparison by AIC (Table 4.3). Five of the seals which retained pile driving as a covariate had also been identified to exhibit behavioural responses in the Mahalanobis distance analysis (Table 4.3). All of the models converged; however, there was evidence of residual autocorrelation in the observed dive

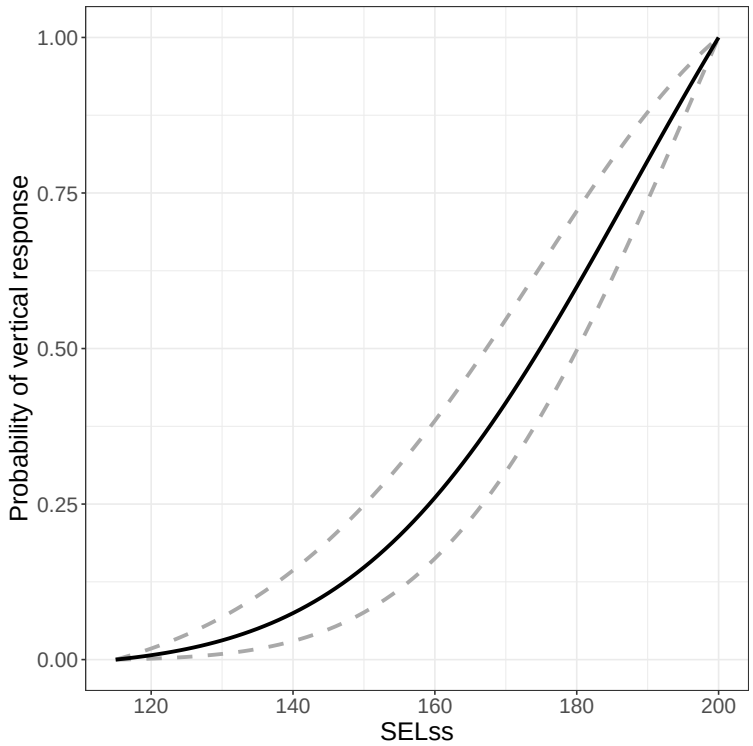


Figure 4.6 Estimated dose-response relationship for vertical responses in harbour seals as a function of single-strike sound exposure level (SELss in dB re $1\mu\text{Pa}^2 \cdot \text{s}$, median across depth) of pile driving. Shown are the mean estimated function (black) and the 95% credible intervals (grey dashed).

Table 4.3 Summary of HMM results for each of the 20 seals which encountered pile driving. Individuals for which pile driving was retained as a covariate on the HMM transition probabilities are highlighted in grey, along with the difference in AIC (ΔAIC) between the full model (piling and bathymetry) and the reduced model (bathymetry only). Individuals which were borderline (ΔAIC < 2) on retaining pile driving in the model are highlighted by asterisks. Also shown are the individuals in which dive behavioural responses were identified by the Mahalanobis distance analysis.

Seal ID	Piling retained in HMM by AIC	ΔAIC	M-Distance Dive Response
pv42-162-12	-	-	Y
pv42-165-12	Y	8.2	-
pv42-194-12	-	-	-
pv42-221-12	Y	25.8	-
pv42-266-12	-	-	-
pv42-277-12	Y	41.6	Y
pv42-288-12	-	-	-
pv42-289-12	Y	4.5	-
pv42-290-12	Y	10.1	Y
pv42-291-12	Y	701.6	Y
pv42-292-12	-	-	-
pv42-293-12	-	-	Y
pv42-294-12	-	-	-
pv42-295-12	Y	16947.56	Y
pv42-316-12	Y	9.6	-
pv42-317-12	Y	11.1	Y
pv42-318-12	*	(0.4)	Y
pv42-319-12	Y	8.8	-
pv42-320-12	-	-	-
pv40-268-12	*	(0.6)	Y

metrics used (Figure S35), suggesting that dive patterns were not fully captured by the simple three-state model. Given the aim of the study was to assess whether pile driving had an effect on overall behaviour switching, and not on interpreting seal activity budgets, the three-state model was deemed a pragmatic approach for this research question (Pohle et al., 2017; see Discussion).

Across all seals, each of the HMMs fitted identified three distinct behavioural states (Figure 4.7). Variability between individuals was apparent, both in the values corresponding to the different state-dependent distributions and in the abundance of each of the estimated behavioural states (Figures S36, S37, S38). There was, however, high consistency in the

three estimated states across individuals. As dives beyond 100 km from the wind farm were excluded and different durations of data were available for each individual, interpretation of activity budgets from these models is not recommended. State 1 consisted of the longest dives ($\mu_{d,1} = 254$, $\sigma_{d,1} = 50$), with a consistently short post-dive surface duration ($\mu_{s,1} = 48$, $\sigma_{s,1} = 8$). Dives in state 1 had a narrow distribution for proportion of the dive in the bottom phase, centred around ~0.8 ($\alpha_1 = 312945$, $\beta_1 = 68704$, $\epsilon_1 = 5.9 \times 10^{-4}$). State 2 consisted of dives of a moderate dive duration which was more variable in duration ($\mu_{d,2} = 204$, $\sigma_{d,2} = 66$), and surface durations which were short but also more variable ($\mu_{s,2} = 53$, $\sigma_{s,2} = 16$). In state 2, the proportion of the dive in the bottom phase had a wide distribution from 0 to 1, but it was more common for seals to spend more than half of the dive in the bottom phase ($\alpha_2 = 8.1$, $\beta_2 = 4.1$, $\epsilon_2 = 5.4 \times 10^{-3}$). State 3 consisted of dives of a very short duration ($\mu_{d,3} = 78$, $\sigma_{d,3} = 60$), with longer and more variable surface durations ($\mu_{s,3} = 142$, $\sigma_{s,3} = 192$). For state 3, the proportion of the dive in the bottom phase also had a wide distribution from 0 to 1, but it was more common to spend less than half of the dive in the bottom phase ($\alpha_3 = 2.3$, $\beta_3 = 3.1$, $\epsilon_3 = 0.15$). Some of the dives in state 3 were estimated to have 100% of the dive time in the bottom phase, due to the short duration of the dives meaning that the maximum depth (and therefore 85% of the maximum depth, see Figure 2.3) was close to the surface. Values presented in the text are averaged across all 20 individuals for the full model; see the Appendix D Figures S36, S37, S38 for graphical results by individual.

Decoding of the estimated states revealed the different dive shapes associated with each behavioural state (Figure 4.8). Dives in state 1 were U-shaped, with a steep descent, flat bottom, and steep ascent. Dives in state 3 were typically very shallow in depth (<10 m; despite depth not being used in the HMM state estimation), with dive profiles which were either V-shaped or dives in which the seal stayed very close to the surface for the entirety (near to 1.5 m, the threshold at which dives are recorded). State 2 contained the most variation in dive shape, consisting of dives which were either evenly V-shaped, skewed, or undulating in dive profile.

Behaviour switching during baseline

Over all 20 of the full models, there was persistence within each behavioural state during baseline (non-piling) periods. Averaged across all individuals, the estimated probabilities of staying within the same state i on dive $t + 1$ (at the mean bathymetry value of 20 m) were highest in state 1 ($\gamma_{1,1} = 0.86$), moderate in state 3 ($\gamma_{3,3} = 0.83$), and lowest in state

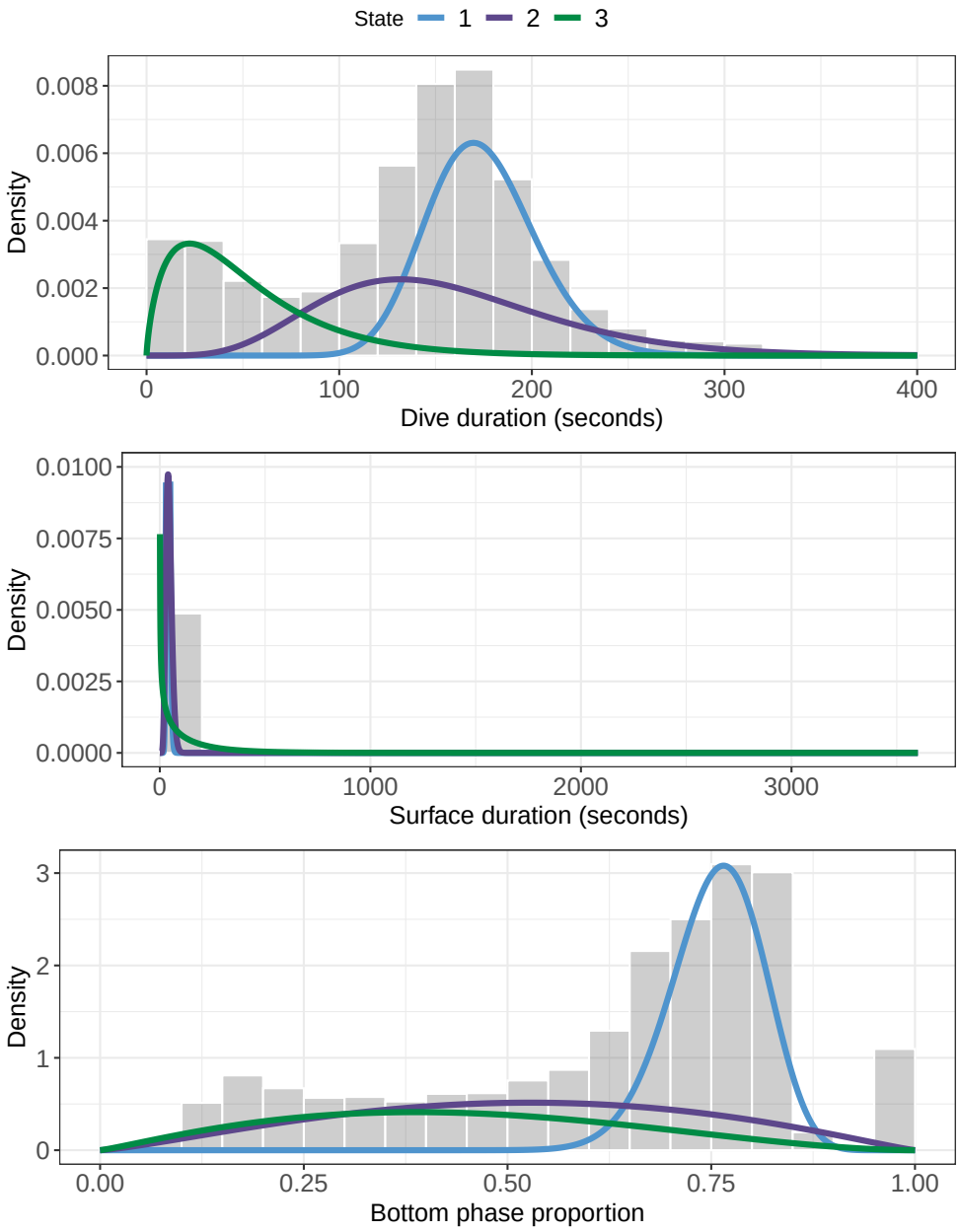


Figure 4.7 Fitted state-dependent distributions of dive duration, surface duration and proportion of the dive in the bottom phase for one example seal (pv42-165-12). Shown are the estimates from a 3-state HMM with bathymetry and pile driving as covariates on the transition probabilities.

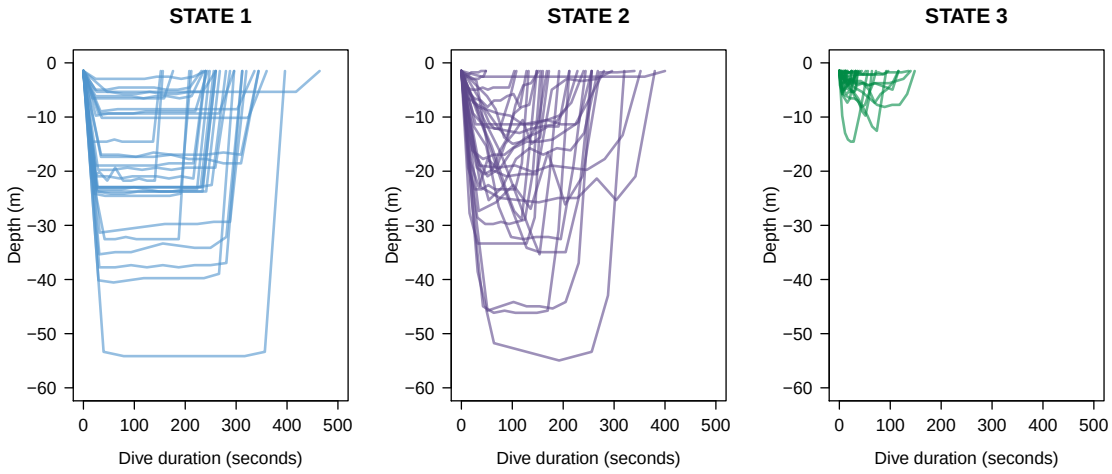


Figure 4.8 Example of dive profiles observed in each of the three estimated behavioural states. States were decoded (using the Viterbi algorithm) from the full HMM containing bathymetry and pile driving for each seal. For each behavioural state, 30 randomly selected dives across all 20 seals are shown.

Table 4.4 Average estimated transition probability matrix across all 20 seals. Probabilities were estimated from the full model, with the covariates of piling=0 and bathymetry=20 m.

	State		
	1	**2**	**3**
1	0.86	0.12	0.03
2	0.13	0.77	0.10
3	0.03	0.15	0.83

2 ($\gamma_{2,2} = 0.77$; Table 4.4). There was limited switching between states 1 and 3 ($\gamma_{1,3} = 0.03$, $\gamma_{3,1} = 0.03$), with seals more likely to switch to state 2 first.

The effect of bathymetry on the transition probabilities differed by individual (Figure S39). Seals travelled to different areas within the study region, and seals which used the same areas did not necessarily use them for the same types of dive behaviour. Notably, some individuals never travelled near to the 100 km distance limit used to examine dive behaviour in the vicinity of the wind farm; others spent a significant proportion of their time near or beyond this limit, excluding some of their behaviours from the analysis. Thus including bathymetry represented an important control for patterns in dive behaviour within each individual.

Effect of pile driving on behaviour switching

Across the 10 seals in which piling was retained in the model, the average probability of remaining in any of the behavioural states was lower overall when seals were exposed to pile driving (Figure 4.9). There was substantial variation in this between individuals, with some seals having increased probabilities of staying in behavioural states or showing limited change. Six of the ten seals showed decreased probabilities of state persistence for all states; eight of the ten seals had decreased probabilities of state persistence for at least two of the states (Figures 4.9, 4.10). There was also high individual variability in the effect of pile driving on the states which seals switch to (Figure 4.10). The only individual which showed increased persistence within all behavioural states during piling (pv42-295-12) was an old female (estimated age of 20.5 years; Lucke et al. (2016)).

4.5 Discussion

In this chapter, data from animal-borne tags were used to track the diving activity of 24 harbour seals during the construction of an offshore wind farm. Twenty of the tagged seals were estimated to have been exposed, on at least one occasion, to pile driving sounds during wind farm construction. Seals were found to exhibit unusual diving behaviour in response to wind farm construction, with nine individuals identified as carrying out significantly unusual dive behaviour during some pile driving bouts. These unusual dives were typically of shorter duration and slower descent rate, with seals using less of the water column, spending less of the dive within the bottom phase, and spending increased time at the surface between dives. The mean response threshold for the population was estimated at piling single-strike sound exposure levels (SELss) of 188 dB re $1\mu Pa^2 \cdot s$, with high variability between (SD=16) and within (SD=25) individuals. There was limited evidence of the effect of individual or encounter-level covariates on the dose-response relationship. Pile driving activity was also found to have a significant effect on behaviour-switching in ten of the tagged seals. In general, most seals showed lower persistence within dive behaviour types (states) and were more likely to switch dive behaviour during piling than during non-piling.

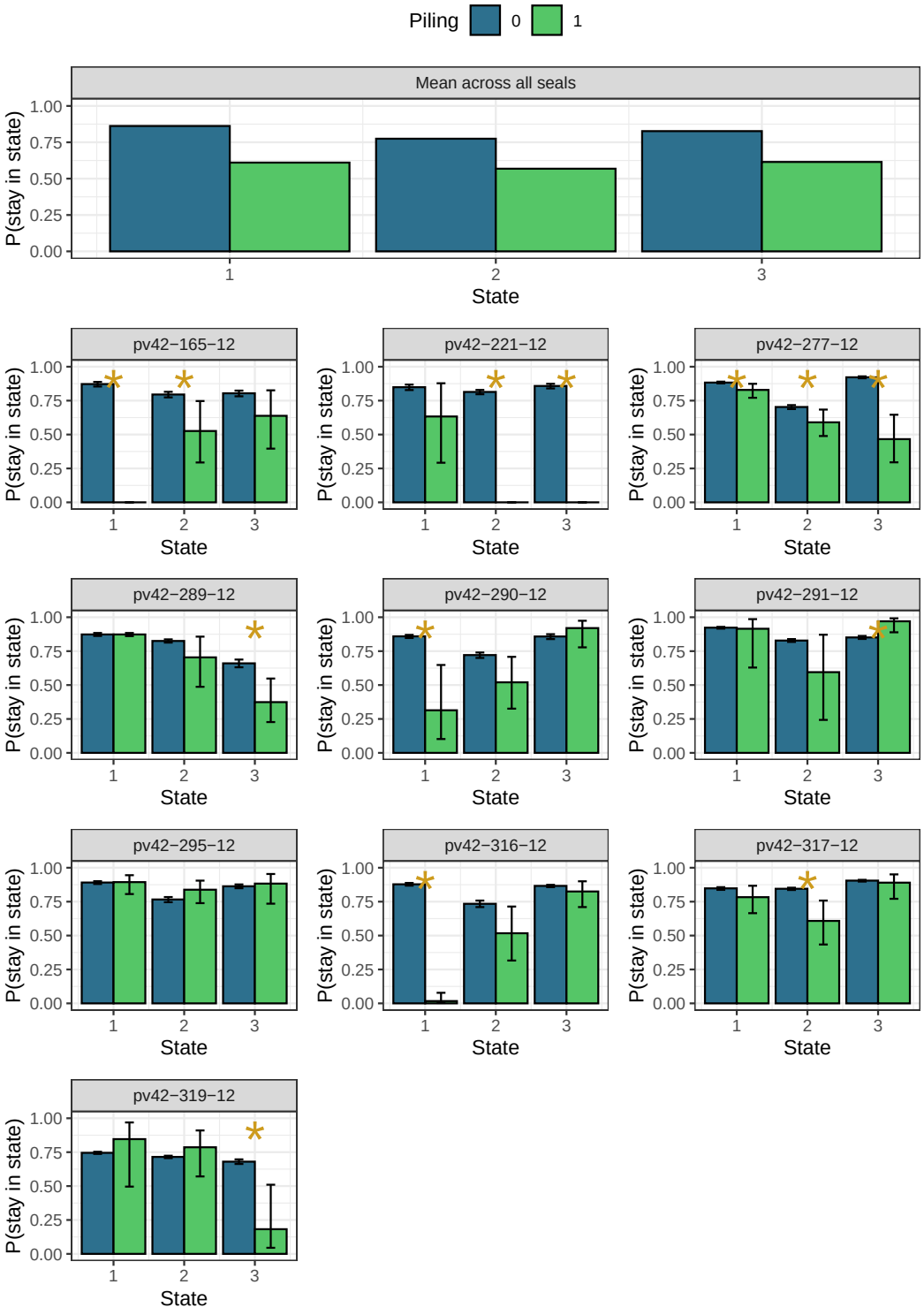


Figure 4.9 Estimated probabilities of staying in a given behavioural state. Probabilities (mean and 95% confidence intervals) were calculated for each of the 10 seals in which pile driving was retained as a covariate in the full model, at the mean bathymetry value of 20 m, for times when piling was inactive (blue) and active (green). Yellow stars denote changes in which the confidence intervals do not overlap. The mean probabilities across all individuals are shown in the top panel.

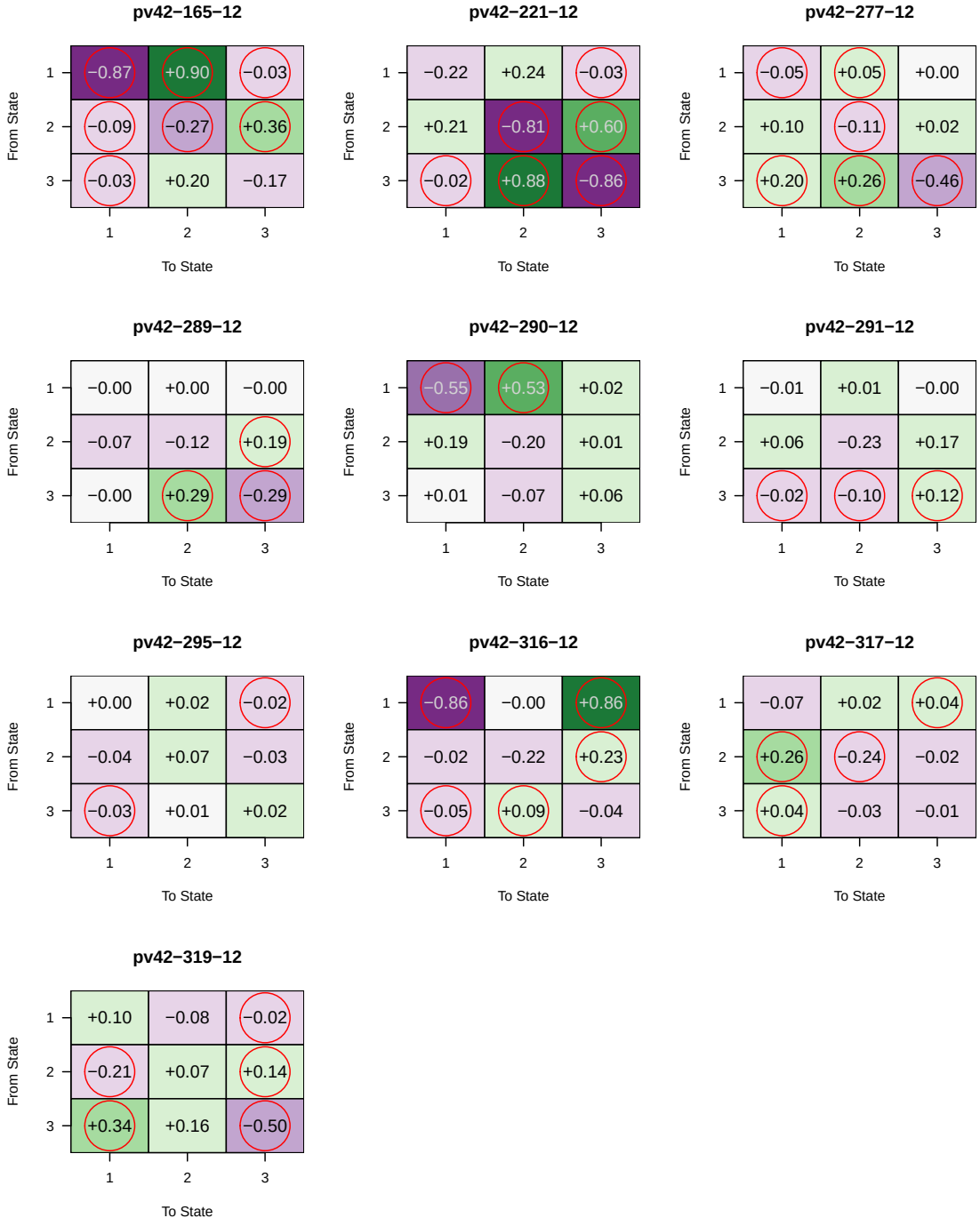


Figure 4.10 The absolute difference in transition probabilities in times of piling exposure, relative to baseline. Probabilities were calculated for each of the 10 seals which retained piling in the three-state HMM, at the mean bathymetry value of 20 m. Red circles denote switches in which the 95% confidence intervals of piling and non-piling probabilities do not overlap. The fill colour denotes the magnitude and direction of the change: increase (green), decrease (purple), or near zero (white). Values of ±0.00 are changes of <0.01.

4.5.1 Unusual dive behaviour during pile driving: Mahalanobis distance analyses

During pile driving, shorter dive durations and longer periods at the surface were observed for some encounters. This is consistent with the findings of several captive studies of sound playback to pinnipeds. Playbacks of pile driving sounds to two captive harbour seals found seals to spend approximately 19–24% of their time at the surface during control periods, and 42–53% of their time at the surface during sound playback periods (Kastelein et al., 2018a). During these playbacks, the seals were often seen swimming actively at the surface, rather than remaining stationary. Similar responses have been observed in captive studies of harbour seals exposed to acoustic deterrent sounds, with seals surfacing more often (Kastelein et al., 2006) and spending increased time swimming with their head above the surface (Kastelein et al., 2017). An increase in jumping behaviour has also been reported (Kastelein et al., 2017, 2018a). Captive studies of other pinniped species during sonar playback have observed increased haulout behaviour in grey seals (Hastie et al., 2014), and rapid swimming at the surface in hooded seals (*Cystophora cristata*) (Kvadsheim et al., 2010).

In this chapter, differences in dive shape were also observed during pile driving, including decreased descent rate, decreased time within the bottom phase of the dive, and decreased use of the water column. In a previous unpublished study, Aarts et al. (2018) recorded the movements and dive behaviour of twenty tagged grey seals during pile driving activity at wind farms in the North Sea (Netherlands). Grey seals were estimated to have significant decreases in dive descent speed, and the most typical dive behaviours included decreases in dive depth, decreases in bottom time (a combination of the proportion of the water column and bottom phase metrics used here), and irregular diving patterns. In Aarts et al. (2018), each metric of dive behaviour was considered separately; in this chapter, I incorporated multiple metrics of dive behaviour together (using Mahalanobis distance), accounting for correlation between different metrics and considering changes in dive behaviour as a whole. This chapter presents important evidence of how wild seals dive during exposure to sound from human activity. The consistency between the responses observed here in harbour seals with the responses observed in grey seals (Aarts et al., 2018) suggests that these behavioural changes may also be observed in other pinniped species and/or in response to other types of sound disturbance in marine environments (e.g. shipping, sonar).

Understanding how seals respond to sound can provide insights into the behavioural motivations for the responses, and there are several possible reasons why seals may respond

in this way. First, diving shallower and shorter, with more time at the surface and less time at the bottom, is likely to directly reduce the level of sound exposure. Sound from sub-surface activities, such as pile driving, is typically louder at depth and quieter towards the surface. Received levels of piling sound at depth may come close to the pain threshold or be unpleasant to experience (e.g. Götz and Janik, 2010). By not diving to the seafloor, seals stay further away from areas of high noise exposure and reduce the risk of temporary or permanent hearing damage (Southall et al., 2019b). Second, coming to the surface more frequently or for longer durations could provide a mechanism for seals to assess the nature, distance, and direction of the potential threat and/or exhibit curiosity towards the sound source. Australian fur seals (*Arctocephalus pusillus doriferus*) exposed to playbacks of vessel noise at a breeding colony were most alert during the playback periods (Tripovich et al., 2012). The majority of responses by fur seals, particularly at low and medium sound levels, consisted of looking or orienting their body towards the sound source. Playback studies of anthropogenic sounds (including white noise) to captive pigs (Talling et al., 1996) and horses (Christensen et al., 2005) also led to increased alertness and attention towards the sound source, and wild polar bears (*Ursus maritimus*) can respond to snowmobile presence by stopping moving and looking (Andersen and Aars, 2008). It is conceivable that coming to the surface to visually assess potential threats, such as oncoming vessels, may be a useful tactic for seals in urbanised environments. Third, in some cases, the reduced diving behaviour observed may be a type of flee response (e.g. such as those seen in beaked whales; Tyack et al., 2011), with seals travelling near the water surface, and perhaps "porpoising" (leaping out of the water; e.g. Yoda et al. (1999)) as they move away from the sound source. In practice, responses may be motivated by a combination of these factors, be short (e.g. 1 dive) or long-term (e.g. several hours), and these proposed functions are likely a small part in a long and complex sequence of behaviour observed before, during and after sound exposure.

In this chapter, dive responses were detected as far as 33 km from pile driving. This is supported by an unpublished report on grey seals and pile driving, which reported changes in dive behaviour up to 36 km away (Aarts et al., 2018). In this chapter, the piling sound level at which a random individual from the population has a 50% chance of exhibiting a dive response (p_{50}) was estimated at a SELss of 175 (95% CI: 167–180) dB re $1\mu\text{Pa}^2\cdot\text{s}$. The high variability observed in the likelihood of a response contributes to the long tail in the dose-response relationship, with low (but above zero) probabilities of response predicted for low–moderate sound levels (Figure 4.6; Table S2). At these lower sound levels, the probability of response may be more closely related to the context in which the exposure

occurred (e.g. current behaviour, hunger and energy levels) than the level of sound received (Ellison et al., 2012). Here, there was not enough evidence to support the inclusion of context-related covariates on the dose-response relationship, for the selection of covariates tested (Figure S34; Table 4.2); however, this is not unexpected as drivers of responses are likely to be multifaceted, change over time (due to habituation or experience), and potentially differ between individuals. The relatively low number of detected responses also provided limited statistical power to examine the effects of additional covariates. This study showed that, in harbour seals, measurable changes in dive behaviour can occur across a wide range of piling sound levels and in a variety of contexts.

Mahalanobis distance randomisation tests were used to compare the average dive behaviour observed in baseline to that observed during pile driving encounters. By this approach, it is assumed that unusual or extreme dive behaviours (above the 95^{th} percentile) observed during pile driving are associated with the disturbance effect of this activity. A more stringent approach would be to assume that responses to pile driving must be more extreme than all behaviours observed during baseline periods (above the 100^{th} percentile). In practice, this stringent approach would assume that a seal's response to pile driving is the most extreme behaviour ever observed over a 2–5 month period. This assumption seems unrealistic, in particular given that seals could potentially be exposed to variety of other threats and stimuli during this period (e.g. boats, killer whales) which may alter their behaviour in similar, or more extreme, ways. Undertaking the assumption that seal behaviour was likely to be unusual, but not the most extreme behaviour observed, was determined to be a practical compromise in using this method to identify periods of unusual dive behaviour. The relatively low number of dive responses identified was likely to be a combination of a variety of factors (see General Discussion), including low numbers of observed encounters at high piling sound levels, the baseline periods potentially containing other sources of disturbance, and limitations in detecting responses which are present (see Chapter 2). As multiple randomisation tests were carried out, it is possible that some of the detected responses were false positives (see Chapter 2) as, by chance, $\sim$5% of behaviours will be above the 95^{th} percentile used to identify unusual behaviour. In this chapter, more encounters than would be expected were above the 99^{th} percentile ($\sim$5% compared to an expected $\sim$1%), highlighting the truly unusual dive behaviour observed. As the dive metrics recorded here only measure one aspect of behaviour, at a dive-by-dive resolution, some type of behavioural change (e.g. horizontal movement, Chapter 3; or changes in body posture) would never be detected by this approach.

4.5.2 Behaviour-switching during pile driving: HMM analyses

HMMs were used to investigate whether pile driving activity also influenced behaviour-switching. In the absence of pile driving, the average persistence within each of the three states was relatively high (0.77–0.86), indicating the presence of dive bouts. For air-breathing animals, it is commonly observed that dives occur in bouts, where animals conduct sequences of similar dives separated by gaps at the surface, resting on land, or switching to a different type of dive (Boyd et al., 1994; Ramasco et al., 2014). Diving in bouts may be an optimal strategy to, for example, manage oxygen stores to enable foraging opportunities to be maximised once prey is discovered. Bouts of dive behaviour have been observed from biologging studies in a number of species including short-finned pilot whales (*Globicephala macrorhynchus*) (Quick et al., 2017), narwhals (*Monodon monoceros*) (Ngô et al., 2019), grey seals (van Beest et al., 2019) and harbour seals (Ramasco et al., 2014).

The HMM approach clustered dives into three behavioural states. State 1 consisted of U-shaped (flat-bottomed) dives, state 2 consisted of more irregular curved, V-shaped and skewed dives, and state 3 consisted of short and shallow dives with longer post-dive surface durations. Although HMMs cannot determine the true biological function of each behavioural state, the observed state-dependent dives can be used to infer the potential activity (or activities) associated with each estimated state. In pinnipeds, U-shaped dives are widely considered to be associated with foraging activity (Carter et al., 2016; Lesage et al., 1999). During these dives, seals travel directly to a depth (often near to the seafloor), remain at that depth for the majority of the dive, and then travel directly to the surface. This type of dive may be an efficient strategy to maximise time at depth when prey are present. Ground-truthing of foraging activity, whilst not possible in this study, can be attempted through the use of stomach temperature sensors which record a decrease in temperature when prey is consumed. Lesage et al. (1999) showed that feeding was observed in all dive types in harbour seals; however, most (75%) of the observed foraging (from stomach sensors and behavioural observations) was during U-shaped dive types. Kuhn et al. (2009) also detected an association between U-shaped dives and feeding in northern elephant seals (*Mirounga angustirostris*), where feeding was observed in all dive types but most often (74%) during a U-shaped dive. The use of accelerometer tags can also provide insights for improving ground-truthing, by recording the fine-scale movement and posture of seals during diving. Vance et al. (2021) used measurements of acceleration to estimate prey capture attempt rates in harbour seals; comparable rates of prey capture were observed both offshore and during transits to and from the potential foraging site.

Overall, current research suggests that U-shaped dives have a strong association with foraging; however, it is worth noting that (1) successful feeding may not occur in all dives, (2) these dives may incorporate other activities associated with feeding such as prey search, and (3) foraging may occur opportunistically in other dive types.

The dives observed in state 3 were short and shallow, with longer periods at the surface. The true biological function of these types of dives is not fully understood, and could be one of a variety of functions including resting, digestion, and prey handling (Carter et al., 2016; Watanabe et al., 2015). In general, these dives were considered a distinct dive type as they were times when seals spent the majority of their time in a relatively low activity state, near the surface and not diving fully. The dives observed in state 2 were the most variable of the states fitted. Dives were of a moderate duration (∼2–5 minutes), but dive shape was not always consistent (Figure 4.8). Dives assigned to state 2 were often more V-shaped or curved, a dive type often associated with travel or transit behaviour (Carter et al., 2016; Kuhn et al., 2009). Some of the dives observed in state 2 were also right or left-skewed (asymmetrical), indicating that seals may be passively drifting within the water column. Drift dives, seen in a variety of pinnipeds including harbour seals, may represent an opportunity for resting or digestion at sea (Mitani et al., 2010; Ramasco et al., 2014; Watanabe et al., 2015). Alternatively, in some cases, irregular dive shapes may represent failed or aborted attempts at foraging dives (e.g. due to irregular bathymetry or prey not being located). Future work to subdivide state 2 into two (or more) states may therefore provide added insight into behavioural-switching and activity budgets of harbour seals during pile driving and baseline periods.

The decision to model harbour seal diving as three behavioural states was a compromise between biological interpretability, statistical stability, and taking account of differences between individuals. Data exploration and initial modelling identified two clear dive types present in all seals: U-shaped dives (state 1) and short dives (state 3). The number of other potential dive types differed by individual, with limited consistency between seals; dive types observed regularly in one seal were limited (or never observed) in others within the study area, leading to duplicate or spurious states being identified. To make comparisons across individuals, the models needed to identify the same numbers and types of states. An alternative approach would be to fit all individuals within the same model (complete pooling; Langrock et al. (2012)); however, this would not respect the individual variability in dive behaviour (observed in data exploration) and would assume that all transition rates were the same between individuals (for baseline and piling exposure). Therefore, HMMs were fit to each individual separately (no pooling) with the aim of identifying the same

three states that could be compared post-hoc (Langrock et al., 2012). In general, AIC can be used to determine the appropriate number of states in a HMM; however, in ecological data AIC typically favours higher numbers of states (Pohle et al., 2017). In practice, these large numbers of states can become biologically uninterpretable. Fitting models with fewer, interpretable states may be used as a pragmatic compromise, at the expense of poorer model fit (DeRuiter et al., 2017; Pohle et al., 2017; Quick et al., 2017). Overall, the choice of the number of states to fit generally depends on the aim of the study (Pohle et al., 2017); in this case, as the aim was to examine whether pile driving affected state switching, a simpler model which fit consistently across individuals was chosen. Here, the residual autocorrelation observed is likely an indication that more states would improve model fit, or that the switching between states is not strictly first-order (i.e. the next state only depends on the previous state). A natural development of the model presented in this chapter would be to determine the number of potential sub-types of dive within state 2, and identify more sophisticated dive metrics (potentially related to dive shape) to aid in state characterisation and identification across seals. van Beest et al. (2019) accounted for residual autocorrelation in a HMM of grey seal diving by dividing dives into groups using the average metrics across each batch of 10 dives; however, this is at the expense of reduced interpretability in state inference and a loss of fine-scale information on behaviour-switching, a factor that is potentially important in behavioural response detection.

Overall, pile driving was found to have a significant effect on behaviour-switching (i.e. piling was retained in the model by AIC) in ten (42%) of the tagged seals and, in most cases, led to decreased persistence within each behavioural state. In nine of the ten seals, overall persistence decreased, leading to more variable and erratic dive behaviour. This is consistent with the findings of Aarts et al. (2018), which found that grey seal diving patterns became more irregular during pile driving. There are two possible behavioural patterns which may contribute to these observed patterns. Seals may be directly disturbed by the sound source, leading to behaviour-switching as a result of a startle reaction (Götz and Janik, 2011), an attempt to assess the "threat" of the sound source, or a change to reduce sound exposure (see previous Discussion Section 4.5.1). Alternatively, the observed changes may also represent aborted attempts at switching into different behaviours later in the pile driving bout (e.g. state 1 is attempted for a couple of dives and then the seal switches back to another type of dive).

4.5.3 Overall findings and implications for renewable developments

Overall, some seals were observed to spend less time at depth and dive more irregularly during pile driving. In the short-term, modifying dive behaviour in this way could provide benefits, in particular a reduced risk of hearing damage, to individuals close to pile driving activity. For example, it has been predicted that harbour seals may experience piling sounds at a level high enough to cause a temporary threshold shift in hearing (Hastie et al., 2015; Whyte et al., 2020). However, in the long-term, if individuals were exposed to sustained noise activity, or regularly exposed to noise from a variety of sources, establishing sustained and efficient bouts of behaviour (e.g. foraging, resting, travelling) could be challenging. If individuals spend less time at the bottom and/or exhibit less sustained behaviours, opportunities to successfully forage or rest may be limited. Understanding how responses to noise may accumulate to impact the health of individuals, and ultimately populations, is a key area of future research (Booth et al., 2020; New et al., 2014).

To inform the assessment of the potential impacts of renewable energy developments, the dose-response relationship presented (Figure 4.6; Table S2) can be used to predict the probability of seals exhibiting unusual dive behaviour in response to pile driving activity at different sound levels. The 95% credible intervals provided could be used in these assessments to account for the uncertainty in the probability of response at each sound level. It should be noted that these doses are single-strike sound exposure levels (SELss) averaged across depths, and that uncertainty in predicted received levels from acoustic modelling may lead to additional uncertainty in response prediction. Comparison of the acoustic models used here with field measurements of piling sound are presented in Whyte et al. (2020) and discussed in the previous Chapter. Whilst this analysis was restricted to a single population and study site (discussed further in General Discussion), the use of these results as an initial estimate of the impact of pile driving sounds on individual dive behaviour fills an important knowledge gap in environmental impact assessment. This population was not naïve to pile driving sounds (they occurred prior to the tagging study), the habitat was relatively shallow and constricted around a highly used haulout site, and the study was restricted to ∼half of the year (January–May); consideration and caution should be applied if applying these results to seals within the breeding season (June–July), to populations that differ strongly from these characteristics, or to piling operations that differ strongly from those measured here.

4.5.4 Conclusions

In conclusion, this chapter used tag data to provide the first insights into the effects of pile driving sounds on harbour seal dive behaviour. By comparing the dive behaviour of individuals during pile driving to their behaviour during baseline periods, there was evidence that some seals may dive less fully (shallower, shorter, decreased bottom time, increased surface time) and dive more irregularly (switch between different dive behaviours more often). Understanding how changes in behaviour may translate into impacts on the stress, health and fitness of populations is critical for future management of wildlife populations and human activities in the increasingly urbanised marine environment. The findings from this study are a key step in achieving this goal.

Chapter 5

General Discussion

5.1 Thesis summary

This thesis was motivated by a need to understand the effects of offshore renewable energy developments on seals, with a focus on the high intensity sounds produced during pile driving for offshore wind farm construction. More broadly, this thesis aimed to fill a knowledge gap in understanding how individual seals move and behave during at-sea anthropogenic disturbance. This thesis was also motivated by a relative paucity of appropriate analytical methods to quantify behavioural responses by seals, and so the work of this thesis aimed to develop and evaluate potential approaches to do this.

In Chapter 2, a comprehensive simulation study was designed to test the performance of Mahalanobis distance methods in biologging studies of behavioural responses. The detection and false positive rates of the approach were quantified for a variety of simulated responses and for different implementations of the method. Performance of the approach varied by the response simulated; critically, the simulation study highlighted (1) the importance of sufficient baseline data in minimising false positives, (2) the need for the covariance matrix used in calculations to only include the baseline period, and (3) the need for biological understanding of the potential duration of responses prior to analysis. Overall, Mahalanobis distance can be used to effectively detect extreme events in movement data (up to 79% of simulated extreme horizontal responses, up to 100% of simulated extreme dive responses) but, as with any statistical method, should always be coupled with biological knowledge and sense-checking. Simulated weaker and shorter responses were detected less reliably. As part of the work of this chapter, the theoretical and computational frameworks for two data simulators were also developed. The resulting `R` functions (see Appendix A) can generate realistic seal GPS locations and dive summary data, with customisable options for specifying the nature, severity, and duration of behavioural responses to simulate. These functions can be used in future simulation studies, e.g. to test the performance of other statistical methods. As the parameters underpinning the baseline and response behaviour (e.g. step length, dive duration, state transition probabilities) are specified by the user as input parameters, these functions present a valuable tool for studies interested in adapting the simulators for different study species.

In Chapters 3 and 4, I examined the individual movements (Chapter 3) and dive behaviours (Chapter 4) of 24 tagged harbour seals during intermittent exposure to *in-situ* pile driving sounds. Mahalanobis distance-based approaches were used to identify responses, in the form of statistically unusual instances of movement and dive behaviour. The detected horizontal responses consisted of either (1) instances of unusually high speed, (2) the cessation

of horizontal movement, or (3) initiating travel after a long period of stationary behaviour. The detected dive responses generally consisted of reduced dive depths, durations and descent rates, increased post-dive surface intervals, and reduced time in the bottom phase of dives. Using provided estimates of the acoustic propagation of pile driving sounds across the study area, Bayesian hierarchical models were used to estimate the dose-response relationship between received sound levels and the probability of behavioural response, accounting for variability between and within individuals. These dose-response relationships can be used in future environmental impact assessments to predict the probability of horizontal and dive behavioural changes as a function of piling sound level. Classification of the dive behaviour into different dive types (using hidden Markov models) also revealed that seals were generally more likely to change dive type during pile driving, highlighting that seal dive behaviour may be more erratic during disturbance.

Although the estimated dose-response relationships for each type of response were similar (estimated p_{50} of horizontal response of 175 (95% CI: 166–181) dB re $1\mu\text{Pa}^2\cdot\text{s}$; estimated p_{50} of dive response of 175 (95% CI: 167–180) dB re $1\mu\text{Pa}^2\cdot\text{s}$), none of the encounters between seals and pile driving were identified as having both a significant horizontal response and a significant dive response (by Mahalanobis distance). Changing horizontal movements or reducing dive behaviour may both potentially reduce the sound exposure experienced by the seal. It is theoretically possible that seals may adopt different strategies depending on the context of the disturbance event (e.g. Ellison et al., 2012). For example, if a seal intended to forage when pile driving starts, it may make sense to move away horizontally from the sound source so that diving and foraging can continue. Alternatively, if the seal is in a high quality foraging area, it may be beneficial to cease diving, remain stationary and wait until the pile driving activity has ended so that foraging can resume. If a seal was travelling past the wind farm, the best approach may be to continue travelling, but at a faster horizontal speed to get past and away from the sound quicker. Alternatively, the best approach may be to continue travelling, but dive shallower to reduce noise exposure. The choice of which approach to take may depend on the direction of travel (i.e. whether they are travelling out to forage, or in to the haulout site), and the current energetic and motivational state of the individual (e.g. hunger levels, energy levels, oxygen management levels). Speculatively, seals which have lower energy levels (e.g. after a long foraging trip) may be more likely to change their dive behaviour than increase their travel speed. Alternatively, if the seal is aware that it is almost approaching its haulout site, it may be sensible to spend a final burst of energy in getting to an area of perceived safety, compared to a seal travelling out which may be preserving its energy for a long foraging trip. If a seal was resting at-sea,

it may make sense to continue this behaviour but reduce dive depth (if they were indeed underwater). Alternatively, as resting can potentially take place elsewhere, it may make sense to move horizontally away from the sound source to resume resting later.

Seals may also adopt different behavioural response strategies depending on their interpretation of the sound source. Seals which interpret the sound source as a potential threat or predator sound (e.g. Deecke et al., 2002; Tyack et al., 2011) may, theoretically, be more likely to respond by moving horizontally away from the sound. This would be an effective anti-predator response if the sound was associated with a predation risk. Seals which respond by either reducing dive behaviour or ceasing horizontal movement, may be curious or uncertain of how to interpret this sound cue (e.g. Blackwell et al., 2004; Kastelein et al., 2012; Kastelein et al., 2018a). By pausing their current behaviour, or coming to the surface, they may be attempting to assess the nature of the sound and its direction.

The findings of this thesis substantially improve understanding of the potential effects of pile driving on seals. For example, Russell et al. (2016) only examined changes in broad patterns of seal density across the study area. In this thesis, the horizontal movement behaviours of individual encounters between seals and pile driving were quantified and examined in detail, providing information on the nature of individual behavioural responses. Additionally, Russell et al. (2016) did not consider the dive behaviour of the tagged seals. In this thesis, the dive behaviour of seals was analysed using both Mahalanobis distance and HMMs, identifying unusual events of dive behaviour and quantifying the probabilities of seals switching dive types during pile driving. In Russell et al. (2016), a binary comparison was made between piling and non-piling, combining multiple pile driving events together. In this thesis, the specific sound levels and durations of each piling event were accounted for separately, and differences between individuals and between encounters could be considered and examined.

Overall, the work of this thesis contributes the first in-depth study on the individual movement behaviour of seals exposed to *in-situ* sound disturbance. In the remainder of this chapter, I discuss the wider significance of this work and highlight key areas of future research in relation to the use of observational data in disturbance studies (Section 5.2), behavioural response research (Section 5.3), renewable energy generation (Section 5.4), and seal movement ecology and disturbance (Section 5.5).

5.2 Using observational data as disturbance experiments: challenges and benefits

During the course of this thesis, observational data on the horizontal movement and dive activity of tagged seals were used to examine the effects of pile driving on these behaviours. As data were not collected as part of a designed experimental protocol using controlled exposures (in contrast to studies such as Miller et al. (2014)), the analysis approach necessarily required dividing the data into times of possible disturbance (so-called 'encounters' with pile driving), and selected times of non-disturbance (baseline). Undertaking this approach highlighted a number of potential challenges and benefits; these insights are discussed further here. While primarily based on studies aiming to examine the effects of sound disturbance on animal behaviour, these insights may prove useful to any biological study aiming to use observational animal tracking data to examine the effects of a potential disturbance source.

5.2.1 Challenges

The challenges encountered by observational studies are likely to vary in nature and severity between different studies, species, data types, locations, and types of disturbance; however, the key challenges presented here are generally applicable to a range of circumstances. In combination, some or all of these factors likely contributed to the relatively low numbers of detected behavioural responses in Chapters 3 and 4.

Defining a standardised trial

When conducting a traditional experimental approach, the disturbance activity can be designed to take place (once or multiple times) at a defined level over a defined duration of time. In contrast, using observational data often requires these decisions to be made post-hoc, once it can be examined what disturbances the animals have potentially been exposed to. In practice, this can lead to challenges in defining what constitutes a potential disturbance event, and increased variability in the nature of these events. For example, in this thesis, quantitative decisions were made to divide the data into different encounters between seals and pile driving; however, each resulting encounter was unique in that it was of a different duration and sequence of piling energies, different initial distance and behaviour of the seals, different times of day and year, etc. There can also be challenges

in determining when these potential disturbance events start and finish. In this thesis, as animals were continually moving around an environment with intermittent pile driving activity, the threshold of 118 dB re $1\mu Pa^2 \cdot s$ was chosen to divide seal behaviour into times when they could potentially hear the sounds from pile driving, and times when it was likely below the median ambient noise level of 118 dB re $1\mu Pa^2 \cdot s$. It is recommended that future studies using this observational approach devote time to determining careful rules, such as these, for deciding what portions of the data can be considered as disturbance trials. Sensitivity analysis could be used to examine whether the choice of rule affects the conclusions drawn. Scientists should also aim to work closely with the relevant industry or noise source to obtain complete and detailed datasets of the timing and extent of the human activities of interest, and, if possible, work together to standardise some of these potential disturbance events.

Contents of the baseline period

When using observational animal tracking data, animals are assumed to be undisturbed and interacting normally with their environment between potential disturbance events. Understanding what the animal potentially experienced during these 'baseline' periods is another challenge (with the tag types used here), as it is possible that the animal experienced other unknown types of disturbance, altering its behaviour in a manner that means the baseline dataset does not only contain their "natural" behaviours. For example, the area surrounding The Wash (the study area considered in this thesis) is an area of shipping activity, and it is has been predicted that harbour seals may have a high number of encounters with vessels in this area (>100 total daily occurrences per 5×5 km grid square (across multiple individuals); Jones et al., 2017). Seals may alter their movement behaviour in response to these encounters (Mikkelsen et al., 2019), as well as encounters with competitors or predators such as grey seals (Brownlow et al., 2016) or killer whales (Bolt et al., 2009). Future studies should consider, in addition to the required data on the animals of interest and the disturbance source, conducting concurrent data collection on pertinent aspects of the environment. Microphone or hydrophone arrays, and animal-attached tags capable of recording sound (Johnson and Tyack, 2003; Mikkelsen et al., 2019), can be used to measure levels of background noise (to determine if disturbance sounds are audible at different locations, see *'Defining a standardised trial'*) and to detect the presence of animal vocalisations (e.g. Riera et al., 2019) or human activities (e.g. Kline et al., 2020). Visual surveys, camera traps and time-lapse photography can be used to monitor the presence of

conspecifics, as well as potential competitor and predator species.

Moderate and weak responses by animals

When reacting to disturbances, animals may not always respond by exhibiting extreme behaviours; instead, they may respond by modifying their behaviour in more minimal ways. Alternatively, animals may respond in ways which are not easy to measure with the resolution and nature of the data collected. For example, in this thesis, the median gap between GPS locations was 10 minutes; detecting very short-term responses would be challenging with this data. While, to an observer, weaker responses may seem less extreme, these responses may still be important in terms of the potential biological impacts on the animal (e.g. induced stress, missed feeding opportunities). If the available measured responses by animals are relatively weak, then statistical testing may have limited ability to distinguish behavioural responses from normal baseline data. For example, in Chapter 2, the simulation study of Mahalanobis distance performance showed that simulated strong horizontal responses were detected up to 79% of the time, whereas weak and short horizontal responses were only detected up to 43% of the time. For studies which expect animals to exhibit more subtle behavioural modifications during disturbance, these studies should consider using statistical methods which look at broader patterns of behaviour and activity (e.g. HMMs or other movement models) and not solely on methods which focus on identifying extreme events (e.g. Mahalanobis distance). Future research into incorporating the effects of context (see Section 5.3.1) may also aid in the detection of moderate and weak responses.

Limited observations at high disturbance levels

In an experimental setup, trials can be designed so that there are even sample sizes across different levels of disturbance (e.g. for trials with difference source levels of a disturbance sound). Animals can be targeted with a disturbance stimuli when they are a fixed distance away (e.g. Gordon et al., 2019), and/or the disturbance stimuli can be ramped up until a high "dose" of the disturbance stimuli has been experienced by the animal (e.g. Miller et al., 2014). In the observational data of this thesis, there were limited observations of encounters at the highest sound exposure levels; $\sim 50\%$ of the encounters did not exceed estimated recieved levels of 137 dB re $1\mu\text{Pa}^2\cdot\text{s}$, and $\sim 75\%$ of the encounters did not exceed 151 dB re $1\mu\text{Pa}^2 \cdot \text{s}$. This challenge may be encountered in observational studies due to

the inability to control spatial and temporal overlap with animals and high disturbance events (and so resulting in low sample sizes), or due to animals already having responded to low levels of disturbance and therefore not coming closer to the disturbance source to experience the higher levels. In this study, if seals exhibited long-range subtle changes in heading to avoid the sound source, these types of responses would also be unlikely to be detected. Prior to data collection, it is recommended that future studies examine the predicted spatial and temporal overlap of the intended study population with the disturbance source, through examining historical data sets, conducting pilot studies, or designing simulation studies and power analyses to check the ability of the intended data collection (and sample size of tags) to answer the intended research questions. For studies which are likely to experience low sample sizes at high disturbance levels, researchers may need to consider whether the observational study should be complemented with a few high-level disturbance experiments, to confirm that the behaviours seen at moderate levels of disturbance can be extrapolated to high disturbance scenarios.

5.2.2 Benefits

Using observational data also provides a number of key benefits and opportunities for improving studies of animal behaviour and disturbance.

Long study periods

Compared to an experimental trial approach (that often requires intensive human involvement), using observational data provides the opportunity to collect data over longer study periods (weeks–months vs. hours–days). Longer study periods mean that disturbance effects can be observed over a wider range of conditions, such as across different seasons, times of day, and habitats. For example, in this thesis, pile driving and seal data collection took place over a four-month study period and across all hours of the day, and confirmed that seal responses to pile driving take place across a variety of external conditions.

Large sample sizes

As observational tracking studies using tags do not need to actively follow individual animals during the study (as data collection is remote), they have the capacity to deploy tags on larger numbers of animals. Larger sample sizes mean that a larger proportion of the

study population is included in data collection, aiding in ensuring that the animals that have been studied are representative of the wider population. Larger sample sizes also enable animals with different individual-level characteristics (e.g. age, sex) to be sufficiently represented in data collection. In this thesis, tags were deployed on 24 harbour seals (20 of which were considered to have encounters with pile driving), compared to sonar exposure experiments which typically involve < 5 individuals (DeRuiter et al., 2013; Miller et al., 2015) or < 10 individuals (Antunes et al., 2014; Miller et al., 2014). Here, having this sample size revealed the substantial difference in the number of exposures between male and female seals, with females encountering pile driving a factor of 10 greater times than males, possibly due to differences in distribution and movement patterns.

Improved understanding of behavioural repertoires

Long-term observational data also allows improved understanding of natural behavioural variation, both between and within individuals. As illustrated in the simulation study in Chapter 2, having a longer baseline period leads to less false-positives in behavioural response detection. This is partly because the animal has more time to exhibit its full behavioural repertoire (its range of behaviours), and therefore studies can be more confident that the changes observed during disturbance are indeed unusual. Understanding the types, durations and activity patterns of different baseline behaviours is an important consideration for any study and statistical approach.

In-situ disturbance

Perhaps one of the most significant benefits of observational studies is that they can take place *in-situ*, with the real disturbance source, in the intended environment, with wild animals that are undertaking their natural behaviours. *In-situ* studies generally require less extrapolation and assumptions about how results observed in experiments may be transferable to real-life situations, e.g. compared to captive studies such as Kastelein et al. (2018a). While experimental approaches with wild animals are considerably closer to an *in-situ* approach, these studies may still have to make compromises that could alter how animals perceive or respond to the disturbance. For example, experimental studies of sound disturbance may have to use artificial versions or recordings of sounds (Hastie et al., 2018; Koschinski et al., 2003) or play sounds at quieter levels than normal operation (Mikkelsen et al., 2017). In this thesis, the pile driving activity that took place was part

of the construction activity of the wind farm, providing insights into how wild free-ranging seals behave during this potential disturbance period.

5.3 Behavioural response studies

5.3.1 The importance of context

This thesis is part of a growing field of research aiming to examine the effects of different stimuli or events on animal behaviour (Doherty et al., 2021; Dominoni et al., 2020; Harris et al., 2018). In any behavioural response study, it is essential to consider the context, i.e. the intrinsic and extrinsic factors which may affect when, how and why animals alter their behaviour. Intrinsic factors are features of the individual animal which may affect their response (e.g. age, sex, current behaviour, previous experience with the stimuli). Extrinsic factors include the spatial context (e.g. habitat), temporal context (e.g. time of day), and features of the stimuli itself (e.g. duration and pattern of its presence, distance to the animal, sound level and characteristics). Here, a number of contextual factors were considered as candidate variables in the dose-response Bayesian modelling in Chapters 3 and 4; however, in this study, there was not enough evidence to support the inclusion of these extra variables. Detecting a clear effect of contextual variables is challenging, likely because context is multi-faceted and unique to each disturbance encounter.

Some studies have found quantitative evidence of the role of context in behavioural response studies. Goldbogen et al. (2013) investigated the behavioural responses of tagged blue whales to playbacks of simulated mid-frequency active sonar signals and pseudo-random noise. They found that the responses observed depended on the current behavioural state of the whale, with individuals more likely to respond when they were deep-feeding or not feeding, compared to responses being rare when whales were surface feeding. Zoratto et al. (2014) measured the behaviours of European starlings (*Sturnus vulgaris*) during video playbacks of predator (peregrine falcon; *Falco peregrinus*) attacks on starling flocks. Starlings responded to the predator video by decreasing the amount of body turns (remaining still), but increasing the amount of head turns (increased vigilance), relative to a control video without a predator. The response of the exposed starlings depended on social context, with individual starlings remaining still for longer than when starlings were in a mini-flock with other birds. Investigating the effects of context can also produce more unexpected results. For example, Beale and Monaghan (2004) conducted disturbance experiments on turnstones (*Arenaria interpres*), where some bird groups were provisioned with supplemen-

tary food in their normal habitat. Birds which were provided with supplementary food were more vigilant and responded to the disturbance at a greater distance; therefore, birds in greater need of acquiring food took greater risks to tolerate disturbance and continue foraging. In Ellison et al. (2012) it was proposed that although responses to sound are likely to be dependent on both the received level of sound and the signal-to-noise ratio (relative to background noise), when sound (or disturbance) level is lower, contextual factors such as these are likely to become more important in driving responses. Particular contextual factors which should be investigated in future studies of seal responses to sound include aspects of the foraging trip (e.g. proportion of trip elapsed, travelling in vs. out), information on foraging success, estimates of habitat quality, energetic and physiological state of the animal, and the estimated exposure of the seal to other disturbance events (e.g. in pristine vs. polluted environments).

Biologging technology opportunities

New developments in biologging technology present key opportunities for further understanding behavioural responses and their context (Figure 5.1). Accelerometers, devices which record fine-scale and high-frequency (>once per second) measurements of activity levels and body posture, may provide more detailed information on the current behaviour of animals prior to disturbance. In particular, accelerometers can be used to monitor sudden body, head, and jaw movements ('jerks'), which can indicate foraging and prey capture attempts in seals (Adachi et al., 2019; Vance et al., 2021; Volpov et al., 2015). Incorporating these measurements into behavioural response studies may provide important information on the motivation for, and disruption of, foraging during disturbance events. Recent technological developments have led to non-invasive animal-attached devices which can now record aspects of seal physiology underwater (McKnight et al., 2019, 2021). In particular, the ability to measure oxygen concentration levels within the animal may provide improved understanding on whether oxygen management during dives is an important contextual factor in behavioural responses. Tags which record aspects of the environment (e.g. sound, Mikkelsen et al., 2019; water temperature and salinity, Photopoulou et al., 2020) can be used to measure the presence of other natural and anthropogenic drivers of behaviour. Future studies may also wish to consider the use of wearable video technology, which can be used to verify the types of prey being consumed by seals (Adachi et al., 2021; Yoshino et al., 2020), or wearable sonar devices which can be used to measure prey encounters (Goulet et al., 2019).

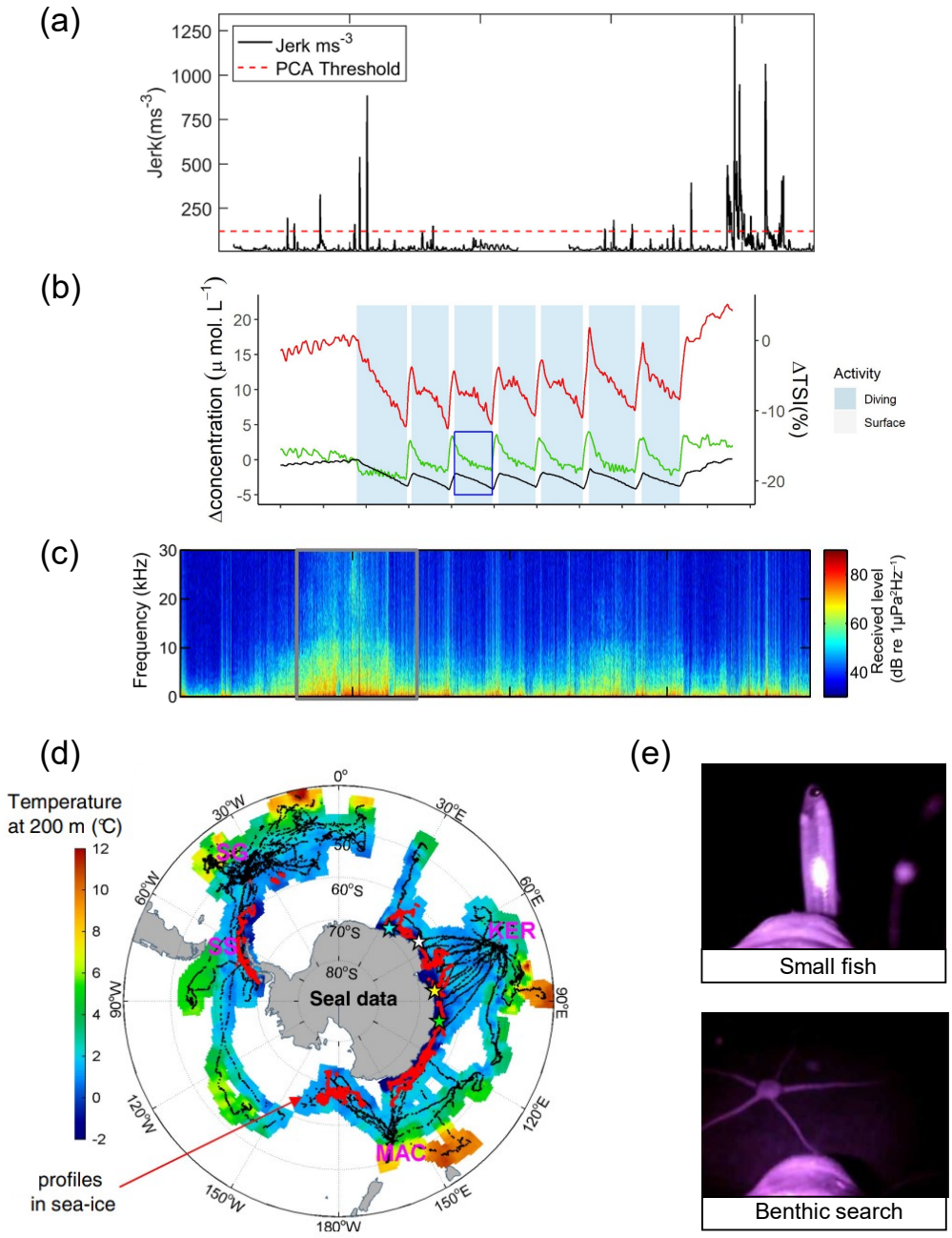


Figure 5.1 Biologging tools to understand context in future behavioural response studies for pinnipeds. (a) Acceleration jerks of harbour seals measured over time, using a threshold (red line) to determine prey capture events (PCAs). (b) Measurements of harbour seal physiology over time, including tissue saturation index (TSI; red line), blood volume (green line), and haemoglobin oxygenation (black line). (c) Recorded sound over time on a grey seal tag, plotted as the received power spectrum density level. (d) Ocean temperatures recorded by biologging devices on southern elephant seals. (e) Video footage of prey encounters in foraging dives of northern elephant seals. Figures adapted from (a) Vance et al. (2021), (b) McKnight et al. (2019), (c) Mikkelsen et al. (2019), (d) Charrassin et al. (2008), and (e) Adachi et al. (2021).

Statistical opportunities

As context may influence both the occurrence and nature of the behavioural responses observed, modelling contextual variables may be more more challenging when using two-stage modelling approaches (e.g. Mahalanobis distance followed by Bayesian hierarchical models). One-stage 'holistic' models (e.g. HMMs, and other options discussed in Section 5.5.2) may provide a stronger alternative for studies interested in context. The Mahalanobis distance-based approach compares behaviours observed at a given time to one overall measurement of "average" behaviour during baseline. One extension of the work in this thesis, to account further for behavioural context, would be to first divide the baseline data into a chosen number of behaviours (e.g. shallow dives, deep dives) and use Mahalanobis distance to then examine changes within each behavioural category. In this thesis, available water depth was used as a proxy for habitat in the fitted HMMs, to account for changes in dive behaviour as seals move between different environments. Another extension of this work would be to include additional context-related covariates (e.g. time of day) on the HMM transition probabilities (as individual and interaction terms), to investigate how different factors affect behavioural switching.

5.3.2 Challenges and future directions in behavioural response studies

Temporal mismatch of responses

A key remaining challenge of behavioural response studies is in instances of temporal mismatch between disturbances and observed behaviour changes. If clear behavioural changes occur during the disturbance activity, then it is relatively straightforward to relate these changes to a known level of disturbance; however, if behavioural changes occur shortly after the disturbance activity ends, it is challenging to assign a "dose" to these disturbances and determine if this is a true response to the activity. For example, it is plausible to imagine that an animal may tolerate a disturbance activity if it is carrying out an important behaviour (e.g. foraging), but that it may exhibit a delayed response by deciding to move away from the area after its behavioural goal has been completed. Examining this robustly within a statistical framework is difficult, as the main tool which is used to help distinguish normal vs. disturbed behaviour is typically the known timing of the potential disturbance activity. In the data used in this thesis, there were anecdotal observations of a few occasions in which seals appeared to show visibly unusual movement behaviour between two bouts of pile driving that were several hours apart. If the seal didn't behave unusually until after the

first pile driving bout ended, these responses would likely be undetected. If, theoretically, the seal responded to the first pile driving bout and was still behaving unusually by the time of the second pile driving bout, the response was likely caused by the high received levels of piling at the end of the first piling bout, but an incorrectly low "dose" from the ramp-up of the second piling bout would be assigned to this response (the piling energies are lower and the seal may also have responded and moved further away). Alternatively, if measured changes in movement behaviour take time to reach a level of significance for response detection (when they exceed the Mahalanobis distance 95^{th} percentile threshold), the dose at the time when the response was detected may not be the same as the dose that elicited the response. In this thesis, both of these concerns were somewhat accounted for by using intermediate window sizes (which would record a wider range of possible doses) and by looking for horizontal responses in an additional 30-minute period after pile driving ceased; however, temporal mismatch remains a key statistical and practical challenge in any behavioural response study.

The relationship between behaviour and data collection

Any biological study should consider how animal behaviour may influence the data collected; however, this is particularly important for studies that aim to make inferences about changes in behaviour. For example, Mikkelsen et al. (2017) examined the responses of harbour seals to playbacks of simulated acoustic deterrent device sounds, using theodolites to track seal locations at the surface. During exposure to the sounds, there were more sightings and seals were observed closer to sound source, relative to baseline periods. From this result, the study proposed that seals may have been attracted to the sound source; however, it is equally possible that seals were spending more time at (or near) the surface as a response to the sound. This finding would be consistent with the behavioural responses identified in Chapter 4, in which seal dives were shorter and shallower, with longer post-dive surface intervals. If some seals respond to sound disturbance by spending more time at the surface, then this may have important implications for visual surveys and avoiding bias. Similarly, future research is also required into understanding whether the amount or quality of data collected from GPS tags varies by behaviour. This could be approached by further investigating the data used in this thesis. For diving animals, GPS fixes are necessarily limited to times when the animal is at the surface, and so it is possible that this limitation may bias data collection and miss important behaviours (e.g. if an animal is fleeing with very short surface intervals).

5.4 Sustainable renewable energy generation

In this thesis, dose-response curves were estimated in Chapters 3 and 4, which quantify the relationship (and associated uncertainty) between received piling sound levels and behavioural responses. These functions can be used to predict the probability of behavioural changes occurring in harbour seals during at-sea pile driving for renewable energy or other industrial developments. Environmental impact assessments using these relationships should also use the 95% credible intervals provided to account for the estimated uncertainty in the relationships. In using these functions, it should be noted that uncertainty in the acoustic propagation modelling predictions was not accounted for in the dose-response modelling, and that these predictions are based on a single population of harbour seals at a particular time of year (January–May). The pertinent aspects of this study population are discussed in more detail below (Section 5.4.1); extrapolation of these results to populations which differ highly from this study population should be approached with caution. It is recommended that similar studies are carried out on different populations, at different times of year, to validate these results more widely. If the impact assessment only allows for one type of dose-response curve, the estimated relationship presented here (Chapter 3) for horizontal responses is recommended, as it is more conservative with higher levels of uncertainty. These relationships also provide a basis for comparison with other similar studies examining different causes of disturbance in different species and habitats. More broadly, this thesis provides improved understanding of how individual seals may move and dive during encounters with pile driving. Here, a number of key caveats and considerations for further research are discussed, in relation to the future of environmentally-responsible renewable energy generation.

5.4.1 Representativeness of this study

The results presented in this thesis are based on 24 tagged harbour seals from populations in The Wash and The Thames on the east coast of England, UK. In 2020, the STRANGE Framework (Webster and Rutz, 2020) was proposed as a means of reducing sampling biases in studies of animal behaviour, by improving data collection methods, and the declaration and discussion of potential biases. Here, I use the STRANGE framework to highlight important features of the study population and sampled individuals (Table 5.1), which should be considered when applying these results more widely.

Table 5.1 Summary of the STRANGE Framework assessment for the 24 harbour seals used within this study. For a full description of the STRANGE Framework, see Webster and Rutz (2020).

STRANGE framework	**Study animals**
Social background	Unknown
Trappability and self-selection	Possible tagging bias of bolder, less cautious animals
Rearing history	Unknown; possible prior exposure to shipping
Acclimation and habituation	Prior pile driving exposure; no known prior tagging
Natural changes in responsiveness	Study of one 4-month period (Jan–May)
Genetic make-up	Distinct from Scotland and N Ireland populations
Experience	Possible age and sex differences in level of experience

Social background

As the tagged individuals were not visually observed prior to tagging, and were not known to the study, the social status or background of the tagged seals in this study is unknown. During tag deployment, it was attempted to obtain an approximately even split of males (n=11) and females (n=13), to ensure that different demographics of the population were represented. Three juveniles were also tagged.

Trappability and self-selection

As seals were required to be caught at or near haulout sites to deploy the tags, it is possible that the individuals caught were of a certain behavioural type or physiological state. This is a common concern with trapping and tagging studies (Carter et al., 2012; Webster and Rutz, 2020), where bolder or less cautious individuals are more prone to being caught and sampled. The seal capture method used in this study (seine netting), reduces the likely levels of behavioural bias compared to traditional capture methods; however, there may be biases in the availability of seals to being captured (e.g. in terms of differences in the amount of time spent hauled out). In the data collected in this study, there was high individual variation in the movement patterns and in the responses observed; however, it should be noted that there may be "sensitive" or "shy" individuals (who may be more responsive to disturbance) in the population which were not sampled.

Rearing history

While the rearing and developmental history of the tagged individuals is unknown, it is known that shipping occurs within the study area (Jones et al., 2017), and so some or all of the seals studied here are likely to have been previously exposed to this. The tagged seals may therefore already be acclimatised to regular human disturbance.

Acclimation and habituation

During tag deployment, the seals were checked for signs of other tags and marks from previous potential captures; the seals tagged in this study had no known previous capture or tag deployment experience. To control for the effects of potential disturbance from the tag deployment, and to avoid biases in measuring movements when seals are first fitted with a tag, the first week of all deployments was removed from the analysis (McKnight, 2011). Although the seals had not been tagged before, they were likely not naïve to pile driving sounds. Pile driving for Lincs wind farm construction began in May 2011 (eight months prior to tag deployment), and other wind farms had been built in the study area in previous years. The behaviour and responses observed in this population may therefore not be as extreme as they were when pile driving initially began, or compared to a population in a quieter environment. Alternatively, it is also possible that seals may have become sensitised to pile driving sounds over time.

Natural changes in responsiveness

The data collected here covered a four-month period from January to May; however, there were no observations of behaviour at other times of the year. Data on both baseline and disturbance were collected across all hours of the day and across tidal states, and there was no evidence of differences in responses between these times.

Genetic make-up (and other population-level specificities)

The genetic make-up of the individuals studied here has not been examined in detail; however, there is evidence that the south-east England harbour seal population is distinct from those in Scotland and Northern Ireland (using genetic, movement, and demographic data; Carroll et al., 2020). One notable feature of this population is the geography of the haulout site and surrounding coastlines. The Wash is a relatively enclosed area, with lots of shallow habitats and sandbanks for seals to haul out on at low tide. In particular, the location of the wind farm is on the edge of the only entrance to the haulout site, and thus seals must pass within ∼20 km of the site to enter and leave the haulout site. This also means that seals are most likely to be exhibiting travel-like behaviour when within close-range of the wind farm, which may influence the detectability and types of behavioural responses observed in this study.

Experience

In this study, seals of both sexes and a range of ages were tagged (ages are known from tooth analysis of 15 individuals; a range of 2.0–23.0 years). The older individuals may be more experienced in encountering disturbance, and it was observed that females encoun-

tered pile driving more often than males in this study. It is unclear whether experienced individuals would respond more weakly (habituated) or more strongly (improved knowledge or sensitisation) to disturbance; however, more "experienced" individuals may also have accumulated temporary or permanent hearing damage (Hastie et al., 2015; Whyte et al., 2020). From the HMM analysis, the only individual which exhibited significantly increased persistence within dive types during pile driving was one of the old females (20.5 years old). The relative hearing abilities of some of the tagged individuals was assessed (at the time of tagging) in Lucke et al. (2016). Differences in hearing sensitivity between individuals were observed; however, hearing sensitivities were broadly similar to those measured in captive animals (Lucke et al., 2016). Further studies of hearing sensitivity before, and if possible after, sound disturbance are required to further understand how hearing ability may be related to behavioural changes and avoidance of anthropogenic sounds.

5.4.2 Behaviour and pile driving: Scaling-up of possible effects

Population consequences of disturbance

To monitor and predict the non-lethal effects of disturbance on wildlife populations, it is important to understand the mechanisms of these effects at both an individual and population level. The Population Consequences of Disturbance (PCoD) framework provides a quantitative approach for assessing the effects of disturbance (Figure 5.2; Pirotta et al., 2018a), with a particular focus on marine mammal applications. The PCoD framework has been used to predict effects through a series of quantitative relationships, comprising (1) the effect of a stressor on physiology and behaviour, (2) the effect of physiological and behavioural changes on health and vital rates, and (3) the effect of individual vital rates on wider population dynamics (Figure 5.2). Where quantitative information is limited, stages of the framework can be supplemented with information gathered from expert elicitation activities (King et al., 2015). For example, experts can be asked to provide an estimate of the number of days an animal may tolerate disturbance without an effect on survival (King et al., 2015), and groups of experts can be asked to come to a consensus on their estimate and its associated uncertainty.

In the data on harbour seals used in this thesis, predicted impacts differed by individual (Table 5.2). Out of the 24 tagged seals, 20 were estimated to have encountered pile driving sounds (above ambient noise levels) at least once during the study. The four seals which did not encounter pile driving were all adult males. Of those encountering piling sounds,

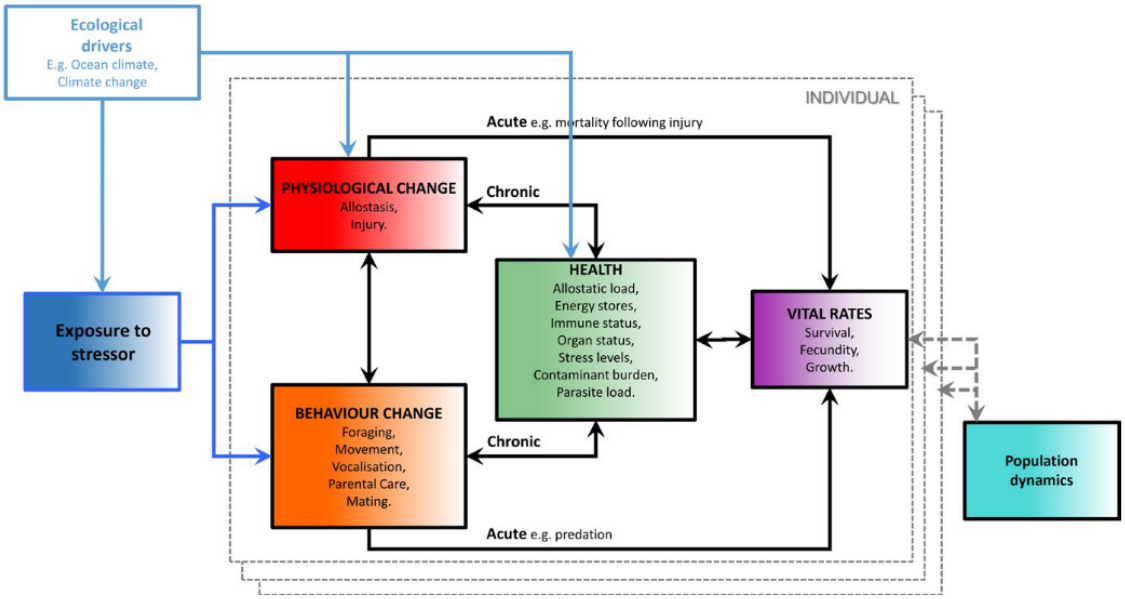


Figure 5.2 The Population Consequences of Disturbance (PCoD) conceptual framework. Aspects of the framework are denoted by boxes, with arrows showing the potential linkages between different aspects. Effects are examined at the individual level (within grey dashed box), and then accumulated across individuals to predict effects on population dynamics. Figure from Pirotta et al. (2018a).

in 16 seals an effect on behaviour was detected, either in the form of a horizontal movement response (Mahalanobis distance analysis, Chapter 3), a vertical movement response (Mahalanobis distance analysis, Chapter 4), or an alteration to dive behaviour switching (HMM analysis, Chapter 4). The Mahalanobis distance analyses identified responses in 12 seals, and six of these seals exhibited both horizontal and vertical responses during different encounters with pile driving. Additionally, Whyte et al. (2020) estimated that four of the seals received sound levels sufficient to cause temporary threshold shifts (TTS) in hearing; all four of these seals were adult females. Understanding whether the proportion of tagged individuals predicted to be affected by pile driving sounds in this study is representative of the proportion of individuals affected in the wider population remains an important consideration for further research, and must be coupled with knowledge of the pertinent aspects of the study population and tagged individuals (see previous Section 5.4.1).

The results of this thesis provide information for the first stage of the PCoD framework, linking levels of piling sound to probabilities of behavioural changes in harbour seals. Gathering quantitative information on the further stages of the PCoD framework is typically more challenging, as these changes become harder to directly observe; however, understand-

Table 5.2 Summary of the different predicted impacts on the 24 tagged harbour seals. For each seal, shown are the proportion of encounters in which the seal was detected to respond horizontally (Chapter 3), the proportion of encounters in which the seal was detected to respond vertically (Chapter 4), and whether the seal was predicted to alter its dive behaviour switching during pile driving (Chapter 4). Also highlighted are the seals which were predicted to have recieved sound levels sufficient to cause a temporary threshold shift in hearing (TTS; from Whyte et al., 2020), and the sex and age class of each individual. The four individuals which did not encounter any piling bouts above ambient sound level are denoted by dashes (-). Individuals with the prefix 'pv42-' were tagged in The Wash; individuals with the prefix 'pv40-' were tagged in the Thames.

Seal reference	**Horizontal response**	**Vertical response**	**Dive switch**	**Hearing**	**Sex**	**Age class**
pv42-162-12	0.06	0.07	-	-	F	Adult
pv42-165-12	0.00	0.00	Y	TTS	F	Juvenile
pv42-194-12	0.00	0.00	-	-	M	Adult
pv42-198-12	-	-	-	-	M	Adult
pv42-220-12	-	-	-	-	M	Adult
pv42-221-12	0.00	0.00	Y	-	M	Adult
pv42-266-12	1.00	0.00	-	-	F	Adult
pv42-277-12	0.06	0.03	Y	TTS	F	Adult
pv42-287-12	-	-	-	-	M	Adult
pv42-288-12	0.00	0.00	-	-	F	Adult
pv42-289-12	0.00	0.00	Y	-	M	Adult
pv42-290-12	0.14	0.25	Y	-	F	Adult
pv42-291-12	0.00	0.17	Y	-	F	Adult
pv42-292-12	0.00	0.00	-	-	M	Adult
pv42-293-12	0.25	0.33	-	-	F	Adult
pv42-294-12	0.00	0.00	-	-	M	Adult
pv42-295-12	0.10	0.05	Y	-	F	Adult
pv42-316-12	0.33	0.00	Y	-	M	Juvenile
pv42-317-12	0.00	0.07	Y	-	F	Adult
pv42-318-12	0.00	0.09	-	-	F	Adult
pv42-319-12	0.00	0.00	Y	-	M	Juvenile
pv42-320-12	0.14	0.00	-	TTS	F	Adult
pv40-268-12	0.12	0.12	-	TTS	F	Adult
pv40-270-12	-	-	-	-	M	Adult

ing how behavioural changes in seal movement and dive behaviour may affect health and vital rates is an important area of future research. This could be approached by simulation studies (e.g. Chudzinska et al., 2021; Farmer et al., 2018; New et al., 2014), and the observed behavioural changes in this thesis provide key information to help parameterise these simulations. Monitoring of wild populations under times of disturbance (e.g. visual and acoustic surveys, capture-recapture) is also a useful avenue (Booth et al., 2020); however, these methods may require long-term studies with sufficient baseline and large numbers of sampled individuals, and may need to be coupled with concurrent individual-level measurements of disturbance. A promising avenue of future data collection is in assessments of body condition and energetic expenditure of individuals. Future work should investigate whether pinnipeds experience changes in foraging success, energy used, and body condition during disturbance, e.g. using drift dive assessments or measures of prey capture attempts (Schick et al., 2013, Vance et al., 2021).

Multiple stressors

Pile driving is one of a number of potential stressors that may impact marine wildlife. Sound pollution is produced from a variety of other human activities including shipping, seismic surveys, and sonar (Duarte et al., 2021). Other potential anthropogenic stressors include chemical pollutants, bycatch, and collisions with boats (Nelms et al., 2021). Humans may also alter aspects of the natural environment, putting increased pressure on marine animal populations, e.g. climate change, prey availability, predator and competitor presence, and invasive species (Halpern et al., 2007). Understanding how these different potential stressors act individually and cumulatively on marine animal populations is a key area of further research (National Academies of Sciences Engineering and Medicine, 2017). The interaction between multiple stressors may lead to effects which are greater (often termed "synergism") or smaller (often termed "antagonism") than the effects of the individual stressors (Folt et al., 1999; Piggott et al., 2015). For example, a population viability analysis of Swedish-Danish harbour seals (Silva et al., 2021) indicated a low risk of extinction during moderate levels of the single stressors considered (hunting, decreased fecundity from chemical exposure, epizootics such as phocine distemper virus); however, there was a high risk of extinction for the most vulnerable subpopulations when all three stressors together occurred at low levels.

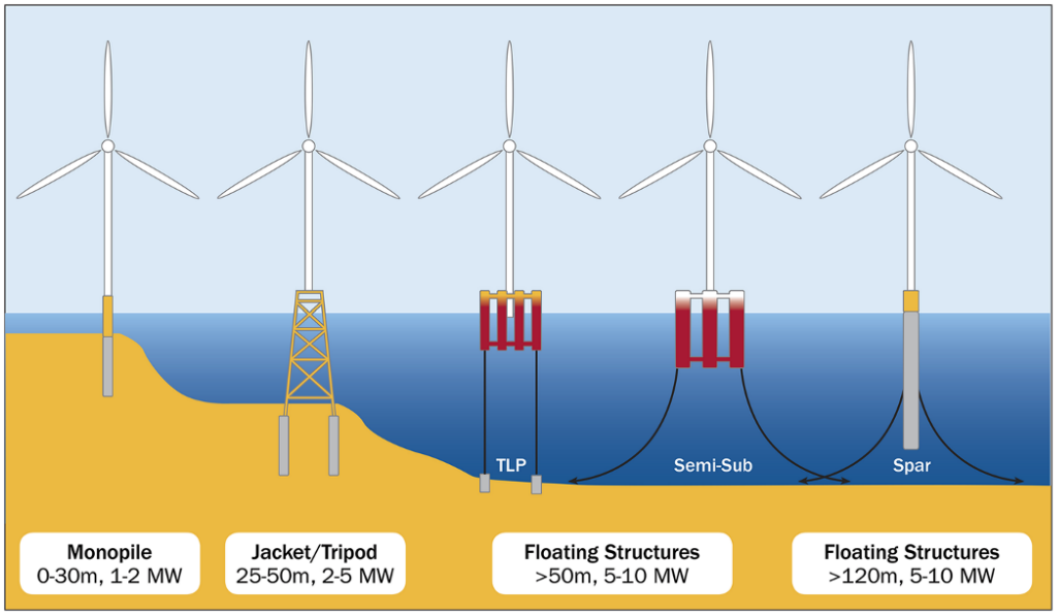


Figure 5.3 Types of offshore wind turbine foundations. Monopile foundations were used at the wind farm studied in this thesis, but jacket/tripod foundations are also widely used. Different types of floating foundation structures are in initial use or development: Tension Leg Platform (TLP), semi-submersible (Semi-sub), Spar Buoy (Spar). Figure from Bailey et al. (2014).

5.4.3 Technological and industry considerations

In the wind farm construction described in this thesis, pile driving was used to install steel monopile turbines (diameter of 5.2 metres) during construction of Lincs wind farm in 2011–2012. Offshore renewable energy is developing rapidly, and there are a number of novel technologies now being considered. The world's first commercial wind farm using floating foundation structures (Figure 5.3) was installed in Scotland and began operating in 2017 (WindEurope, 2017). Floating turbines can be constructed in deeper water than traditional turbine foundations (Bailey et al., 2014; Breton and Moe, 2009), and construction may produce less noise pollution than pile driving for monopiles. Floating wind may therefore aid in increasing options for wind farm site location and reducing disturbance; however, further research is required into the potential environmental effects of this new technology, in particular the risk of entanglement for large cetaceans or the risk of exclusion from key areas of habitat for marine mammals.

Noise abatement measures, such as bubble curtains, can be used to reduce the levels of sound propagated out from a pile driving site (Dähne et al., 2017); however, there are currently limited applications to turbines in deeper waters ($>$ 45 m) and with larger ($>$8 m)

pile sizes (Verfuss et al., 2019). Developers should consider whether noise abatement measures would be applicable at the intended construction site, and if they would effectively reduce noise pollution for the wildlife populations concerned. Another potential mitigation measure is the use of acoustic deterrent devices to deter marine mammals from the immediate vicinity of the site, with the aim of reducing auditory injury from exposure to pile driving. Harbour porpoises have been found to have a higher probability of response (by decreasing their occurrence) when acoustic deterrent devices are used prior to pile driving, in comparison to pile driving alone (Graham et al., 2019). Controlled exposure experiments of acoustic deterrent devices to tagged harbour seals (Gordon et al., 2019) detected responses in all seals within 1 km of the device; however, seals did not always move away, and the efficacy of deterrent devices appears variable across studies (Coram et al., 2014; Götz and Janik, 2016; Graham et al., 2009; Mikkelsen et al., 2017). Future research is required to understand whether the benefits of using deterrent devices prior to pile driving outweighs the potential risks of noise pollution, disturbance, and hearing damage posed by the device itself. Results observed from deterrence studies in aquaculture may not be directly applicable, as seals may behave differently when there is not a clear prey source to be attracted to (e.g. at salmon farms). For example, captive studies of grey seals found that seals tolerated anthropogenic sound playbacks to feed at high-quality prey patches, but not at low-quality prey patches (Hastie et al., 2021). This could be approached by *in-situ* studies of seal movements at pile-driving sites that use deterrents, to examine whether deterrence from these devices alters the predicted levels of risk from pile driving sound exposure.

5.4.4 Short-term construction vs. long-term operation

There is currently limited evidence to suggest that operational offshore wind farms pose a significant risk to marine mammals. Koschinski et al. (2003) conducted a playback study of simulated operational wind turbine generator sounds to harbour porpoises and harbour seals using theodolite location tracking. Significant differences in the distribution of both species were detected; however, the scale of these differences was relatively small. For example, in seals, the median distance of seals from the sound source was 284 m (relative to 239 m during control), the closest approach was 12.0 m (relative to 9.6 m during control), and no extreme responses such as fast swimming were observed. Notably, as part of the processing and playback of the artificial sounds used, high-frequency components may have been introduced into the original recording, potentially altering the behavioural responses

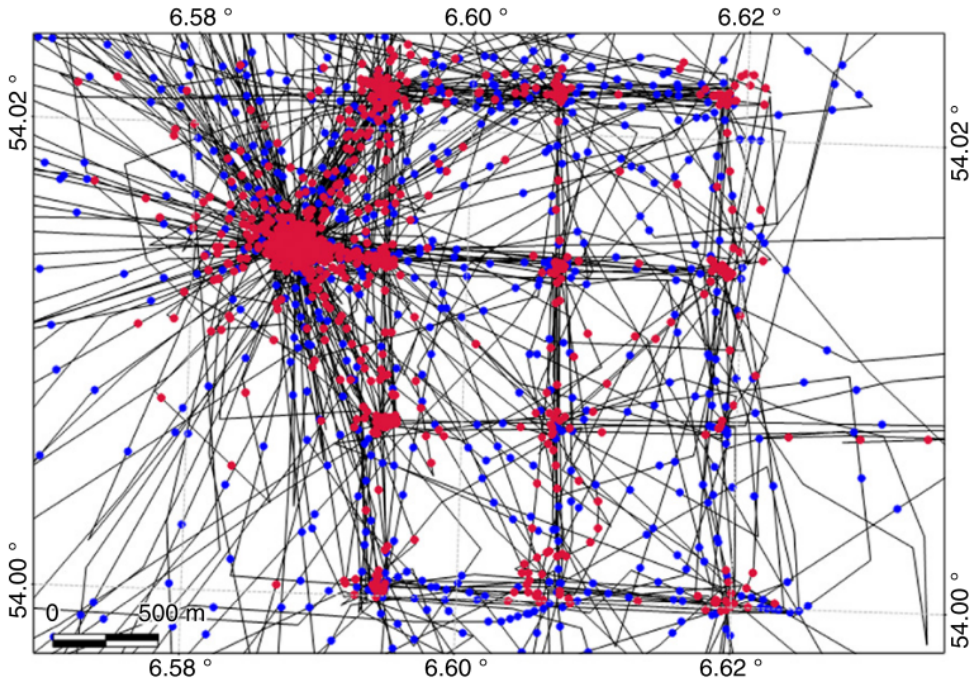


Figure 5.4 Movement tracks of a tagged harbour seal at Alpha Ventus wind farm, Germany. A state-space model was used to estimate times when seals were most likely to be foraging (red dots) or travelling (blue dots). Figure from Russell et al. (2014).

observed (Madsen et al., 2006; Koschinski et al., 2003).

In contrast, in the long-term, it is possible that the offshore structure of the turbines may be ecologically neutral or beneficial (Inger et al., 2009). The hard substrate of the turbine foundation can provide a surface for biofouling species (e.g. mussels, macroalgae, barnacles, anemones) to establish, building up over time into an artificial reef (Degraer et al., 2020). As the community becomes established, fish can be attracted to the turbine sites to forage, and fishing activities are likely to be excluded from wind farm sites (De Troch et al., 2013; Inger et al., 2009). Larger marine predators may also be attracted to the sites. Russell et al. (2014) tracked the movements of seals around offshore structures, and found that the movement of some individuals was highly associated with the turbine locations and undersea pipelines (Figure 5.4). State-space modelling also suggested that some individual seals appeared to be foraging at these sites. Understanding the potential risks and benefits of offshore renewables for different species is an important area of future research, to ensure that new developments in technology can be appropriately assessed for potential impact. This is essential to allow future renewable developments to be built in a timely and environmentally-responsible manner.

5.5 Seal movement ecology, behaviour and disturbance

5.5.1 Effects on prey

An aspect not investigated in this thesis is the effects of sound on the prey species of marine top predators. Harbour seals are considered to be generalist predators, feeding on a variety of fish, cephalopods and other marine invertebrates (Wilson and Hammond, 2019). In understanding how predators such as seals respond to sound disturbance, it is important to also consider how sound may alter the behaviour or availability of prey. A growing number of studies are documenting the effects of anthropogenic sound on marine invertebrates. Jones et al. (2020) examined the behaviour of longfin squid (*Doryteuthis pealeii*) during captive playback experiments of different pile driving recordings. Almost all (15 out of 16) of the squid responded to the sounds, by exhibiting alarm and anti-predator behaviours including inking, jetting away, and changes in body pattern appearance. Squid exhibited similar responses to both repeats of the experiment, but responses were all short-term and constrained to the first 30 pulses (∼1 minute) of the playback. In the long-term, there is the potential for these disturbance events to alter the levels of alertness, or the response of squid to real predator events, if individuals become habituated to disturbance. Solan et al. (2016) also conducted captive playback studies of both pile driving sounds and shipping noise to Manila clam (*Ruditapes philippinarum*), Norway lobster (*Nephrops norvegicus*), and Ophiuroid brittlestars (*Amphiura filiformis*). The invertebrate species exhibited multiple types of stress-related responses, including less movement activity by *N. norvegicus* and the cessation of suspension-feeding by *R. philippinarum*; however, there was no evidence of an effect on the organisms' tissue chemistry and the responses differed depending on the sound type. Hermit crabs (*Pagurus bernhardus*) have also been observed to spend less time assessing which new shell to move into when they are exposed to playbacks of white noise (Walsh et al., 2017). Overall, the understanding of how invertebrates respond to sound disturbance is relatively limited, and therefore this is an important area for further research. The responses observed so far suggest that sound disturbance of prey species may have implications for the overall fitness of prey populations and important ecological processes such as bioturbation.

A number of studies have investigated the effect of anthropogenic sound disturbance on fish (for comprehensive reviews see Popper and Hastings, 2009 and Slabbekoorn et al., 2010). During exposure to an air gun in a large sea cage, schooling fish (travelly, *Pseudocaranx dentex*; pink snapper, *Pagrus auratus*) have been seen to increase aggregation

and move down in the water column (Fewtrell and McCauley, 2012). For pile driving sounds in particular, captive seabass (*Dicentrarchus labrax*) have been found to alter their shoaling behaviour during captive playback studies (Herbert-Read et al., 2017). During playback, fish within groups were spaced further apart, decreased their movement speed, and exhibited less movement coordination with the other fish in the shoal. Seabass have also been found to increase their ventilation rates during pile driving playbacks, indicating a stress response to these impulsive sounds (Radford et al., 2016). Studies of free-ranging responses of fish to sound disturbance are less common, and understanding whether responses observed in captivity are applicable to wild populations is not always clear. In a promising approach for future studies, Hawkins et al. (2014) used echosounders to track the movements of schools of free-ranging fish during pile driving sound playbacks; sprat (*Sprattus sprattus*) schools were generally seen to disperse (horizontally), whereas mackerel (*Scomber scombrus*) schools generally changed their depth during playback.

It is possible that sound disturbance of prey species may increase foraging opportunities for predators, if prey become disorientated, or if schooling and anti-predator behaviours are less efficient. The responses observed in this thesis do not provide any clear evidence that this has occurred in this study; however, this aspect of predator-prey relationships was not investigated in detail. Future research should investigate the movement patterns (horizontal and vertical) of prey species during *in-situ* disturbance events, the prey capture attempts of predators during disturbance, and the long-term population health of prey populations that experience disturbance.

5.5.2 Movement models

Building on the work in this thesis, a useful extension for further work is in the field of animal movement modelling. Discrete-time random walk models consider the movement of individual animals as a sequence of movement steps and turn angles (Hooten et al., 2017). Typically, these models are based on observations of these movement metrics over regularly spaced time intervals. Due to ecological interest in modelling and estimating the attraction of animals to particular landscape features (e.g. nests or haulout sites, feeding grounds), the effects of spatial attraction have been incorporated into a variety of movement models (e.g. McClintock et al., 2012). Movement models which also include the ability to estimate the location of a repulsion centre, and its effect on movement parameters such as steps and turns (e.g. Tracey et al., 2005) would be useful methods to apply to future studies of behavioural disturbance. Continuous-time correlated random walks do not

require regularly-spaced data observations, and instead model movement as a process on a continuous time scale, e.g. by modelling the instantaneous velocity of the animal (Johnson et al., 2008). As these approaches do not require interpolation (or other modification) of the original recorded data to fit the modelling framework, they may provide additional insights into fine-scale behavioural responses which may be lost if data were interpolated. Another possible approach is the use of potential functions (Preisler et al., 2013); these may prove useful for estimating the strength of different forces influencing animal movement and improving understanding of how animals balance different benefits and risks as they move through their environment.

Diving and flying animals move in three-dimensions. In the majority of animal movement studies, and in the chapters of this thesis, the horizontal and vertical movement of animals are considered separately. Both aspects of movement are part of the same fundamental biological process, and considering these aspects separately may influence the inferences made on both the behavioural states of animals and the overall levels of energetic expenditure. This dichotomy between studies of vertical (diving or flying) movement and studies of horizontal movement has likely been driven by previous technological limitations allowing only one type of movement to be recorded thoroughly, and (more recently) by the practical and statistical challenge of robustly integrating data recorded at different temporal scales. One approach to this challenge is to incorporate both dimensions of the data into the same discrete-time framework, by summarising or adjusting one of the data streams so that both are available at the same temporal resolution. This approach can increase the reliability of behavioural state classification and improve understanding of the relationship between animal behaviour and environmental covariates (McClintock et al., 2013, 2017). Another possible approach is the use of hierarchical hidden Markov models, which can incorporate multiple time series of data recorded at different resolutions (Adam et al., 2019). These models can also be used to estimate hierarchical behavioural states, using the models to quantify both high-resolution (e.g. hourly) and low-resolution (e.g. daily) patterns of behaviour (Adam et al., 2019; Leos-Barajas et al., 2017). More broadly, there is a need for the development of statistical approaches which can efficiently model the movements of animals through three-dimensional space; these approaches would be a key step forward in examining the fine-scale interactions of animals with obstacles in their environment (e.g. harbour porpoises and tidal turbines (Gillespie et al., 2021); marine birds and wind turbines (Furness et al., 2013)).

5.6 Concluding remarks

Understanding the behaviour of individual animals is a critical step in estimating the potential effects of human disturbance on wildlife populations. In this thesis, I have shown that biologging devices present a powerful tool to do this, and that statistical approaches which model the movement of individuals can provide quantitative and biological insights into animal behaviour during these disturbance events. In particular, the application of these methods to a harbour seal population exposed to pile driving activity improve our understanding of the movements and dive behaviour of individual seals during sound disturbance. These results fill a key gap in environmental impact assessments, enabling future renewable energy developments to predict the effects of offshore activities on animal behaviour, and ensuring that decisions of development consent are supported by robust scientific information. Overall, this thesis contributes the first in-depth study of the individual at-sea behaviour of seals during an *in-situ* disturbance event, and provides advice and insights for future studies pursuing this important area of research.

References

Aarts, G., Brasseur, S., and Kirkwood, R. (2018). Behavioural response of grey seals to pile-driving. Technical report, Wageningen Marine Research, Report No. C006/18.

Adachi, T., Hückstädt, L. A., Tift, M. S., Costa, D. P., Naito, Y., and Takahashi, A. (2019). Inferring prey size variation from mandible acceleration in northern elephant seals. *Marine Mammal Science*, 35(3):893–908.

Adachi, T., Takahashi, A., Costa, D. P., Robinson, P. W., Hückstädt, L. A., Peterson, S. H., Holser, R. R., Beltran, R. S., Keates, T. R., and Naito, Y. (2021). Forced into an ecological corner: Round-the-clock deep foraging on small prey by elephant seals. *Science Advances*, 7(20):21–25.

Adam, T., Griffiths, C. A., Leos-Barajas, V., Meese, E. N., Lowe, C. G., Blackwell, P. G., Righton, D., and Langrock, R. (2019). Joint modelling of multi-scale animal movement data using hierarchical hidden Markov models. *Methods in Ecology and Evolution*, 10(9):1536–1550.

Agostinelli, C. and Lund, U. (2018). *CircStats: Circular Statistics, from "Topics in Circular Statistics" (2001)*. S-plus original by Ulric Lund and R port by Claudio Agostinelli, R package version 0.2-6.

Aguilar Soto, N., Johnson, M., Madsen, P. T., Tyack, P. L., Bocconcelli, A., and Fabrizio Borsani, J. (2006). Does intense ship noise disrupt foraging in deep-diving cuvier's beaked whales (Ziphius cavirostris)? *Marine Mammal Science*, 22(3):690–699.

Ahlén, I., Baagøe, H. J., and Bach, L. (2009). Behavior of scandinavian bats during migration and foraging at sea. *Journal of Mammalogy*, 90(6):1318–1323.

Ainslie, M. (2010). *Principles of sonar performance modelling*. Springer-Praxis, Chichester, UK.

Andersen, M. and Aars, J. (2008). Short-term behavioural response of polar bears (Ursus maritimus) to snowmobile disturbance. *Polar Biology*, 31(4):501–507.

Antunes, R., Kvadsheim, P. H., Lam, F. P. A., Tyack, P. L., Thomas, L., Wensveen, P. J., and Miller, P. J. O. (2014). High thresholds for avoidance of sonar by free-ranging long-finned pilot whales (Globicephala melas). *Marine Pollution Bulletin*, 83(1):165–180.

Arnett, E. B., Brown, W. K., Erickson, W. P., Fiedler, J. K., Hamilton, B. L., Henry, T. H., Jain, A., Johnson, G. D., Kerns, J., Koford, R. R., Nicholson, C. P., O'Connell, T. J., Piorkowski, M. D., and Tankersley, R. D. (2008). Patterns of bat fatalities at wind energy facilities in North America. *Journal of Wildlife Management*, 72(1):61–78.

Arso Civil, M., Hague, E., Langley, I., and Scott-Hayward, L. (2021). Allo-suckling occurrence and its effect on lactation and nursing duration in harbour seals (Phoca vitulina) in Orkney, Scotland. *Behavioral Ecology and Sociobiology*, 75(8):121.

Au, W. and Hastings, M. (2008). *Principles of marine bioacoustics*. Springer, New York.

Auger-Méthé, M., Newman, K., Cole, D., Empacher, F., Gryba, R., King, A. A., Leos-Barajas, V., Mills Flemming, J., Nielsen, A., Petris, G., and Thomas, L. (2021). A guide to state–space modeling of ecological time series. *Ecological Monographs*, 0(0):e01470.

Baechler, J., Beck, C. A., and Bowen, W. D. (2002). Dive shapes reveal temporal changes in the foraging behaviour of different age and sex classes of harbour seals (Phoca vitulina). *Canadian Journal of Zoology*, 80(9):1569–1577.

Bagniewska, J. M., Hart, T., Harrington, L. A., and MacDonald, D. W. (2013). Hidden Markov analysis describes dive patterns in semiaquatic animals. *Behavioral Ecology*, 24(3):659–667.

Bailey, H., Brookes, K. L., and Thompson, P. M. (2014). Assessing environmental impacts of offshore wind farms: Lessons learned and recommendations for the future. *Aquatic Biosystems*, 10(1):1–13.

Bailey, H., Senior, B., Simmons, D., Rusin, J., Picken, G., and Thompson, P. M. (2010). Assessing underwater noise levels during pile-driving at an offshore windfarm and its potential effects on marine mammals. *Marine Pollution Bulletin*, 60(6):888–897.

Barber, J. R., Crooks, K. R., and Fristrup, K. M. (2010). The costs of chronic noise exposure for terrestrial organisms. *Trends in Ecology and Evolution*, 25(3):180–189.

Beale, C. M. and Monaghan, P. (2004). Behavioural responses to human disturbance: a matter of choice? *Animal Behaviour*, 68(5):1065–1069.

Bekkby, T. and Bjørge, A. (2000). Diving behaviour of harbour seal Phoca vitulina pups from nursing to independent feeding. *Journal of Sea Research*, 44(3-4):267–275.

Bell, W. (1990). Central place foraging. In *Searching Behaviour. Chapman and Hall Animal Behaviour Series*, pages 171–187. Springer, Dordrecht.

Benhemma-Le Gall, A., Graham, I. M., Merchant, N. D., and Thompson, P. M. (2021). Broad-scale responses of harbor porpoises to pile-driving and vessel activities during offshore windfarm construction. *Frontiers in Marine Science*, 8:664724.

Berta, A. (2018). Pinnipeds. In Wursig, B., Thewissen, J. G. M. H., and Kovacs, K., editors, *Encyclopedia of Marine Mammals*, pages 733–740. Elsevier Science and Technology, 3rd edition.

Bivand, R., Keitt, T., and Rowlingson, B. (2021). *rgdal: Bindings for the 'geospatial' data abstraction library*. R package version 1.5-23.

Bivand, R. and Lewin-Koh, N. (2021). *maptools: Tools for handling spatial objects*. R package version 1.1-1.

Blackwell, S. B., Lawson, J. W., and Williams, M. T. (2004). Tolerance by ringed seals (Phoca hispida) to impact pipe-driving and construction sounds at an oil production island. *The Journal of the Acoustical Society of America*, 115(5):2346–2357.

Blanchet, M.-A., Lydersen, C., Ims, R. A., and Kovacs, K. M. (2015). Seasonal, oceanographic and atmospheric drivers of diving behaviour in a temperate seal species living in the high arctic. *PLoS ONE*, 10(7):e0132686.

Blix, A. S. (2018). Adaptations to deep and prolonged diving in phocid seals. *Journal of Experimental Biology*, 221(12):jeb182972.

Bolt, H. E., Harvey, P. V., Mandleberg, L., and Foote, A. D. (2009). Occurrence of killer whales in Scottish inshore waters: temporal and spatial patterns relative to the distribution of declining harbour seal populations. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 19:671–675.

Booth, C. G., Sinclair, R. R., and Harwood, J. (2020). Methods for monitoring for the population consequences of disturbance in marine mammals: a review. *Frontiers in Marine Science*, 7:115.

Boyd, A. I. L., Arnould, J. P. Y., Barton, T., and Croxall, J. P. (1994). Foraging behaviour of Antarctic fur seals during periods of contrasting prey abundance. *Journal of Animal Ecology*, 63(3):703–713.

BP (2021). *Statistical review of world energy 2021*. London, UK.

Braimoh, B., Iwajomo, S., Wilson, M., Chaskda, A., Ajang, A., and Cresswell, W. (2018). Managing human disturbance: factors influencing flight-initiation distance of birds in a West African nature reserve. *Ostrich*, 89(1):59–69.

Brandt, M., Diederichs, A., Betke, K., and 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.

Breton, S. P. and Moe, G. (2009). Status, plans and technologies for offshore wind turbines in Europe and North America. *Renewable Energy*, 34(3):646–654.

Brownie, C., Boos, D. D., and Hughes-Oliver, J. (1990). Modifying the t and ANOVA F tests when treatment is expected to increase variability relative to controls. *Biometrics*, 46(1):259–266.

Brownlow, A., Onoufriou, J., Bishop, A., Davison, N., and Thompson, D. (2016). Corkscrew seals: Grey seal (Halichoerus grypus) infanticide and cannibalism may indicate the cause of spiral lacerations in seals. *PLoS ONE*, 11(6):e0156464.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York, 2nd edition.

Carroll, E. L., Hall, A., Olsen, M. T., Onoufriou, A. B., Gaggiotti, O. E., and Russell, D. J. (2020). Perturbation drives changing metapopulation dynamics in a top marine predator. *Proceedings of the Royal Society B: Biological Sciences*, 287(1928):20200318.

Carter, A. J., Heinsohn, R., Goldizen, A. W., and Biro, P. A. (2012). Boldness, trappability and sampling bias in wild lizards. *Animal Behaviour*, 83(4):1051–1058.

Carter, M. I., McClintock, B. T., Embling, C. B., Bennett, K. A., Thompson, D., and Russell, D. J. (2020a). From pup to predator: generalized hidden Markov models reveal rapid development of movement strategies in a naïve long-lived vertebrate. *Oikos*, 129(5):630–642.

Carter, M. I., Russell, D. J., Embling, C. B., Blight, C. J., Thompson, D., Hosegood, P. J., and Bennett, K. A. (2017). Intrinsic and extrinsic factors drive ontogeny of early-life at-sea behaviour in a marine top predator. *Scientific Reports*, 7(1):1–14.

Carter, M. I. D., Bennett, K. A., Embling, C. B., Hosegood, P. J., and Russell, D. J. F. (2016). Navigating uncertain waters: a critical review of inferring foraging behaviour from location and dive data in pinnipeds. *Movement Ecology*, 4(1):25.

Carter, M. I. D., Boehme, L., Duck, C. D., Grecian, W. J., Hastie, G. D., Mcconnell, B. J., Miller, D. L., Morris, C. D., Moss, S. E. W., Thompson, D., Thompson, P. M., and Russell, D. J. F. (2020b). Habitat-based predictions of at-sea distribution for grey and harbour seals in the British Isles. *Sea Mammal Research Unit, University of St Andrews, Report to BEIS, OESEA-16-76/OESEA-17-78*.

Charrassin, J. B., Hindell, M., Rintoul, S. R., Roquet, F., Sokolov, S., Biuw, M., Costa, D., Boehme, L., Lovell, P., Coleman, R., Timmermann, R., Meijers, A., Meredith, M., Park, Y. H., Bailleul, F., Goebel, M., Tremblay, Y., Bost, C. A., McMahon, C. R., Field, I. C., Fedak, M. A., and Guinet, C. (2008). Southern Ocean frontal structure and sea-ice formation rates revealed by elephant seals. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33):11634–11639.

Christensen, J. W., Keeling, L. J., and Nielsen, B. L. (2005). Responses of horses to novel visual, olfactory and auditory stimuli. *Applied Animal Behaviour Science*, 93(1-2):53–65.

Christiansen, B., Denda, A., and Christiansen, S. (2020). Potential effects of deep seabed mining on pelagic and benthopelagic biota. *Marine Policy*, 114:103442.

Chudzinska, M., Nabe-Nielsen, J., Smout, S., Aarts, G., Brasseur, S., Graham, I., Thompson, P., and McConnell, B. (2021). AgentSeal: Agent-based model describing movement of marine central-place foragers. *Ecological Modelling*, 440:109397.

Cooke, S. J., Hinch, S. G., Wikelski, M., Andrews, R. D., Kuchel, L. J., Wolcott, T. G., and Butler, P. J. (2004). Biotelemetry: A mechanistic approach to ecology. *Trends in Ecology and Evolution*, 19(6):334–343.

Coram, A., Gordon, J., Thompson, D., and Northridge, S. (2014). Evaluating and Assessing the Relative Effectiveness of Acoustic Deterrent Devices and other Non-Lethal Measures on Marine Mammals. *Scottish Government, Marine Scotland Report*.

Costa, D. P., Crocker, D. E., Gedamke, J., Webb, P. M., Houser, D. S., Blackwell, S. B., Waples, D., Hayes, S. A., and Le Boeuf, B. J. (2003). The effect of a low-frequency sound source (acoustic thermometry of the ocean climate) on the diving behavior of juvenile northern elephant seals, Mirounga angustirostris . *The Journal of the Acoustical Society of America*, 113(2):1155–1165.

Cox, S. L., Orgeret, F., Gesta, M., Rodde, C., Heizer, I., Weimerskirch, H., and Guinet, C. (2018). Processing of acceleration and dive data on-board satellite relay tags to investigate diving and foraging behaviour in free-ranging marine predators. *Methods in Ecology and Evolution*, 9(1):64–77.

Cox, T., Ragen, T., Read, A., Vos, E., Baird, R., Balcomb, K., Barlow, J., Caldwell, J., Cranford, T., Crum, L., D'Amico, A., D'Spain, G., Fernandez, A., Finneran, J., Gentry, R., Gerth, W., Gulland, F., Hildebrand, J., Houser, D., Hullar, T., Jepson, P., Ketten, D., MacLeod, C., Miller, P., Moore, S., Mountain, D., Palka, D., Ponganis, P., Rommel, S., Rowles, T., Taylor, B., Tyack, P., Wartzok, D., Gisiner, R., Mead, J., and Benner, L. (2006). Understanding the impacts of anthropogenic sound on beaked whales. *Journal of Cetacean Research and Management*, 7(3):177–187.

Crawford, J. A., Frost, K. J., Quakenbush, L. T., and Whiting, A. (2019). Seasonal and diel differences in dive and haul-out behavior of adult and subadult ringed seals (Pusa hispida) in the Bering and Chukchi seas. *Polar Biology*, 42(1):65–80.

Crocker, D. and Champagne, C. (2018). Pinniped physiology. In Wursig, B., Thewissen, J., and Kovacs, K. M., editors, *Encyclopedia of Marine Mammals*, pages 726–733. Elsevier Science and Technology, 3rd edition.

Culloch, R. M., Anderwald, P., Brandecker, A., Haberlin, D., McGovern, B., Pinfield, R., Visser, F., Jessopp, M., and Cronin, M. (2016). Effect of construction-related activities and vessel traffic on marine mammals. *Marine Ecology Progress Series*, 549:231–242.

Cunningham, L., Baxter, J. M., Boyd, I. L., Duck, C. D., Lonergan, M., Moss, S. E., and McConnell, B. (2009). Harbour seal movements and haul-out patterns: Implications for monitoring and management. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 19(4):398–407.

Dähne, M., Tougaard, J., Carstensen, J., Rose, A., and Nabe-Nielsen, J. (2017). Bubble curtains attenuate noise from offshore wind farm construction and reduce temporary habitat loss for harbour porpoises. *Marine Ecology Progress Series*, 580:221–237.

Daly, E. and White, M. (2021). Bottom trawling noise: Are fishing vessels polluting to deeper acoustic habitats? *Marine Pollution Bulletin*, 162:111877.

Davies, N. B., West, S. A., and Krebs, J. R. (2012). *An Introduction to Behavioural Ecology*. John Wiley and Sons, Oxford, 4th edition.

de Jong, C., Binnerts, B., Prior, M., Colin, M., Ainslie, M., Mulder, I., and Hartstra, I. (2019). Wozep – WP2: update of the Aquarius models for marine pile driving sound predictions. In *TNO Report (2018), number R11671*, page 94. The Hague, Netherlands.

De Troch, M., Reubens, J. T., Heirman, E., Degraer, S., and Vincx, M. (2013). Energy profiling of demersal fish: A case-study in wind farm artificial reefs. *Marine Environmental Research*, 92:224–233.

Deecke, V. B., Slater, P. J., and Ford, J. K. (2002). Selective habituation shapes acoustic predator recognition in harbour seals. *Nature*, 420(6912):171–173.

Degraer, S., Carey, D. A., Coolen, J. W., Hutchison, Z. L., Kerckhof, F., Rumes, B., and Vanaverbeke, J. (2020). Offshore wind farm artificial reefs affect ecosystem structure and functioning: A synthesis. *Oceanography*, 33(4):48–57.

Delignette-Muller, M. L. and Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64(4):1–34, R package version 1.0–14.

DeLong, R. L. and Stewart, B. S. (1991). Diving patterns of northern elephant seal bulls. *Marine Mammal Science*, 7(4):369–384.

DeRuiter, S. L., Langrock, R., Skirbutas, T., Goldbogen, J. A., Calambokidis, J., Friedlaender, A. S., and Southall, B. L. (2017). A multivariate mixed hidden markov model for blue whale behaviour and responses to sound exposure. *Annals of Applied Statistics*, 11(1):362–392.

DeRuiter, S. L., Southall, B. L., Calambokidis, J., Zimmer, W. M. X., Sadykova, D., Falcone, E. A., Friedlaender, A. S., Joseph, J. E., Moretti, D., Schorr, G. S., Thomas, L., and Tyack, P. L. (2013). First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. *Biology Letters*, 9(4):20130223.

Díaz, H. and Guedes Soares, C. (2020). Review of the current status, technology and future trends of offshore wind farms. *Ocean Engineering*, 209:107381.

Dietz, R., Teilmann, J., Andersen, S. M., Rigét, F., and Olsen, M. T. (2013). Movements and site fidelity of harbour seals (Phoca vitulina) in Kattegat, Denmark, with implications for the epidemiology of the phocine distemper virus. *ICES Journal of Marine Science*, 70(1):186–195.

Doherty, T. S., Hays, G. C., and Driscoll, D. A. (2021). Human disturbance causes widespread disruption of animal movement. *Nature Ecology and Evolution*, 5(4):513–519.

Dominoni, D. M., Halfwerk, W., Baird, E., Buxton, R. T., Fernández-Juricic, E., Fristrup, K. M., McKenna, M. F., Mennitt, D. J., Perkin, E. K., Seymoure, B. M., Stoner, D. C., Tennessen, J. B., Toth, C. A., Tyrrell, L. P., Wilson, A., Francis, C. D., Carter, N. H., and Barber, J. R. (2020). Why conservation biology can benefit from sensory ecology. *Nature Ecology and Evolution*, 4(4):502–511.

Doniol-Valcroze, T., Lesage, V., Giard, J., and Michaud, R. (2011). Optimal foraging theory predicts diving and feeding strategies of the largest marine predator. *Behavioral Ecology*, 22(4):880–888.

D'Orazio, M. (2019). *StatMatch: Statistical Matching or Data Fusion*. R package version 1.3.0.

Dowle, M. and Srinivasan, A. (2021). *data.table: Extension of 'data.frame'*. R package version 1.14.0.

Duarte, C. M., Chapuis, L., Collin, S. P., Costa, D. P., Devassy, R. P., Eguiluz, V. M., Erbe, C., Gordon, T. A., Halpern, B. S., Harding, H. R., Havlik, M. N., Meekan, M., Merchant, N. D., Miksis-Olds, J. L., Parsons, M., Predragovic, M., Radford, A. N., Radford, C. A., Simpson, S. D., Slabbekoorn, H., Staaterman, E., Van Opzeeland, I. C., Winderen, J., Zhang, X., and Juanes, F. (2021). The soundscape of the Anthropocene ocean. *Science*, 371(6529).

Durbach, I. N., Harris, C. M., Martin, C., Helble, T. A., Henderson, E. E., Ierley, G., Thomas, L., and Martin, S. W. (2021). Changes in the movement and calling behavior of minke whales (Balaenoptera acutorostrata) in response to navy training. *Frontiers in Marine Science*, 8:880.

Dvorkin, D. (2012). *lcmix: Layered and chained mixture models*. R package version 0.3.

Edrén, S. M. C., Andersen, S. M., Teilmann, J., Carstensen, J., Harders, P. B., Dietz, R., and Miller, L. A. (2010). The effect of a large Danish offshore wind farm on harbor and gray seal haul-out behavior. *Marine Mammal Science*, 26(3):614–634.

Ellison, W. T., Southall, B. L., Clark, C. W., and Frankel, A. S. (2012). A new context-based approach to assess marine mammal behavioral responses to anthropogenic sounds. *Conservation Biology*, 26(1):21–28.

Erbe, C., Marley, S. A., Schoeman, R. P., Smith, J. N., Trigg, L. E., and Embling, C. B. (2019). The effects of ship noise on marine mammals—a review. *Frontiers in Marine Science*, 6(October).

Estabrook, B. J., Ponirakis, D. W., Clark, C. W., and Rice, A. N. (2016). Widespread spatial and temporal extent of anthropogenic noise across the northeastern Gulf of Mexico shelf ecosystem. *Endangered Species Research*, 30(1):267–282.

Fahlman, A., Tyack, P. L., Miller, P. J., and Kvadsheim, P. H. (2014). How man-made interference might cause gas bubble emboli in deep diving whales. *Frontiers in Physiology*, 5(13):1–6.

Falcone, E. A., Schorr, G. S., Watwood, S. L., DeRuiter, S. L., Zerbini, A. N., Andrews, R. D., Morrissey, R. P., and Moretti, D. J. (2017). Diving behaviour of Cuvier's beaked whales exposed to two types of military sonar. *Royal Society Open Science*, 4(8):170629.

Farmer, N. A., Baker, K., Zeddies, D. G., Denes, S. L., Noren, D. P., Garrison, L. P., Machernis, A., Fougères, E. M., and Zykov, M. (2018). Population consequences of disturbance by offshore oil and gas activity for endangered sperm whales (Physeter macrocephalus). *Biological Conservation*, 227:189–204.

Fewtrell, J. L. and McCauley, R. D. (2012). Impact of air gun noise on the behaviour of marine fish and squid. *Marine Pollution Bulletin*, 64(5):984–993.

Findlay, C. R., Ripple, H. D., Coomber, F., Froud, K., Harries, O., van Geel, N. C., Calderan, S. V., Benjamins, S., Risch, D., and Wilson, B. (2018). Mapping widespread and increasing underwater noise pollution from acoustic deterrent devices. *Marine Pollution Bulletin*, 135:1042–1050.

Folt, C., Chen, C., Moore, M., and Burnaford, J. (1999). Synergism and antagonism among multiple stressors. *Limnology and Oceanography*, 44(3, part 2):864–877.

Fregosi, S., Klinck, H., Horning, M., Costa, D. P., Mann, D., Sexton, K., Hückstädt, L. A., Mellinger, D. K., and Southall, B. L. (2016). An animal-borne active acoustic tag for minimally invasive behavioral response studies on marine mammals. *Animal Biotelemetry*, 4(1):1–15.

Frisk, G. V. (2012). Noiseonomics: The relationship between ambient noise levels in the sea and global economic trends. *Scientific Reports*, 2(1):1–4.

Furness, R. W., Wade, H. M., and Masden, E. A. (2013). Assessing vulnerability of marine bird populations to offshore wind farms. *Journal of Environmental Management*, 119:56–66.

Gallon, S. L., Sparling, C. E., Georges, J. Y., Fedak, M. A., Biuw, M., and Thompson, D. (2007). How fast does a seal swim? Variations in swimming behaviour under differing foraging conditions. *Journal of Experimental Biology*, 210(18):3285–3294.

Garnier, S. (2018). *viridis: Default Color Maps from 'matplotlib'*. R package version 0.5.1.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.

Gillespie, D., Palmer, L., Macaulay, J., Sparling, C., and Hastie, G. (2021). Harbour porpoises exhibit localized evasion of a tidal turbine. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 31(9):2459–2468.

Goldbogen, J. A., Southall, B. L., DeRuiter, S. L., Calambokidis, J., Friedlaender, A. S., Hazen, E. L., Falcone, E. A., Schorr, G. S., Douglas, A., Moretti, D. J., Kyburg, C., McKenna, M. F., and Tyack, P. L. (2013). Blue whales respond to simulated mid-frequency military sonar. *Proceedings of the Royal Society B: Biological Sciences*, 280(1765):20130657.

Gordon, J., Blight, C., Bryant, E., and Thompson, D. (2019). Measuring responses of harbour seals to potential aversive acoustic mitigation signals using controlled exposure behavioural response studies. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 29(S1):157–177.

Götz, T. and Janik, V. M. (2010). Aversiveness of sounds in phocid seals: psycho-physiological factors, learning processes and motivation. *The Journal of Experimental Biology*, 213(9):1536–1548.

Götz, T. and Janik, V. M. (2011). Repeated elicitation of the acoustic startle reflex leads to sensitisation in subsequent avoidance behaviour and induces fear conditioning. *BMC Neuroscience*, 12:30.

Götz, T. and Janik, V. M. (2016). Non-lethal management of carnivore predation: long-term tests with a startle reflex-based deterrence system on a fish farm. *Animal Conservation*, 19(3):212–221.

Goulet, P., Guinet, C., Swift, R., Madsen, P. T., and Johnson, M. (2019). A miniature biomimetic sonar and movement tag to study the biotic environment and predator-prey interactions in aquatic animals. *Deep-Sea Research Part I: Oceanographic Research Papers*, 148:1–11.

Graham, I. M., Harris, R. N., Denny, B., Fowden, D., and Pullan, D. (2009). Testing the effectiveness of an acoustic deterrent device for excluding seals from Atlantic salmon rivers in Scotland. *ICES Journal of Marine Science*, 66(5):860–864.

Graham, I. M., Merchant, N. D., Farcas, A., Barton, T. R., Cheney, B., Bono, S., and Thompson, P. M. (2019). Harbour porpoise responses to pile-driving diminish over time. *Royal Society Open Science*, 6(6):190335.

Grolemund, G. and Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3):1–25, R package version 1.7.10.

Halfwerk, W. and Slabbekoorn, H. (2015). Pollution going multimodal: The complex impact of the human-altered sensory environment on animal perception and performance. *Biology Letters*, 11(4):20141051.

Halpern, B. S., Selkoe, K. A., Micheli, F., and Kappel, C. V. (2007). Evaluating and ranking the vulnerability of global marine ecosystems to anthropogenic threats. *Conservation Biology*, 21(5):1301–1315.

Hanggi, E. B. and Schusterman, R. J. (1994). Underwater acoustic displays and individual variation in male harbour seals, Phoca vitulina. *Animal Behaviour*, 48(6):1275–1283.

Härkönen, T. and Heide-Jørgensen, M. P. (1990). Comparative life histories of East Atlantic and other harbour seal populations. *Ophelia*, 32(3):211–235.

Harris, C. M., Thomas, L., Falcone, E. A., Hildebrand, J., Houser, D., Kvadsheim, P. H., Lam, F. P. A., Miller, P. J., Moretti, D. J., Read, A. J., Slabbekoorn, H., Southall,

B. L., Tyack, P. L., Wartzok, D., and Janik, V. M. (2018). Marine mammals and sonar: Dose-response studies, the risk-disturbance hypothesis and the role of exposure context. *Journal of Applied Ecology*, 55(1):396–404.

Harris, C. M., Thomas, L., Sadykova, D., DeRuiter, S. L., Tyack, P. L., Southall, B. L., Read, A. J., and Miller, P. J. (2016). The challenges of analyzing behavioral response study data: An overview of the MOCHA (Multi-study OCean acoustics Human effects Analysis) Project. *The Effects of Noise on Aquatic Life II, Advances in Experimental Medicine and Biology*, pages 399–407.

Harris, R. E., Miller, G. W., and Richardson, W. J. (2001). Seal responses to airgun sounds during summer seismic surveys in the Alaskan Beaufort Sea. *Marine Mammal Science*, 17(4):795–812.

Hart, T., Mann, R., Coulson, T., Pettorelli, N., and Trathan, P. (2010). Behavioural switching in a central place forager: Patterns of diving behaviour in the macaroni penguin (Eudyptes chrysolophus). *Marine Biology*, 157(7):1543–1553.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A K-Means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108.

Hastie, G. D., Donovan, C., Gotz, T., and Janik, V. M. (2014). Behavioral responses by grey seals (Halichoerus grypus) to high frequency sonar. *Marine Pollution Bulletin*, 79:205–210.

Hastie, G. D., Lepper, P., McKnight, J. C., Milne, R., Russell, D. J., and Thompson, D. (2021). Acoustic risk balancing by marine mammals: anthropogenic noise can influence the foraging decisions by seals. *Journal of Applied Ecology*, 58(9):1854–1863.

Hastie, G. D., Russell, D. J., Lepper, P., Elliott, J., Wilson, B., Benjamins, S., and Thompson, D. (2018). Harbour seals avoid tidal turbine noise: Implications for collision risk. *Journal of Applied Ecology*, 55(2):684–693.

Hastie, G. D., Russell, D. J. F., McConnell, B., Moss, S., Thompson, D., and Janik, V. M. (2015). Sound exposure in harbour seals during the installation of an offshore wind farm: Predictions of auditory damage. *Journal of Applied Ecology*, 52(3):631–640.

Hastings, K. K., Frost, K. J., Simpkins, M. A., Pendleton, G. W., Swain, U. G., and Small, R. J. (2004). Regional differences in diving behavior of harbor seals in the Gulf of Alaska. *Canadian Journal of Zoology*, 82(11):1755–1773.

Hawkins, A. D., Roberts, L., and Cheesman, S. (2014). Responses of free-living coastal pelagic fish to impulsive sounds. *The Journal of the Acoustical Society of America*, 135(5):3101–3116.

Heggenes, J., Krog, O., Lindas, O., Dokk, J., and Bremnes, T. (1993). Homeostatic behavioural responses in a changing environment: Brown trout (Salmo trutta) become nocturnal during winter. *Journal of Animal Ecology*, 62(2):295–308.

Heithaus, M. R. and Frid, A. (2003). Optimal diving under the risk of predation. *Journal of Theoretical Biology*, 223(1):79–92.

Herbert-Read, J. E., Kremer, L., Bruintjes, R., Radford, A. N., and Ioannou, C. C. (2017). Anthropogenic noise pollution from pile-driving disrupts the structure and dynamics of fish shoals. *Proceedings of the Royal Society B: Biological Sciences*, 284(1863):20171627.

Hijmans, R. J. (2020). *raster: Geographic data analysis and modeling*. R package version 3.1-5.

Hildebrand, J. A. (2005). Impacts of Anthropogenic Sound. In Reynolds, J. E. I., Perrin, W. F., Reeves, R. R., Montgomery, S., and Ragen, T. J., editors, *Marine Mammal Research: Conservation Beyond Crisis*, chapter 7, pages 101–124. The Johns Hopkins University Press, Baltimore, Maryland.

Holton, M. D., Wilson, R. P., Teilmann, J., and Siebert, U. (2021). Animal tag technology keeps coming of age: An engineering perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1831):20200229.

Hooker, S. K., Baird, R. W., and Fahlman, A. (2009). Could beaked whales get the bends? Effect of diving behaviour and physiology on modelled gas exchange for three species: Ziphius cavirostris, Mesoplodon densirostris and Hyperoodon ampullatus. *Respiratory Physiology and Neurobiology*, 167(3):235–246.

Hooker, S. K., Biuw, M., McConnell, B. J., Miller, P. J. O., and Sparling, C. E. (2007). Bio-logging science: Logging and relaying physical and biological data using animal-attached tags. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 54:177–182.

Hooten, M. B., Johnson, D. S., McClintock, B. T., and Morales, J. M. (2017). *Animal movement: statistical models for telemetry data*. CRC Press.

Hüppop, O., Dierschke, J., Exo, K. M., Fredrich, E., and Hill, R. (2006). Bird migration and offshore wind turbines. *Ibis*, 148:90–109.

Hussey, N. E., Kessel, S. T., Aarestrup, K., Cooke, S. J., Cowley, P. D., Fisk, A. T., Harcourt, R. G., Holland, K. N., Iverson, S. J., Kocik, J. F., Flemming, J. E., and Whoriskey, F. G. (2015). Aquatic animal telemetry: A panoramic window into the underwater world. *Science*, 348(6240):1255642.

Iliffe, J., Ziebart, M., Turner, J., Talbot, A., and Lessnoff, A. (2013). Accuracy of vertical datum surfaces in coastal and offshore zones. *Survey Review*, 45(331):254–262.

Inger, R., Attrill, M. J., Bearhop, S., Broderick, A. C., James Grecian, W., Hodgson, D. J., Mills, C., Sheehan, E., Votier, S. C., Witt, M. J., and Godley, B. J. (2009). Marine renewable energy: Potential benefits to biodiversity? An urgent call for research. *Journal of Applied Ecology*, 46(6):1145–1153.

IPCC (2021). *Climate Change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Isojunno, S., Sadykova, D., DeRuiter, S., Cure, C., Visser, F., Thomas, L., Miller, P. J., and Harris, C. M. (2017). Individual, ecological, and anthropogenic influences on activity budgets of long-finned pilot whales. *Ecosphere*, 8(12):e02044.

Jara, F. G. and Perotti, M. G. (2010). Risk of predation and behavioural response in three anuran species: Influence of tadpole size and predator type. *Hydrobiologia*, 644(1):313–324.

Jensen, S. K., Lacaze, J. P., Hermann, G., Kershaw, J., Brownlow, A., Turner, A., and Hall, A. (2015). Detection and effects of harmful algal toxins in Scottish harbour seals and potential links to population decline. *Toxicon*, 97:1–14.

Johnson, D. S., London, J. M., Lea, M.-A., and Durban, J. W. (2008). Continuous-Time Correlated Random Walk Model for Animal Telemetry Data. *Ecology*, 89(5):1208–1215.

Johnson, M. P. and Tyack, P. L. (2003). A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE Journal of Oceanic Engineering*, 28(1):3–12.

Jones, E. L., Hastie, G. D., Smout, S., Onoufriou, J., Merchant, N. D., Brookes, K. L., and Thompson, D. (2017). Seals and shipping: quantifying population risk and individual exposure to vessel noise. *Journal of Applied Ecology*, 54(6):1930–1940.

Jones, I. T., Stanley, J. A., and Mooney, T. A. (2020). Impulsive pile driving noise elicits alarm responses in squid (Doryteuthis pealeii). *Marine Pollution Bulletin*, 150:110792.

Joy, R., Wood, J. D., Sparling, C. E., Tollit, D. J., Copping, A. E., and McConnell, B. J. (2018). Empirical measures of harbor seal behavior and avoidance of an operational tidal turbine. *Marine Pollution Bulletin*, 136:92–106.

Kamrowski, R. L., Limpus, C., Moloney, J., and Hamann, M. (2013). Coastal light pollution and marine turtles: Assessing the magnitude of the problem. *Endangered Species Research*, 19(1):85–98.

Kassambara, A. (2020). *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.3.0.

Kastelein, R. A., Gransier, R., Hoek, L., Macleod, A., and Terhune, J. M. (2012). Hearing threshold shifts and recovery in harbor seals (Phoca vitulina) after octave-band noise exposure at 4 kHz. *The Journal of the Acoustical Society of America*, 132(4):3525–3537.

Kastelein, R. A., Helder-Hoek, L., Kommeren, A., Covi, J., and Gransier, R. (2018a). Effect of pile-driving sounds on harbor seal (Phoca vitulina) hearing. *The Journal of the Acoustical Society of America*, 143(6):3583–3594.

Kastelein, R. A., Horvers, M., Helder-Hoek, L., Van de Voorde, S., ter Hofstede, R., and van der Meij, H. (2017). Behavioral responses of harbor seals (Phoca vitulina) to FaunaGuard seal module sounds at two background noise levels. *Aquatic Mammals*, 43(4):347–363.

Kastelein, R. A., Van de Voorde, S., and Jennings, N. (2018b). Swimming speed of a harbor porpoise (Phocoena phocoena) during playbacks of offshore pile driving sounds. *Aquatic Mammals*, 44(1):92–99.

Kastelein, R. A., van der Heul, S., Terhune, J. M., Verboom, W. C., and Triesscheijn, R. J. V. (2006). Deterring effects of 8-45 kHz tone pulses on harbour seals (Phoca vitulina) in a large pool. *Marine Environmental Research*, 62:356–373.

King, S. L., Schick, R. S., Donovan, C., Booth, C. G., Burgman, M., Thomas, L., and Harwood, J. (2015). An interim framework for assessing the population consequences of disturbance. *Methods in Ecology and Evolution*, 6(10):1150–1158.

Klett-Mingo, J. I., Pavón, I., and Gil, D. (2016). Great tits, Parus major, increase vigilance time and reduce feeding effort during peaks of aircraft noise. *Animal Behaviour*, 115:29–34.

Kline, L. R., DeAngelis, A. I., McBride, C., Rodgers, G. G., Rowell, T. J., Smith, J., Stanley, J. A., Read, A. D., and Van Parijs, S. M. (2020). Sleuthing with sound: Understanding vessel activity in marine protected areas using passive acoustic monitoring. *Marine Policy*, 120:104138.

Klinke, S. (2021). *plot.matrix: Visualizes a matrix as heatmap*. R package version 1.6.

Kooyman, G. L. (1964). Techniques used in measuring diving capacities of Weddell seals. *The Polar Record*, 12(79):391–394.

Kooyman, G. L., McDonald, B. I., Williams, C. L., Meir, J. U., and Ponganis, P. J. (2021). The aerobic dive limit: After 40 years, still rarely measured but commonly used. *Comparative Biochemistry and Physiology Part A : Molecular and Integrative Physiology*, 252:110841.

Koschinski, S., Culik, B. M., Henriksen, O. D., Tregenza, N., Ellis, G., Jansen, C., and Käthe, G. (2003). Behavioural reactions of free-ranging porpoises and seals to the noise of a simulated 2 MW windpower generator. *Marine Ecology Progress Series*, 265:263–273.

Kuhn, C. E., Crocker, D. E., Tremblay, Y., and Costa, D. P. (2009). Time to eat: Measurements of feeding behaviour in a large marine predator, the northern elephant seal Mirounga angustirostris. *Journal of Animal Ecology*, 78(3):513–523.

Kvadsheim, P. H., DeRuiter, S., Sivle, L. D., Goldbogen, J., Roland-Hansen, R., Miller, P. J., Lam, F. P. A., Calambokidis, J., Friedlaender, A., Visser, F., Tyack, P. L., Kleivane, L., and Southall, B. (2017). Avoidance responses of minke whales to 1–4 kHz naval sonar. *Marine Pollution Bulletin*, 121(1-2):60–68.

Kvadsheim, P. H., Sevaldsen, E. M., Folkow, L. P., and Blix, A. S. (2010). Behavioural and physiological responses of hooded seals (Cystophora cristata) to 1 to 7 kHz Sonar Signals. *Aquatic Mammals*, 36(3):239–247.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J. M. (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, 93(11):2336–2342.

Langrock, R., Marques, T. A., Baird, R. W., and Thomas, L. (2014). Modeling the diving behavior of whales: A latent-variable approach with feedback and semi-Markovian components. *Journal of Agricultural, Biological, and Environmental Statistics*, 19(1):82–100.

Leos-Barajas, V., Gangloff, E. J., Adam, T., Langrock, R., van Beest, F. M., Nabe-Nielsen, J., and Morales, J. M. (2017). Multi-scale modeling of animal movement and general behavior data using Hidden Markov models with hierarchical structures. *Journal of Agricultural, Biological, and Environmental Statistics*, 22(3):232–248.

Lesage, V., Hammill, M. O., and Kovacs, K. M. (1999). Functional classification of harbor seal (Phoca vitulina) dives using depth profiles, swimming velocity, and an index of foraging success. *Canadian Journal of Zoology*, 77(1):74–87.

Lonergan, M., Duck, C. D., Thompson, D., Mackey, B. L., Cunningham, L., and Boyd, I. L. (2007). Using sparse survey data to investigate the declining abundance of British harbour seals. *Journal of Zoology*, 271:261–269.

Lucke, K., Hastie, G. D., Ternes, K., McConnell, B., Moss, S., Russell, D. J., Weber, H., and Janik, V. M. (2016). Aerial low-frequency hearing in captive and free-ranging harbour seals (Phoca vitulina) measured using auditory brainstem responses. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 202(12):859–868.

Lyamin, O. I. (1993). Sleep in the harp seal (Pagophilus groenlandica). Comparison of sleep on land and in water. *Journal of sleep research*, 2(3):170–174.

Madsen, P. T., Wahlberg, M., Tougaard, J., Lucke, K., and Tyack, P. (2006). Wind turbine underwater noise and marine mammals: implications of current knowledge and data needs. *Marine Ecology Progress Series*, 309:279–295.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2019). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.0.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India*, 2(1):49–55.

Mardia, K. V. and Jupp, P. E. (1999). *Directional statistics*. John Wiley and Sons.

Marley, S. A., Salgado Kent, C. P., Erbe, C., and Parnum, I. M. (2017). Effects of vessel traffic and underwater noise on the movement, behaviour and vocalisations of bottlenose dolphins in an urbanised estuary. *Scientific Reports*, 7(1):1–14.

McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J., and Morales, J. M. (2012). A general discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs*, 82(3):335–349.

McClintock, B. T., Langrock, R., Gimenez, O., Cam, E., Borchers, D. L., Glennie, R., and Patterson, T. A. (2020). Uncovering ecological state dynamics with hidden Markov models. *Ecology Letters*, 23(12):1878–1903.

McClintock, B. T., London, J. M., Cameron, M. F., and Boveng, P. L. (2017). Bridging the gaps in animal movement: hidden behaviors and ecological relationships revealed by integrated data streams. *Ecosphere*, 8(3):e01751.

McClintock, B. T. and Michelot, T. (2018). momentuHMM: R package for generalized hidden Markov models of animal movement. *Methods in Ecology and Evolution*, 9(6):1518–1530, R package version 1.5.1.

McClintock, B. T., Russell, D. J., Matthiopoulos, J., and King, R. (2013). Combining individual animal movement and ancillary biotelemetry data to investigate population-level activity budgets. *Ecology*, 94(4):838–849.

McConnell, B. and Fedak, M. (1996). Movements of southern elephant seals. *Canadian Journal of Zoology*, 74(8):1485–1496.

McConnell, B. J., Fedak, M. A., Lovell, P., and Hammond, P. S. (1999). Movements and foraging areas of grey seals in the North Sea. *Journal of Applied Ecology*, 36(4):573–590.

McKnight, J. (2011). Short-term effects of capture and tag attachment in common seals, Phoca vitulina. *Masters Thesis, University of St Andrews*.

McKnight, J. C., Bennett, K. A., Bronkhorst, M., Russell, D. J., Balfour, S., Milne, R., Bivins, M., Moss, S. E., Colier, W., Hall, A. J., and Thompson, D. (2019). Shining new light on mammalian diving physiology using wearable near-infrared spectroscopy. *PLoS Biology*, 17(6):1–20.

McKnight, J. C., Ruesch, A., Bennett, K., Bronkhorst, M., Balfour, S., Moss, S. E., Milne, R., Tyack, P. L., Kainerstorfer, J. M., and Hastie, G. D. (2021). Shining new light on

sensory brain activation and physiological measurement in seals using wearable optical technology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1830):20200224.

McSweeney, F. K., Hinson, J. M., and Cannon, C. B. (1996). Sensitization-habituation may occur during operant conditioning. *Psychological Bulletin*, 120(2):256–271.

Mikkelsen, L., Hermannsen, L., Beedholm, K., Madsen, P. T., and Tougaard, J. (2017). Simulated seal scarer sounds scare porpoises, but not seals: species-specific responses to 12 kHz deterrence sounds. *Royal Society Open Science*, 4(170286).

Mikkelsen, L., Johnson, M., Wisniewska, D. M., van Neer, A., Siebert, U., Madsen, P. T., and Teilmann, J. (2019). Long-term sound and movement recording tags to study natural behavior and reaction to ship noise of seals. *Ecology and Evolution*, 9(5):2588–2601.

Miller, P. J., Kvadsheim, P. H., Lam, F. P., Tyack, P. L., Curé, C., DeRuiter, S. L., Kleivane, L., Sivle, L. D., Van Ijsselmuide, S. P., Visser, F., Wensveen, P. J., Von Benda-Beckmann, A. M., Martín López, L. M., Narazaki, T., and Hooker, S. K. (2015). First indications that northern bottlenose whales are sensitive to behavioural disturbance from anthropogenic noise. *Royal Society Open Science*, 2(6):140484.

Miller, P. J. O., Antunes, R. N., Wensveen, P. J., Samarra, F. I. P., Catarina Alves, A., Tyack, P. L., Kvadsheim, P. H., Kleivane, L., Lam, F.-P. A., Ainslie, M. A., and Thomas, L. (2014). Dose-response relationships for the onset of avoidance of sonar by free-ranging killer whales. *The Journal of the Acoustical Society of America*, 135(2):975–993.

Mitani, Y., Andrews, R. D., Sato, K., Kato, A., Naito, Y., and Costa, D. P. (2010). Three-dimensional resting behaviour of northern elephant seals: Drifting like a falling leaf. *Biology Letters*, 6(2):163–166.

Nabe-Nielsen, J., van Beest, F. M., Grimm, V., Sibly, R. M., Teilmann, J., and Thompson, P. M. (2018). Predicting the impacts of anthropogenic disturbances on marine populations. *Conservation Letters*, 11(5):e12563.

National Academies of Sciences Engineering and Medicine (2017). *Approaches to understanding the cumulative effects of stressors on marine mammals.* The National Academies Press, Washington, DC.

National Oceanography Centre (2020). *POLPRED Offshore Tidal Software.* URL: https://noc-innovations.co.uk/software/offshore.

Nedwell, J., Brooker, A., and Barham, R. (2011). *Measurement and assessment of underwater noise during impact piling operations at the Lincs offshore wind farm*. Subacoustech Environmental Report No. E273R0203.

Nelms, S. E., Alfaro-Shigueto, J., Arnould, J. P., Avila, I. C., Nash, S. B., Campbell, E., Carter, M. I., Collins, T., Currey, R. J., Domit, C., Franco-Trecu, V., Fuentes, M. M., Gilman, E., Harcourt, R. G., Hines, E. M., Hoelze, A. R., Hooker, S. K., Kelkar, N., Kiszka, J. J., Kiszka, J. J., Laidre, K. L., Mangel, J. C., Marsh, H., Maxwe, S. M., Onoufriou, A. B., Palacios, D. M., Pierce, G. J., Ponnampalam, L. S., Porter, L. J., Russell, D. J., Stockin, K. A., Sutaria, D., Wambiji, N., Wei, C. R., Wilson, B., and Godley, B. J. (2021). Marine mammal conservation: over the horizon. *Endangered Species Research*, 44:291–325.

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer palettes*. R package version 1.1-2.

New, L. F., Clark, J. S., Costa, D. P., Fleishman, E., Hindell, M. A., Klanjscek, T., Lusseau, D., Kraus, S., McMahon, C. R., Robinson, P. W., Schick, R. S., Schwarz, L. K., Simmons, S. E., Thomas, L., Tyack, P., and Harwood, J. (2014). Using short-term measures of behaviour to estimate long-term fitness of southern elephant seals. *Marine Ecology Progress Series*, 496:99–108.

Ngô, M. C., Heide-Jørgensen, M. P., and Ditlevsen, S. (2019). Understanding narwhal diving behaviour using hidden Markov models with dependent state distributions and long range dependence. *PLoS Computational Biology*, 15(3):e1006425.

Nieukirk, S. L., Stafford, K. M., Mellinger, D. K., Dziak, R. P., and Fox, C. G. (2004). Low-frequency whale and seismic airgun sounds recorded in the mid-Atlantic Ocean. *The Journal of the Acoustical Society of America*, 115(4):1832–1843.

O'Hara, R. B. and Sillanpää, M. J. (2009). A review of bayesian variable selection methods: What, how and which. *Bayesian Analysis*, 4(1):85–118.

Oksanen, S. M., Ahola, M. P., Lehtonen, E., and Kunnasranta, M. (2014). Using movement data of Baltic grey seals to examine foraging-site fidelity: Implications for seal-fishery conflict mitigation. *Marine Ecology Progress Series*, 507:297–308.

Onoufriou, J., Russell, D. J., Thompson, D., Moss, S. E., and Hastie, G. D. (2021). Quantifying the effects of tidal turbine array operations on the distribution of marine mammals: Implications for collision risk. *Renewable Energy*, 180:157–165.

Ordiz, A., Moen, G. K., Sæbø, S., Stenset, N., Swenson, J. E., and Støen, O. G. (2019). Habituation, sensitization, or consistent behavioral responses? Brown bear responses after repeated approaches by humans on foot. *Biological Conservation*, 232(January):228–237.

Owen, M. A., Swaisgood, R. R., and Blumstein, D. T. (2017). Contextual influences on animal decision-making: Significance for behavior-based wildlife conservation and management. *Integrative Zoology*, 12(1):32–48.

Pebesma, E. and Bivand, R. (2005). *sp: Classes and methods for spatial data in R.* R package version 1.4-5.

Peeke, H. V. (1984). *Habituation, sensitization, and behavior.* Elsevier.

Photopoulou, T., Fedak, M. A., Thomas, L., and Matthiopoulos, J. (2014). Spatial variation in maximum dive depth in gray seals in relation to foraging. *Marine Mammal Science*, 30(3):923–938.

Photopoulou, T., Heerah, K., Pohle, J., and Boehme, L. (2020). Sex-specific variation in the use of vertical habitat by a resident Antarctic top predator: Sex-specific variation in diving. *Proceedings of the Royal Society B: Biological Sciences*, 287(1937).

Photopoulou, T., Lovell, P., Fedak, M. A., Thomas, L., and Matthiopoulos, J. (2015). Efficient abstracting of dive profiles using a broken-stick model. *Methods in Ecology and Evolution*, 6(3):278–288.

Piggott, J. J., Townsend, C. R., and Matthaei, C. D. (2015). Reconceptualizing synergism and antagonism among multiple stressors. *Ecology and Evolution*, 5(7):1538–1547.

Pirotta, E., Booth, C. G., Costa, D. P., Fleishman, E., Kraus, S. D., Lusseau, D., Moretti, D., New, L. F., Schick, R. S., Schwarz, L. K., Simmons, S. E., Thomas, L., Tyack, P. L., Weise, M. J., Wells, R. S., and Harwood, J. (2018a). Understanding the population consequences of disturbance. *Ecology and Evolution*, 8(19):9934–9946.

Pirotta, E., Edwards, E. W., New, L., and Thompson, P. M. (2018b). Central place foragers and moving stimuli: A hidden-state model to discriminate the processes affecting movement. *Journal of Animal Ecology*, 87(4):1116–1125.

Planque, Y., Huon, M., Caurant, F., Pinaud, D., and Vincent, C. (2020). Comparing the horizontal and vertical approaches used to identify foraging areas of two diving marine predators. *Marine Biology*, 167(2):1–14.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*, 124(125.10):1–10.

Plummer, M. (2019). *rjags: Bayesian graphical models using MCMC*. R package version 4-10.

Pohle, J., Langrock, R., van Beest, F. M., and Schmidt, N. M. (2017). Selecting the number of states in hidden Markov models: Pragmatic solutions illustrated using animal movement. *Journal of Agricultural, Biological, and Environmental Statistics*, 22(3):270–293.

Popper, A. N. and Hastings, M. C. (2009). The effects of anthropogenic sources of sound on fishes. *Journal of Fish Biology*, 75(3):455–489.

Preisler, H. K., Ager, A. A., and Wisdom, M. J. (2013). Analyzing animal movement patterns using potential functions. *Ecosphere*, 4(3):1–13.

Quick, N. J., Isojunno, S., Sadykova, D., Bowers, M., Nowacek, D. P., and Read, A. J. (2017). Hidden Markov models reveal complexity in the diving behaviour of short-finned pilot whales. *Scientific Reports*, 7:1–12.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Radford, A. N., Lèbre, L., Lecaillon, G., Nedelec, S. L., and Simpson, S. D. (2016). Repeated exposure reduces the response to impulsive noise in European seabass. *Global change Biology*, 22(10):3349–3360.

Ramasco, V., Biuw, M., and Nilssen, K. T. (2014). Improving time budget estimates through the behavioural interpretation of dive bouts in harbour seals. *Animal Behaviour*, 94:117–134.

Riera, A., Pilkington, J. F., Ford, J. K., Stredulinsky, E. H., and Chapman, N. R. (2019). Passive acoustic monitoring off Vancouver Island reveals extensive use by at-risk Resident killer whale (Orcinus orca) populations. *Endangered Species Research*, 39:221–234.

Ries, E. H., Traut, I. M., Paffen, P., and Goedhart, P. W. (1997). Diving patterns of harbour seals (Phoca vitulina) in the Wadden Sea, the Netherlands and Germany, as indicated by VHF telemetry. *Canadian Journal of Zoology*, 75(12):2063–2068.

Riffell, J. A., Shlizerman, E., Sanders, E., Abrell, L., Medina, B., Hinterwirth, A. J., and Kutz, J. N. (2014). Flower discrimination by pollinators in a dynamic chemical environment. *Science*, 344(6191):1515–1518.

Ritchie, H. and Roser, M. (2020). *Energy: Our World in Data*. Published online at 'https://ourworldindata.org/energy' [Accessed 25-08-21].

Russell, D. (2016). Activity budgets: Analysis of seal behaviour at sea. *Report for the Department for Business, Energy and Industrial Strategy (OESEA-15-66)*.

Russell, D., Brasseur, S., Thompson, D., Hastie, G., Janik, V., Aarts, G., McClintock, B. T., Matthiopoulos, J., Moss, S. E., and McConnell, B. (2014). Marine mammals trace anthropogenic structures at sea. *Current Biology*, 24(14):R638–R639.

Russell, D., Hastie, G. D., Thompson, D., Janik, V. M., Hammond, P. S., Scott-Hayward, L. A., Matthiopoulos, J., Jones, E. L., and McConnell, B. J. (2016). Avoidance of wind farms by harbour seals is limited to pile driving activities. *Journal of Applied Ecology*, 53(6):1642–1652.

Russell, D. J. F., McClintock, B. T., Matthiopoulos, J., Thompson, P. M., Thompson, D., Hammond, P. S., Jones, E. L., MacKenzie, M. L., Moss, S., and McConnell, B. J. (2015). Intrinsic and extrinsic drivers of activity budgets in sympatric grey and harbour seals. *Oikos*, 124(11):1462–1472.

Rutz, C. and Hays, G. C. (2009). New frontiers in biologging science. *Biology Letters*, 5(3):289–292.

Schaffeld, T., Schnitzler, J. G., Ruser, A., Woelfing, B., Baltzer, J., and Siebert, U. (2020). Effects of multiple exposures to pile driving noise on harbor porpoise hearing during simulated flights—An evaluation tool. *The Journal of the Acoustical Society of America*, 147(2):685–697.

Schick, R. S., New, L. F., Thomas, L., Costa, D. P., Hindell, M. A., Mcmahon, C. R., Robinson, P. W., Simmons, S. E., Thums, M., Harwood, J., and Clark, J. S. (2013). Estimating resource acquisition and at-sea body condition of a marine predator. *Journal of Animal Ecology*, 82(6):1300–1315.

Sharples, R. J., Moss, S. E., Patterson, T. A., and Hammond, P. S. (2012). Spatial variation in foraging behaviour of a marine top predator (Phoca vitulina) determined by a large-scale satellite tagging program. *PLoS ONE*, 7(5):e37216.

Sih, A. (2013). Understanding variation in behavioural responses to human-induced rapid environmental change: A conceptual overview. *Animal Behaviour*, 85(5):1077–1088.

Silva, W. T., Bottagisio, E., Härkönen, T., Galatius, A., Olsen, M. T., and Harding, K. C. (2021). Risk for overexploiting a seemingly stable seal population: influence of multiple stressors and hunting. *Ecosphere*, 12(1).

Simonis, A. E., Brownell, R. L., Thayre, B. J., Trickey, J. S., Oleson, E. M., Huntington, R., and Baumann-Pickering, S. (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(1921):20200070.

Širović, A., Hildebrand, J. A., and Wiggins, S. M. (2007). Blue and fin whale call source levels and propagation range in the Southern Ocean. *The Journal of the Acoustical Society of America*, 122(2):1208–1215.

Sivle, L. D., Kvadsheim, P. H., Fahlman, A., Lam, F. P., Tyack, P. L., and Miller, P. J. (2012). Changes in dive behavior during naval sonar exposure in killer whales, long-finned pilot whales, and sperm whales. *Frontiers in Physiology*, 3:400.

Skeate, E. R., Perrow, M. R., and Gilroy, J. J. (2012). Likely effects of construction of Scroby Sands offshore wind farm on a mixed population of harbour Phoca vitulina and grey Halichoerus grypus seals. *Marine Pollution Bulletin*, 64(4):872–881.

Slabbekoorn, H., Bouton, N., van Opzeeland, I., Coers, A., ten Cate, C., and Popper, A. N. (2010). A noisy spring: The impact of globally rising underwater sound levels on fish. *Trends in Ecology and Evolution*, 25(7):419–427.

Smith, J. A., Gaynor, K. M., and Suraci, J. P. (2021). Mismatch between risk and response may amplify lethal and non-lethal effects of humans on wild animal populations. *Frontiers in Ecology and Evolution*, 9:604973.

Smith, L. C. and Stephenson, S. R. (2013). New Trans-Arctic shipping routes navigable by midcentury. *Proceedings of the National Academy of Sciences of the United States of America*, 110(13):E1191–E1195.

Solan, M., Hauton, C., Godbold, J. A., Wood, C. L., Leighton, T. G., and White, P. (2016). Anthropogenic sources of underwater sound can modify how sediment-dwelling invertebrates mediate ecosystem properties. *Scientific Reports*, 6(1):20540.

Song, P. X. K. (2000). Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics*, 27(2):305–320.

Southall, B. L., DeRuiter, S. L., Friedlaender, A., Stimpert, A. K., Goldbogen, J. A., Hazen, E., Casey, C., Fregosi, S., Cade, D. E., Allen, A. N., Harris, C. M., Schorr, G., Moretti, D., Guan, S., and Calambokidis, J. (2019a). Behavioral responses of individual blue whales (Balaenoptera musculus) to mid-frequency military sonar. *Journal of Experimental Biology*, 222(5).

Southall, B. L., Finneran, J. J., Reichmuth, C., Nachtigall, P. E., Ketten, D. R., Bowles, A. E., Ellison, W. T., Nowacek, D. P., and Tyack, P. L. (2019b). Marine mammal noise exposure criteria: Updated scientific recommendations for residual hearing effects. *Aquatic Mammals*, 45(2):125–232.

Southall, B. L., Nowacek, D. P., Bowles, A. E., Senigaglia, V., Bejder, L., and Tyack, P. L. (2021). Marine mammal noise exposure criteria: Assessing the severity of marine mammal behavioral responses to human noise. *Aquatic Mammals*, 47(5):421–464.

Sparling, C., Lonergan, M., and McConnell, B. (2018). Harbour seals (Phoca vitulina) around an operational tidal turbine in Strangford Narrows: No barrier effect but small changes in transit behaviour. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 28(1):194–204.

Stafford, K. M., Fox, C. G., and Clark, D. S. (1998). Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean. *The Journal of the Acoustical Society of America*, 104(6):3616–3625.

Stevenson, C. D., Ferryman, M., Nevin, O. T., Ramsey, A. D., Bailey, S., and Watts, K. (2013). Using GPS telemetry to validate least-cost modeling of gray squirrel (Sciurus carolinensis) movement within a fragmented landscape. *Ecology and Evolution*, 3(7):2350–2361.

Stimpert, A. K., DeRuiter, S. L., Southall, B. L., Moretti, D. J., Falcone, E. A., Goldbogen, J. A., Friedlaender, A., Schorr, G. S., and Calambokidis, J. (2014). Acoustic and foraging behavior of a Baird's beaked whale, Berardius bairdii, exposed to simulated sonar. *Scientific Reports*, 4(1):1–8.

Talling, J. C., Waran, N. K., Wathes, C. M., and Lines, J. A. (1996). Behavioural and physiological responses of pigs to sound. *Applied Animal Behaviour Science*, 48(3-4):187–201.

Teilmann, J. and Galatius, A. (2018). Harbor seal (Phoca vitulina). In Wursig, B., Thewissen, J., and Kovacs, K., editors, *Encyclopedia of Marine Mammals*, pages 451–454. Elsevier Science and Technology, 3rd edition.

The United Kingdom Government (2020). New plans to make UK world leader in green energy - GOV.UK. *Department for Business- Gov.UK [Available at https://www.gov.uk/government/news/new-plans-to-make-uk-world-leader-in-green-energy]*.

Thieurmel, B. and Elmarhraoui, A. (2019). *suncalc: Compute sun position, sunlight phases, moon position and lunar phase*. R package version 0.5.0.

Thomas, L., Russell, D. J., Duck, C. D., Morris, C. D., Lonergan, M., Empacher, F., Thompson, D., and Harwood, J. (2019). Modelling the population size and dynamics of the British grey seal. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 29(S1):6–23.

Thompson, D., Duck, C. D., Morris, C. D., and Russell, D. J. (2019). The status of harbour seals (Phoca vitulina) in the UK. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 29(S1):40–60.

Thompson, D. and Fedak, M. A. (2001). How long should a dive last? A simple model of foraging decisions by breath-hold divers in a patchy environment. *Animal Behaviour*, 61(2):287–296.

Thompson, D., Hammond, P. S., Nicholas, K. S., and Fedak, M. A. (1991). Movements, diving and foraging behaviour of grey seals (Halichoerus grypus). *Journal of Zoology*, 224(2):223–232.

Thompson, P. M., Hastie, G. D., Nedwell, J., Barham, R., Brookes, K. L., Cordes, L. S., Bailey, H., and McLean, N. (2013). Framework for assessing impacts of pile-driving noise from offshore wind farm construction on a harbour seal population. *Environmental Impact Assessment Review*, 43:73–85.

Thompson, P. M., Miller, D., Cooper, R., and Hammond, P. S. (1994). Changes in the distribution and activity of female harbour seals during the breeding season : implications for their lactation strategy and mating patterns. *Journal of Animal Ecology*, 63(1):24–30.

Thompson, P. M., Tollit, D. J., Wood, D., Corpe, H. M., Hammond, P. S., and Mackay, A. (1997). Estimating harbour seal abundance and status in an estuarine habitat in north-east Scotland. *The Journal of Applied Ecology*, 34(1):43–52.

Tigas, L. A., Van Vuren, D. H., and Sauvajot, R. M. (2002). Behavioral responses of bobcats and coyotes to habitat fragmentation and corridors in an urban environment. *Biological Conservation*, 108(3):299–306.

Tollit, D. J., Black, A. D., Thompson, P. M., Mackay, A., Corpe, H. M., Wilson, B., Van Parijs, S. M., Grellier, K., and Parlane, S. (1998). Variations in harbour seal Phoca vitulina diet and dive-depths in relation to foraging habitat. *Journal of Zoology*, 244(2):209–222.

Tougaard, J., Madsen, P. T., and Wahlberg, M. (2008). Underwater noise from construction and operation of offshore wind farms. *Bioacoustics*, 17(1-3):143–146.

Tracey, J. A., Zhu, J., and Crooks, K. (2005). A set of nonlinear regression models for animal movement in response to a single landscape feature. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(1):1–18.

Trigg, L. E., Chen, F., Shapiro, G. I., Ingram, S. N., Vincent, C., Thompson, D., Russell, D. J. F., Carter, M. I. D., and Embling, C. B. (2020). Predicting the exposure of diving grey seals to shipping noise. *The Journal of the Acoustical Society of America*, 148(2):1014–1029.

Tripovich, J. S., Hall-Aspland, S., Charrier, I., and Arnould, J. P. Y. (2012). The behavioural response of australian fur seals to motor boat noise. *PLoS ONE*, 7(5):e37228.

Tyack, P. L., Johnson, M., Soto, N. A., Sturlese, A., and Madsen, P. T. (2006). Extreme diving of beaked whales. *Journal of Experimental Biology*, 209(21):4238–4253.

Tyack, P. L. and Thomas, L. (2019). Using dose–response functions to improve calculations of the impact of anthropogenic noise. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 29(S1):242–253.

Tyack, P. L., Zimmer, W. M. X., Moretti, D., Southall, B. L., Claridge, D. E., Durban, J. W., Clark, C. W., D'Amico, A., DiMarzio, N., Jarvis, S., McCarthy, E., Morrissey, R., Ward, J., and Boyd, I. L. (2011). Beaked whales respond to simulated and actual navy sonar. *PLoS ONE*, 6(3):e17009.

van Beest, F. M., Mews, S., Elkenkamp, S., Schuhmann, P., Tsolak, D., Wobbe, T., Bartolino, V., Bastardie, F., Dietz, R., von Dorrien, C., Galatius, A., Karlsson, O., McConnell, B., Nabe-Nielsen, J., Olsen, M. T., Teilmann, J., and Langrock, R. (2019).

Classifying grey seal behaviour in relation to environmental variability and commercial fishing activity - a multivariate hidden Markov model. *Scientific Reports*, 9(1):1–14.

van Beest, F. M., Teilmann, J., Hermannsen, L., Galatius, A., Mikkelsen, L., Sveegaard, S., Balle, J. D., Dietz, R., and Nabe-Nielsen, J. (2018). Fine-scale movement responses of free-ranging harbour porpoises to capture, tagging and short-term noise pulses from a single airgun. *Royal Society Open Science*, 5(1).

Van Parijs, S. M., Thompson, P. M., Tollit, D. J., and Mackay, A. (1997). Distribution and activity of male harbour seals during the mating season. *Animal Behaviour*, 54(1):35–43.

Vance, H. M., Hooker, S. K., Mikkelsen, L., van Neer, A., Teilmann, J., Siebert, U., and Johnson, M. (2021). Drivers and constraints on offshore foraging in harbour seals. *Scientific Reports*, 11(1):1–14.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., and Watson, J. E. (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature Communications*, 7:1–11.

Verfuss, U., Sinclair, R., and Sparling, C. (2019). A review of noise abatement systems for offshore wind farm construction noise, and the potential for their application in Scottish waters. *Scottish Natural Heritage Research Report No. 1070*.

Volpov, B. L., Hoskins, A. J., Battaile, B. C., Viviant, M., Wheatley, K. E., Marshall, G., Abernathy, K., and Arnould, J. P. (2015). Identification of prey captures in Australian fur seals (Arctocephalus pusillus doriferus) using head-mounted accelerometers: Field validation with animal-borne video cameras. *PLoS ONE*, 10(6):1–19.

Walsh, E. P., Arnott, G., and Kunc, H. P. (2017). Noise affects resource assessment in an invertebrate. *Biology Letters*, 13(4):20170098.

Walther, G. R. (2010). Community and ecosystem responses to recent climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1549):2019–2024.

Watanabe, Y. Y., Baranov, E. A., and Miyazaki, N. (2015). Drift dives and prolonged surfacing periods in Baikal seals: Resting strategies in open waters? *Journal of Experimental Biology*, 218(17):2793–2798.

Webster, M. M. and Rutz, C. (2020). How STRANGE are your study animals? *Nature*, 582(7812):337–340.

Weilgart, L. (2007). The impacts of anthropogenic ocean noise on cetaceans and implications for management. *Canadian Journal of Zoology*, 85(11):1091–1116.

Wensveen, P. J., Isojunno, S., Hansen, R. R., von Benda-Beckmann, A. M., Kleivane, L., van IJsselmuide, S., Lam, F. P. A., Kvadsheim, P. H., DeRuiter, S. L., Curé, C., Narazaki, T., Tyack, P. L., and Miller, P. J. (2019). Northern bottlenose whales in a pristine environment respond strongly to close and distant navy sonar signals. *Proceedings of the Royal Society B*, 286(1899):20182592.

Whyte, K. F., Russell, D. J. F., Sparling, C. E., Binnerts, B., and Hastie, G. D. (2020). Estimating the effects of pile driving sounds on seals: Pitfalls and possibilities. *The Journal of the Acoustical Society of America*, 147(6):3948–3958.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, R package version 3.3.0.

Wickham, H., François, R., Henry, L., and Müller, K. (2020). *dplyr: A Grammar of Data Manipulation*. R package version 0.8.5.

Wilmers, C. C., Nickel, B., Bryce, C. M., Smith, J. A., Wheat, R. E., Yovovich, V., and Hebblewhite, M. (2015). The golden age of bio-logging: How animal-borne sensors are advancing the frontiers of ecology. *Ecology*, 96(7):1741–1753.

Wilson, K., Lance, M., Jeffries, S., and Acevedo-Gutiérrez, A. (2014). Fine-scale variability in harbor seal foraging behavior. *PLoS ONE*, 9(4):e92838.

Wilson, L. J. and Hammond, P. S. (2019). The diet of harbour and grey seals around Britain: Examining the role of prey as a potential cause of harbour seal declines. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 29(S1):71–85.

Wilson, M. W., Ridlon, A. D., Gaynor, K. M., Gaines, S. D., Stier, A. C., and Halpern, B. S. (2020). Ecological impacts of human-induced animal behaviour change. *Ecology Letters*, 23(10):1522–1536.

WindEurope (2017). A breakthrough for offshore wind: world's first floating wind farm opens in Scotland.

WindEurope (2021). *Wind energy in Europe: 2020 statistics and the outlook for 2021-2025*. Brussels, Belgium.

Yang, C.-M., Liu, Z.-W., Lü, L.-G., Yang, G.-B., Huang, L.-F., and Jiang, Y. (2018). Observation and comparison of tower vibration and underwater noise from offshore operational wind turbines in the East China Sea Bridge of Shanghai. *The Journal of the Acoustical Society of America*, 144(6):EL522–EL527.

Yoda, K., Sato, K., Niizuma, Y., Kurita, M., Bost, C. A., Le Maho, Y., and Naito, Y. (1999). Precise monitoring of porpoising behaviour of Adelie penguins determined using acceleration data loggers. *Journal of Experimental Biology*, 202(22):3121–3126.

Yoshino, K., Takahashi, A., Adachi, T., Costa, D. P., Robinson, P. W., Peterson, S. H., Huückstädt, L. A., Holser, R. R., and Naito, Y. (2020). Acceleration-triggered animal-borne videos show a dominance of fish in the diet of female northern elephant seals. *Journal of Experimental Biology*, 223(5):1–9.

Zoratto, F., Manzari, L., Oddi, L., Pinxten, R., Eens, M., Santucci, D., Alleva, E., and Carere, C. (2014). Behavioural response of European starlings exposed to video playback of conspecific flocks: Effect of social context and predator threat. *Behavioural Processes*, 103:269–277.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R*. Monographs on Statistics and Applied Probability (Vol 50), CRC Press, 2nd edition.

Appendix A: R functions

Simulate seal tracks `sealtrack_sim.R`

Description

Simulate multi-state correlated random walks for seal movement, with optional attraction and repel locations.

Usage

```
sealtrack_sim(n, n_states, time_res, tpm, repel_tpm=NULL, switch_strength,
    step_params, repel_step_params=NULL, turn_params, home=NULL,
    attract_centres=NULL, repel_centre=NULL, repel_times=NULL)
```

Arguments

`n`	the number of seal locations to simulate
`n_states`	the number of seal movement behaviour states to simulate
`time_res`	the time resolution of simulated data, specified as the time gap (in minutes) between simulated locations
`tpm`	a matrix containing the transition probabilities between the states during normal behaviour
`repel_tpm`	optional. The transition probability matrix between the states during repel times
`switch_strength`	a single number determining the likelihood of switching from state 1 to state 2 as an attraction point becomes closer, determined by a logistic function of distance. Larger numbers increase switching.
`step_params`	a list with a numbered entry for the step parameters for each state. Each state entry has a named list containing `mean` and `sd` for a Gamma distribution, and `corr` (0–1) for correlated movement.
`repel_step_params`	same as `step_params`, but optional specification for step parameters during repel times
`turn_params`	a list with a numbered entry for the turn parameters for each state. Each state entry has a named list containing `location` (initial average movement direction, by default set to 0) and `rho` (concentration parameter for wrapped Cauchy distribution of turn angles).

`home`	a named list containing the location of the animal's first attraction centre (`x` and `y` coordinates), `strength` (attraction source strength, which affects movement direction), and `gamma_a` (concentration constant for attraction, which affects variability in movement direction)
`attract_centres`	a numbered list of any additional attraction centres. Each entry is a named list with coordinates `x` and `y`, and attraction parameters `strength` and `gamma_a` (see description for `home`) for each attraction centre
`repel_centre`	an optional named list containing `distance` (initial distance between seal and repel centre when generated), `strength` (repel source strength, which affects movement speed and direction), and `gamma_r` (concentration constant for repel, which affects variability in movement direction)
`repel_times`	an optional named list containing `start` and `end` (the start and end time (in seconds) of the repel centre being active). If this is unspecified a repel centre will not be simulated.

Output(s)

A dataframe containing simulated movement data, where each row has a recorded `time`, simulated `x` and `y` location, and `response` indicator column (1/0). Key parameter values used and additional outputs generated in simulation (e.g. step, turn, state, attraction/repel strengths) are also saved for each timepoint.

Simulate seal dives `sealdive_sim.R`

Description

Simulate multi-state seal dive biologging data, with optional behavioural response dives.

Usage

```
sealdive_sim(n, n_states, startstate, tpm, par, response_dives=NULL,
    response_tpm=NULL, response_par=NULL, tagdivethreshold=1.5,
    bottompercentile=0.85)
```

Arguments

`n`	the number of seal dives to simulate
`n_states`	the number of dive types (states) to simulate
`startstate`	the initial dive state
`tpm`	a matrix containing the transition probabilities between the states during normal behaviour
`par`	a numbered list containing the parameters describing the dive metrics for each state. Each entry is a named list containing `shape`, `rate`, and `corr`. `shape` and `rate` are vectors containing the shape and rate parameters respectively for a Gamma distribution, where each entry in the vector is a different dive metric: (1) maximum dive depth (metres), (2) duration of descent phase (seconds), (3) duration of bottom phase (seconds), (4) duration of ascent phase (seconds), (5) duration of post-dive surface interval (seconds). `corr` is a $5x5$ matrix containing the correlation between each of the dive metrics.
`response_dives`	an optional vector with two entries containing (1) the dive number when the response starts and (2) the dive number when the response ends. If this is unspecified a response will not be simulated.
`response_tpm`	an optional transition probability matrix for state switching during the response dives
`response_par`	an optional numbered list containing the parameters describing the dive metrics for each state during a response. This should be specified as detailed in `par`.

`tagdivethreshold`	the minimum depth threshold (metres) at which dives begin being recorded from the tag. This is used to calculate the additional derived dive metrics of descent rate, ascent rate, and proportion of dive in the bottom phase. Default: 1.5.
`bottompercentile`	the proportion of the maximum dive depth used as a cutoff for determining the bottom phase of a dive. Default: 0.85.

Output(s)

A dataframe containing simulated dive data, where each row is a dive. The columns are `divenum` (dive number), `max_dep` (maximum depth of dive), `dur_des` (duration of descent phase), `dur_bot` (duration of bottom phase), `dur_asc` (duration of ascent phase), `dur_surf` (duration of post-dive surface phase), `state` (behavioural state), `response` (response indicator, 1/0), `total_divedur` (sum of descent, bottom and ascent phases), `total_eventdur` (sum of descent, bottom, ascent and surface phases), `bottomthreshold` (the dive depth at which the bottom phase began for that dive), `prop_bot` (the proportion of the dive in the bottom phase), `des_rate` (descent rate), `asc_rate` (ascent rate), `time_start` (time in seconds at which descent phase started), and `time_end` (time in seconds at which surface period ended).

Calculate Mahalanobis distance for location data `mdist_xy.R`

Description

Calculate Mahalanobis distance for location data, with different user-specified options.

Usage

```
mdist_xy(data, inputparams, timecol, width, overlap=0, consec=FALSE,
    cum_sum=FALSE, baseline_start=0, baseline_end=max(data[,timecol]),
    baseline_cov=TRUE, baseline_upload=FALSE, baseline_upload_S,
    baseline_upload_data)
```

Arguments

`data`	a dataframe containing animal locations and movement metrics. Each row must contain a time reference (see `timecol`).
`inputparams`	a vector containing the column names of `data` with the input parameters to use for the Mahalanobis distance calculation
`timecol`	the column name of `data` that contains the timestamp of each observation (in seconds)
`width`	the size of the time window (in minutes) to use for each comparison
`overlap`	the amount of overlap (in minutes) between sequential comparison time windows. Default: 0.
`consec`	logical. If FALSE, Mahalanobis distance is calculated between sliding comparison windows and a fixed baseline. If TRUE, Mahalanobis distance is calculated between a pair of consecutive sliding windows. Default: FALSE.
`cum_sum`	logical. If TRUE, Mahalanobis distance values are added cumulatively over time. Default: FALSE.
`baseline_start`	start time (in seconds) of the baseline period. If not specified, it will be assumed to be 0 (start of record).
`baseline_end`	end time (in seconds) of the baseline period, If not specified, the entire dataset will be used as baseline.

`baseline_cov`	logical. If TRUE, the covariance matrix used in Mahalanobis distance calculations is calculated over the baseline data only. If FALSE, the covariance matrix used is calculated over the entire dataset. Default: TRUE.
`baseline_upload`	logical. If TRUE, baseline data must be uploaded separately using `baseline_upload_S` and `baseline_upload_data` If this is used then `baseline_start` and `baseline_end` are discarded. Default: FALSE.
`baseline_upload_S`	optional. The covariance matrix to use for the baseline period. If this is provided then `baseline_cov` is discarded. If this is missing and `baseline_upload`=TRUE, the covariance matrix will be calculated using the data uploaded by `baseline_upload_data`.
`baseline_upload_data`	optional. A dataframe of baseline movement data matching the columns of `data`.

Output(s)

A dataframe containing `t` (time of comparison window in seconds) and `dist` (calculated Mahalanobis distance value).

Calculate Mahalanobis distance for dive data `mdist_dive.R`

Description

Calculate Mahalanobis distance for dive records, with different user-specified options.

Usage

```
mdist_dive(data, inputparams, starttimecol, endtimecol, width=1,
    overlap=0, consec=FALSE, cum_sum=FALSE, baseline_start=0,
    baseline_end=max(data[,timecol]), baseline_cov=TRUE, include_all=TRUE,
    baseline_upload=FALSE, baseline_upload_S, baseline_upload_data,
    max_gap_interdive=60*60, max_gap_consecwindows=60*60)
```

Arguments

`data`	a dataframe containing animal dive data and metrics. Each row must be a dive and contain a time reference (see `starttimecol` and `endtimecol`).
`inputparams`	a vector containing the column names of `data` with the input parameters to use for the Mahalanobis distance calculation
`starttimecol`	the column name of `data` that contains the starting time of each dive observation (in seconds)
`endtimecol`	the column name of `data` that contains the end time of each dive observation (in seconds). This should include any post-dive surface interval.
`width`	the size of the time window (in number of dives) to use for each comparison. Default: 1.
`overlap`	the amount of overlap (in number of dives) between sequential comparison time windows. Default: 0.
`consec`	logical. If FALSE, Mahalanobis distance is calculated between sliding comparison windows and a fixed baseline. If TRUE, Mahalanobis distance is calculated between a pair of consecutive sliding windows. Default: FALSE.
`cum_sum`	logical. If TRUE, Mahalanobis distance values are added cumulatively over time. Default: FALSE.
`baseline_start`	start time (in seconds) of the baseline period. If not specified, it will be assumed to be 0 (start of record).

`baseline_end`	end time (in seconds) of the baseline period, If not specified, the entire dataset will be used as baseline.
`baseline_cov`	logical. If TRUE, the covariance matrix used in Mahalanobis distance calculations is calculated over the baseline data only. If FALSE, the covariance matrix used is calculated over the entire dataset. Default: TRUE.
`include_all`	logical. If TRUE, include all dives in "exposure" period that overlap with any exposure. If FALSE, only include dives that begin after the "exposure" begins. Dives that are not included in exposure are considered baseline. Default: TRUE.
`baseline_upload`	logical. If TRUE, baseline data must be uploaded separately using `baseline_upload_S` and `baseline_upload_data` If this is used then `baseline_start` and `baseline_end` are discarded. Default: FALSE.
`baseline_upload_S`	optional. The covariance matrix to use for the baseline period. If this is provided then `baseline_cov` is discarded. If this is missing and `baseline_upload`=TRUE, the covariance matrix will be calculated using the data uploaded by `baseline_upload_data`.
`baseline_upload_data`	optional. A dataframe of baseline movement data matching the columns of `data`.
`max_gap_interdive`	the maximum time gap allowed (in seconds) between recorded dives within the same time window (between a surface period ending and a new dive starting). Default: 1 hour.
`max_gap_consecwindows`	the maximum time gap allowed (in seconds) between consecutive time windows. This is only applicable when `consec`=TRUE. Default: 1 hour.

Output(s)

A dataframe containing `t` (time of window in seconds, midpoint), `dist` (calculated Mahalanobis distance), `window2startdive` (number of the first dive in the comparison window), `window2enddive` (number of the last dive in the comparison window), `t_start` (start time of the window, in seconds), and `t_end` (end time of the window, in seconds).

Calculate movement metrics from location data `get_speedheading.R`

Description

For a sequence of animal location coordinates, calculate key movement metrics.

Usage

```
get_speedheading(data, time, locations_x, locations_y, speedmethod,
    headingmethod, headingvarmethod, headingvartype, rar_location=NULL)
```

Arguments

`data`	a dataframe containing raw animal location data. Each row must contain a time-referenced location coordinate (see `time`, `locations_x`, `locations_y`).
`time`	the column name of `data` containing the timestamp of the locations (in seconds)
`locations_x`	the column name of `data` containing the x-coordinate of the location (easting in metres)
`locations_y`	the column name of `data` containing the y-coordinate of the location (northing in metres)
`speedmethod`	logical: whether to calculate speed averaged over three locations ("average"), or relative to the previous location only ("previous"; i.e. over two locations). If "average" the estimated speed for location t is the speed between locations $t-1$ and $t+1$. If "previous" the estimated speed for location t is the speed between locations $t-1$ and t.
`headingmethod`	logical: whether to calculate heading direction over three locations ("average"), or relative to the previous location only ("previous"; i.e. over two locations). If "average" the heading direction at location t is the average easting and northing between locations $t-1$ and $t+1$. If "previous" the heading direction at location t is the easting and northing between locations $t-1$ and t.

`headingvarmethod`	logical: whether to calculate heading variation averaged over previous and future locations ("average"), or relative to the previous locations only ("previous"). If "average" the heading variation at location t is the variation from locations $t-1$ to $t+1$. If "previous" the heading variation at location t is the variation from locations $t-2$ to t.
`headingvartype`	logical: whether to calculate hearing variability as circular variance in heading ("circvar") or turn angle ("turnangle")
`rar_location`	optional. To calculate radial avoidance rate (rate at which an animal moves away from a particular location; see Chapter 3), provide a named list with `x` (x-coordinate of the location) and `y` (y-coordinate of the location).

Output(s)

The original location dataframe with additional columns as follows: `speed` (speed between locations, in $m.s^{-1}$), `acc` (acceleration, positive or negative, in $m.s^{-2}$), `absacc` (absolute value of acceleration, in $m.s^{-2}$), `easting` (easting component of heading direction), `northing` (northing component of heading direction), `headingvar` (heading variation, via method specified by user in `headingvartype`), and (optionally) `rar` (radial avoidance rate).

Calculate dive phases and metrics from dive records `get_divephases.R`

Description

For each dive, use interpolation between the recorded depths to calculate the phases of the dive and derive key metrics.

Usage

```
get_divephases(divedf, depthlist, divedur, tagdivethreshold=1.5,
    bottompercentile=0.85)
```

Arguments

`divedf`	a dataframe containing raw dive data. Each row must be a dive record.
`depthlist`	a vector containing (in temporal order) the column names of `divedf` with the recorded depth points (in metres)
`divedur`	the column name of `divedf` which contains the dive durations (in seconds)
`tagdivethreshold`	the minimum depth threshold (metres) at which dives begin being recorded from the tag. Default: 1.5.
`bottompercentile`	the proportion of the maximum dive depth used as a cutoff for determining the bottom phase of a dive. Default: 0.85.

Output(s)

The original dive dataframe with additional columns as follows: `dur_des` (duration of descent phase, in seconds), `dur_bot` (duration of bottom phase, in seconds), `dur_asc` (duration of ascent phase, in seconds), `bottompercentile` (the proportion of the maximum dive depth used to determine the bottom phase), `bottomthreshold` (the dive depth at which the bottom phase began for that dive, in metres), `maxdrecord` (the maximum of the depth points given in `depthlist`, in metres), `des_rate` (descent rate, in $m.s^{-1}$), `asc_rate` (ascent rate, in $m.s^{-1}$), and `prop_bot` (the proportion of the dive time in the bottom phase).

Appendix B: Supplementary material for Chapter 2

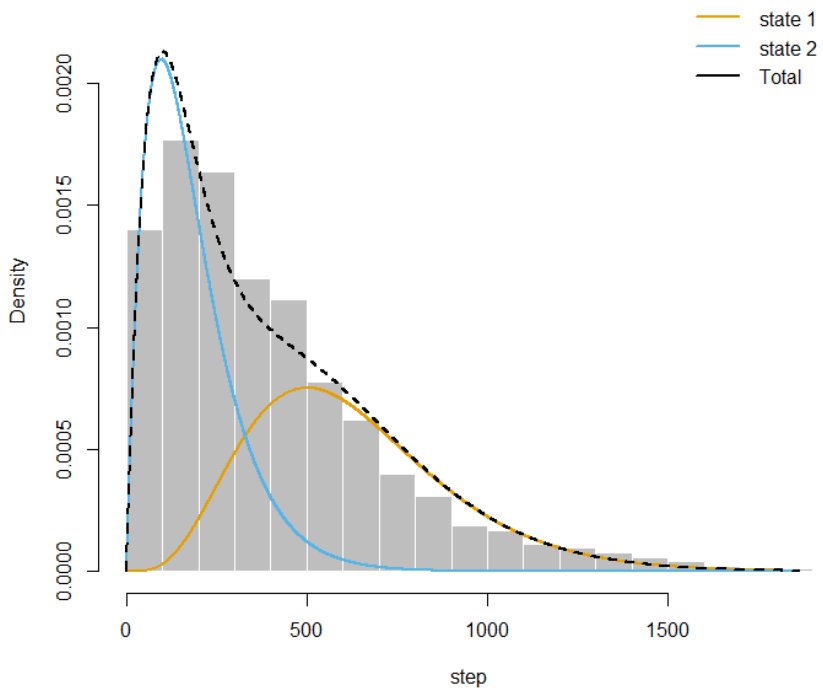


Figure S1 Histogram of seal step lengths in metres (grey bars) with estimated state-dependent distributions for each type of movement behaviour (solid lines) for individual pv42-291-12. Step lengths were fit to a Gamma distribution.

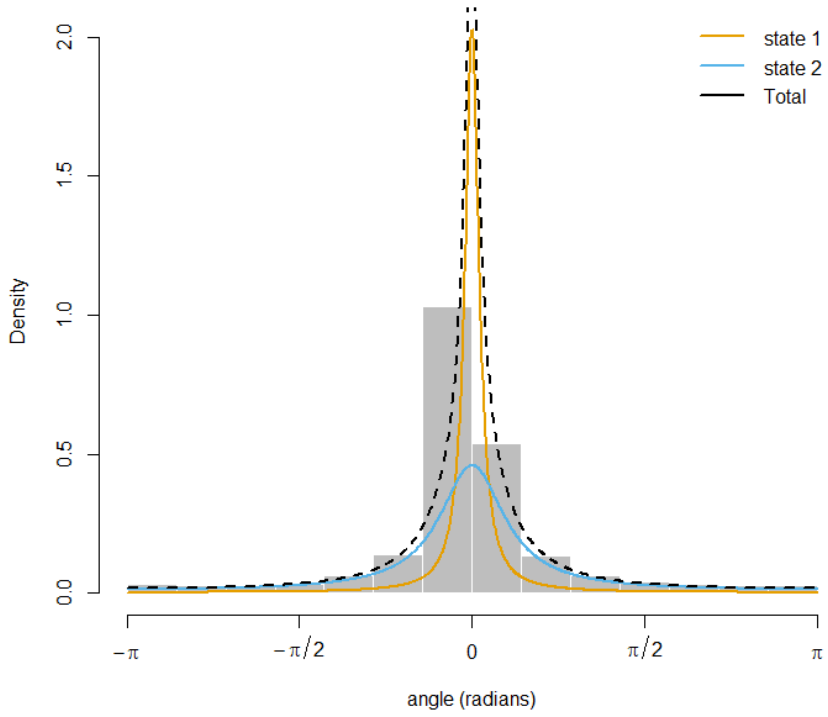


Figure S2 Histogram of seal turn angles (grey bars) with estimated state-dependent distributions for each type of movement behaviour (solid lines) for individual pv42-291-12. Turn angles were fit to a wrapped Cauchy distribution with mean zero.

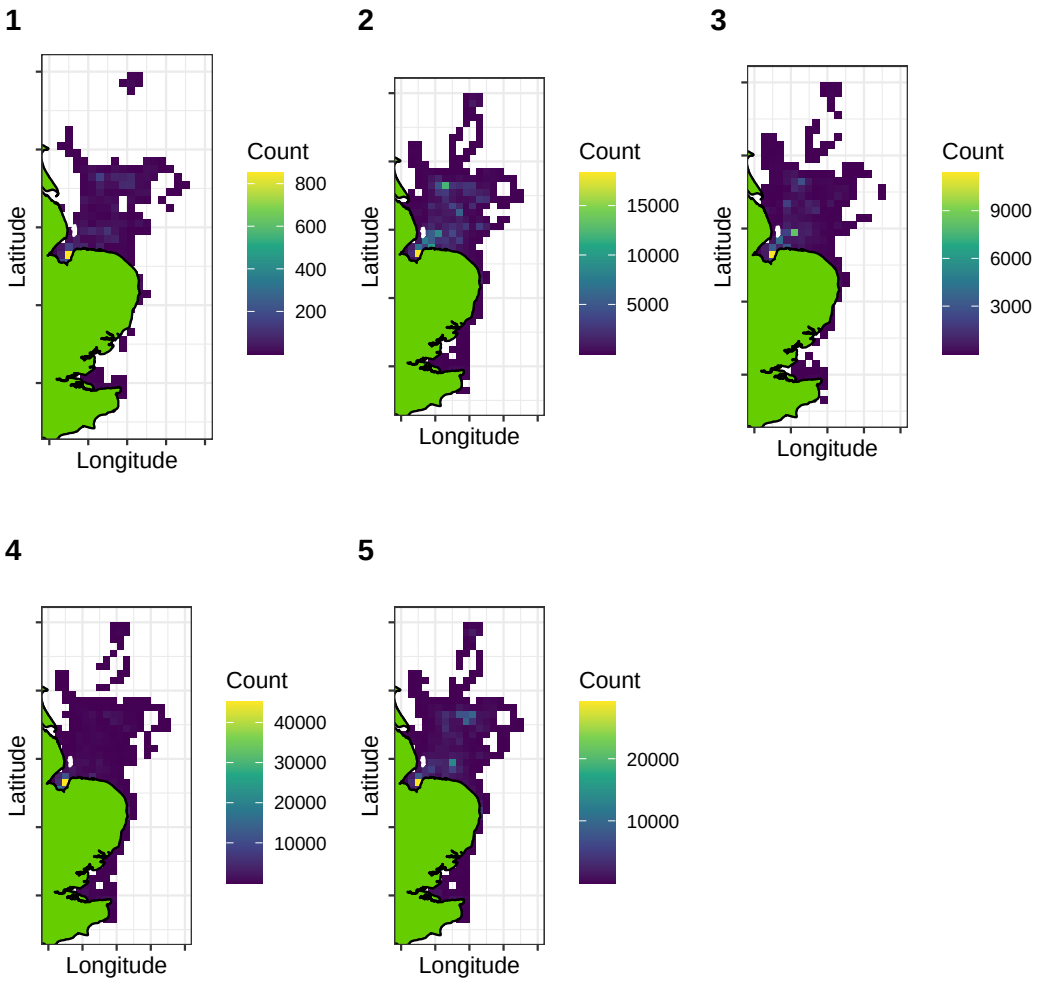


Figure S3 Original locations of each identified dive type (1–5). Each coloured square denotes the count of the number of dives of that type within each 10 x 10 km grid square.

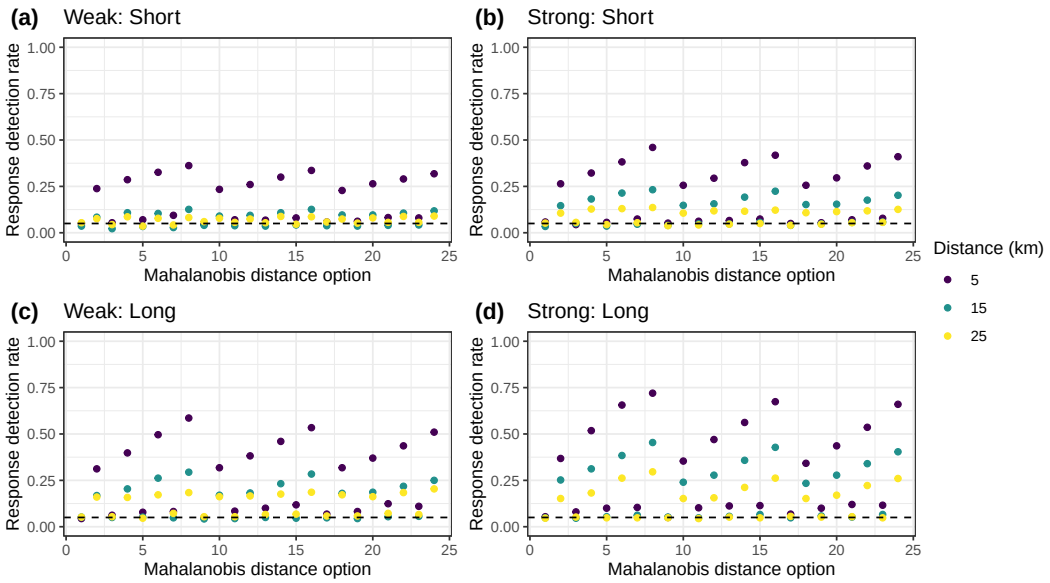


Figure S4 Response detection rates for horizontal simulated responses, with the covariance matrix calculated over all data for the simulated animal. Each dot is the response detection rate (out of 500 simulations) for responses to centres of repulsion of different strengths (weak, strong), durations (short, long), and initial distances (5, 15, 25 km; coloured dots). All simulations had 20 days of baseline data. Each Mahalanobis distance option is detailed in Table 2.6.

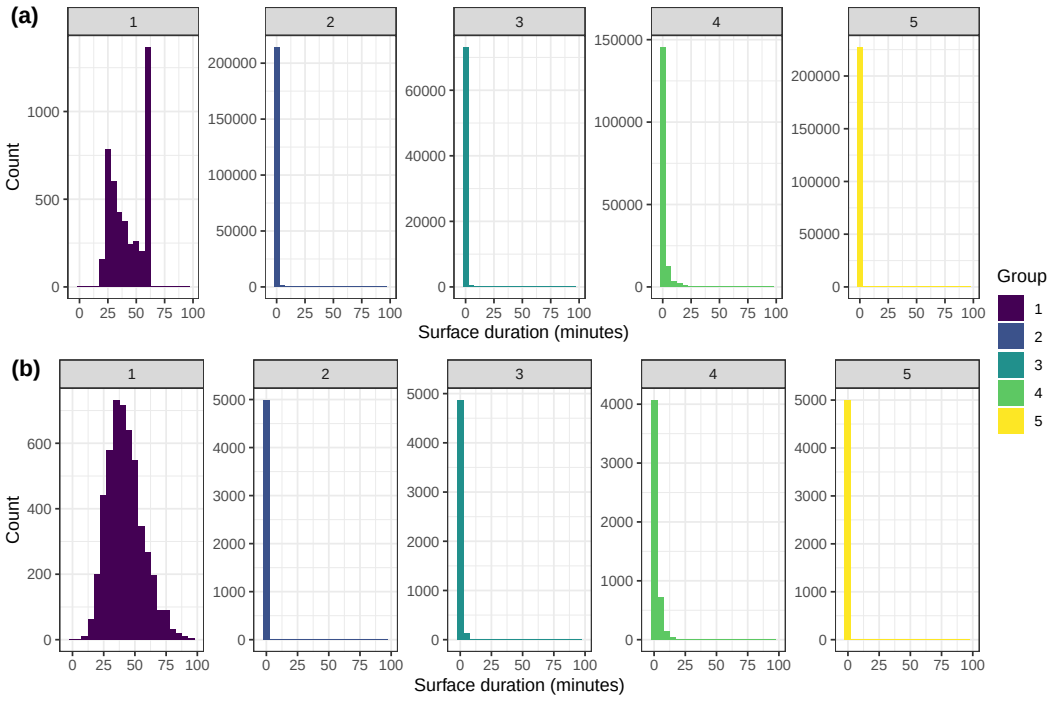


Figure S5 Histograms of post-dive surface duration for each dive type (1–5) for original (a) and simulated (b) dive data. (a) The original data are all recorded dives from 24 tagged harbour seals in the UK. (b) The simulated data consist of 5000 simulated dives of each dive type.

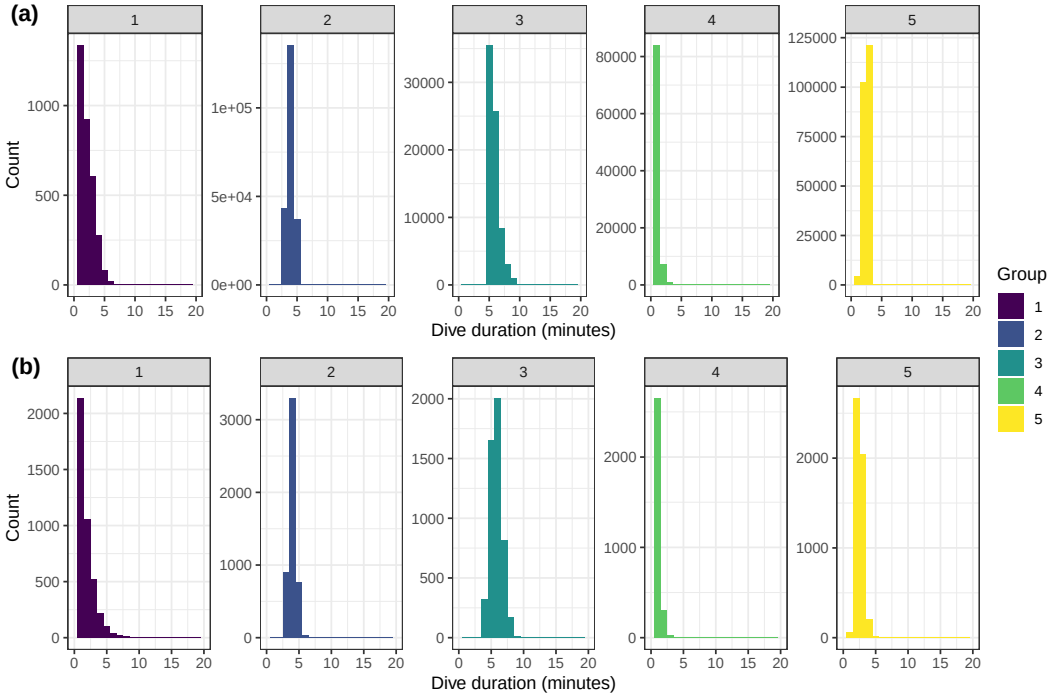


Figure S6 Histograms of total dive duration for each dive type (1–5) for original (a) and simulated (b) dive data. (a) The original data are all recorded dives from 24 tagged harbour seals in the UK. (b) The simulated data consist of 5000 simulated dives of each dive type.

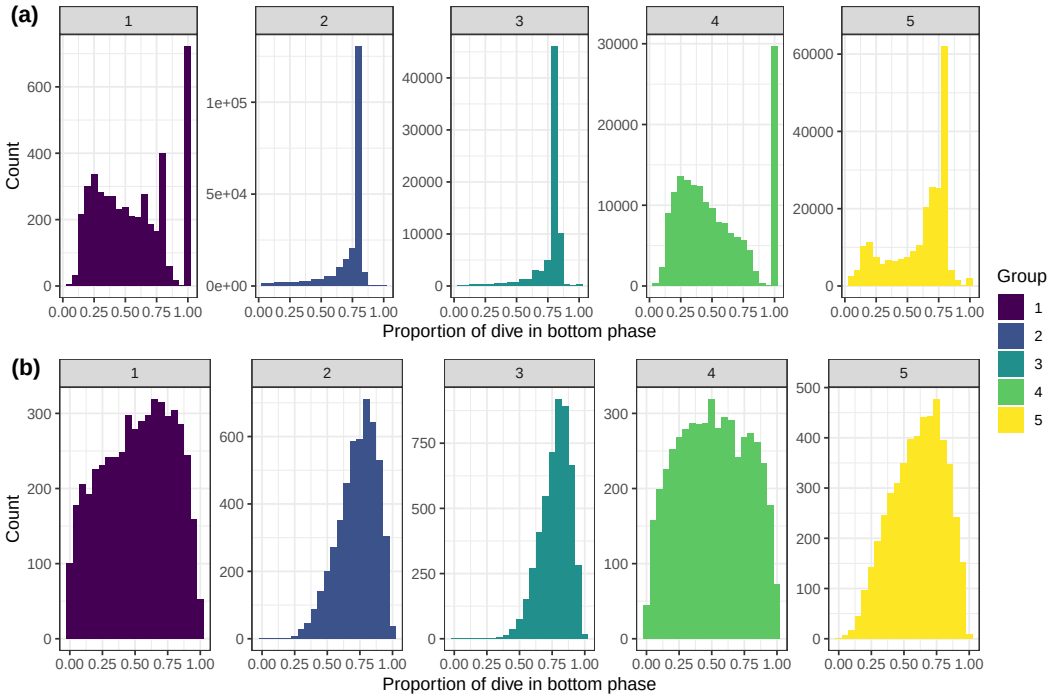


Figure S7 Histograms of proportion of dive in the bottom phase for each dive type (1–5) for original (a) and simulated (b) dive data. (a) The original data are all recorded dives from 24 tagged harbour seals in the UK. (b) The simulated data consist of 5000 simulated dives of each dive type.

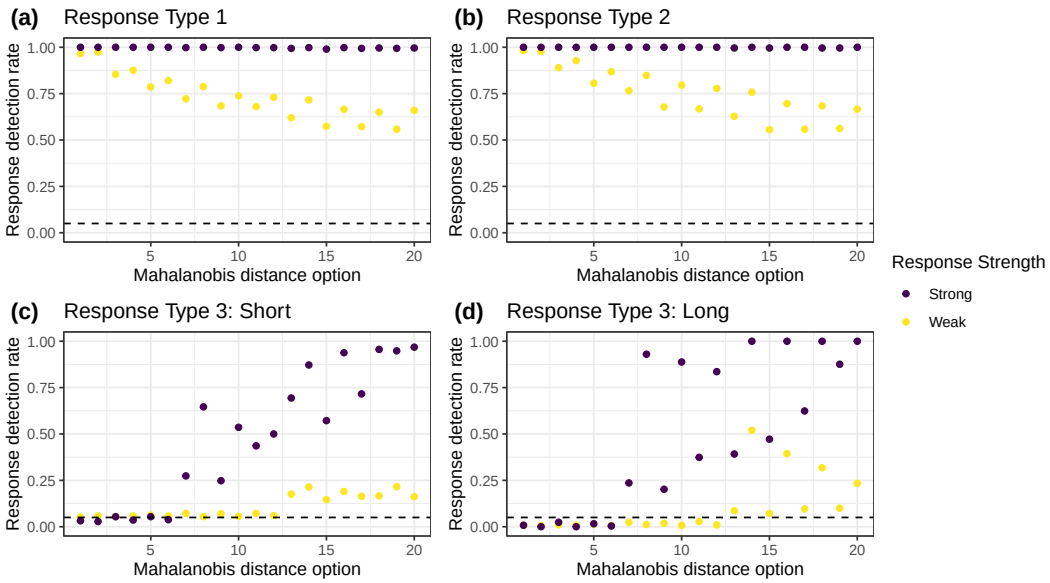


Figure S8 Response detection rates for simulated vertical responses, with the covariance matrix calculated over all data for the simulated animal. Each dot is the response detection rate (out of 500 simulations) for responses of different strengths (coloured dots) and types: (a) Type 1: surface dive; (b) Type 2: surface after a normal dive; (c) and (d) Type 3: travel. All simulations had 20 days of baseline data. Each Mahalanobis distance option is detailed in Table 2.11.

Appendix C: Supplementary material for Chapter 3

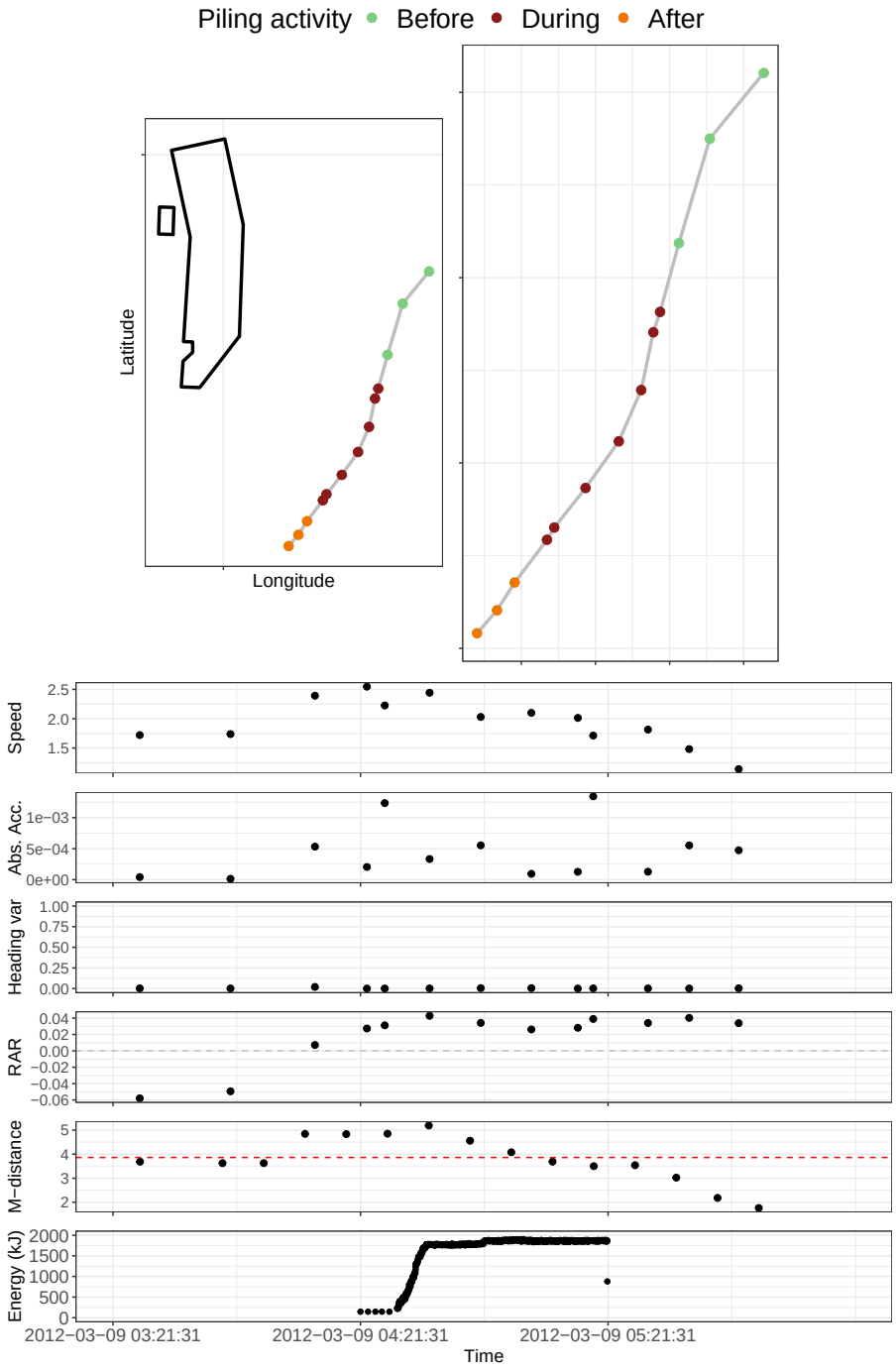


Figure S9 Detected horizontal response 1. Prior to pile-driving, the seal is travelling south-west past the wind farm. As pile driving begins, the seal increases its speed to ~2.5 m s^{-1} (for reference, 99% of speeds measured for this individual were <1.65 m s^{-1}), and continues to travel at high speed away from the wind farm. All panels show 1 hour prior to, and 1 hour post-pile driving.

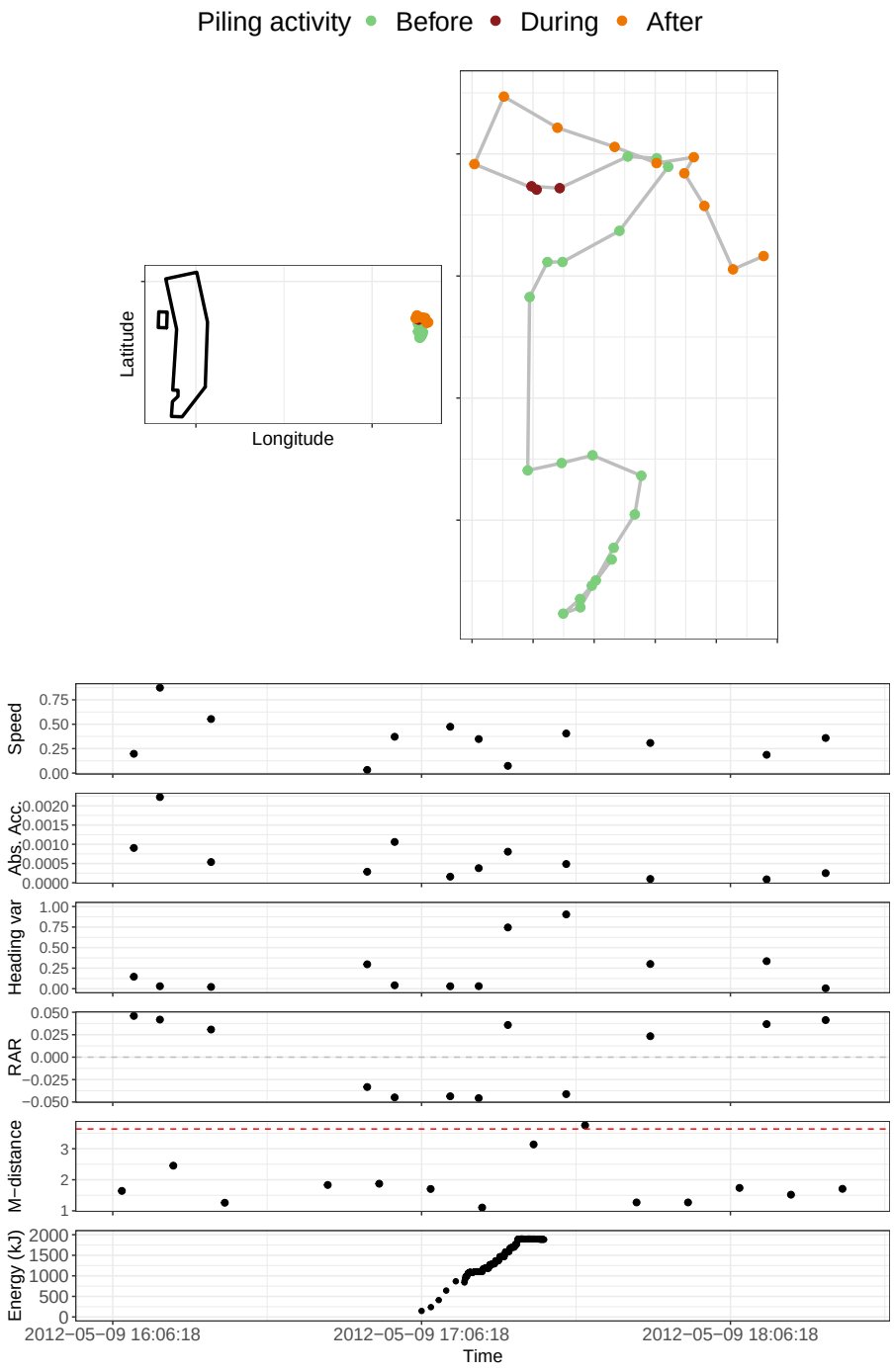


Figure S10 Detected horizontal response 2. Prior to pile driving, the seal is meandering in a northwards direction. During pile driving, the seal slows down and has a high variability in heading. After pile driving ends, the seal turns around and begins travelling east away from the wind farm. The top map panels show 3 hours before and after piling; the bottom movement metric panels show 1 hour before and after piling.

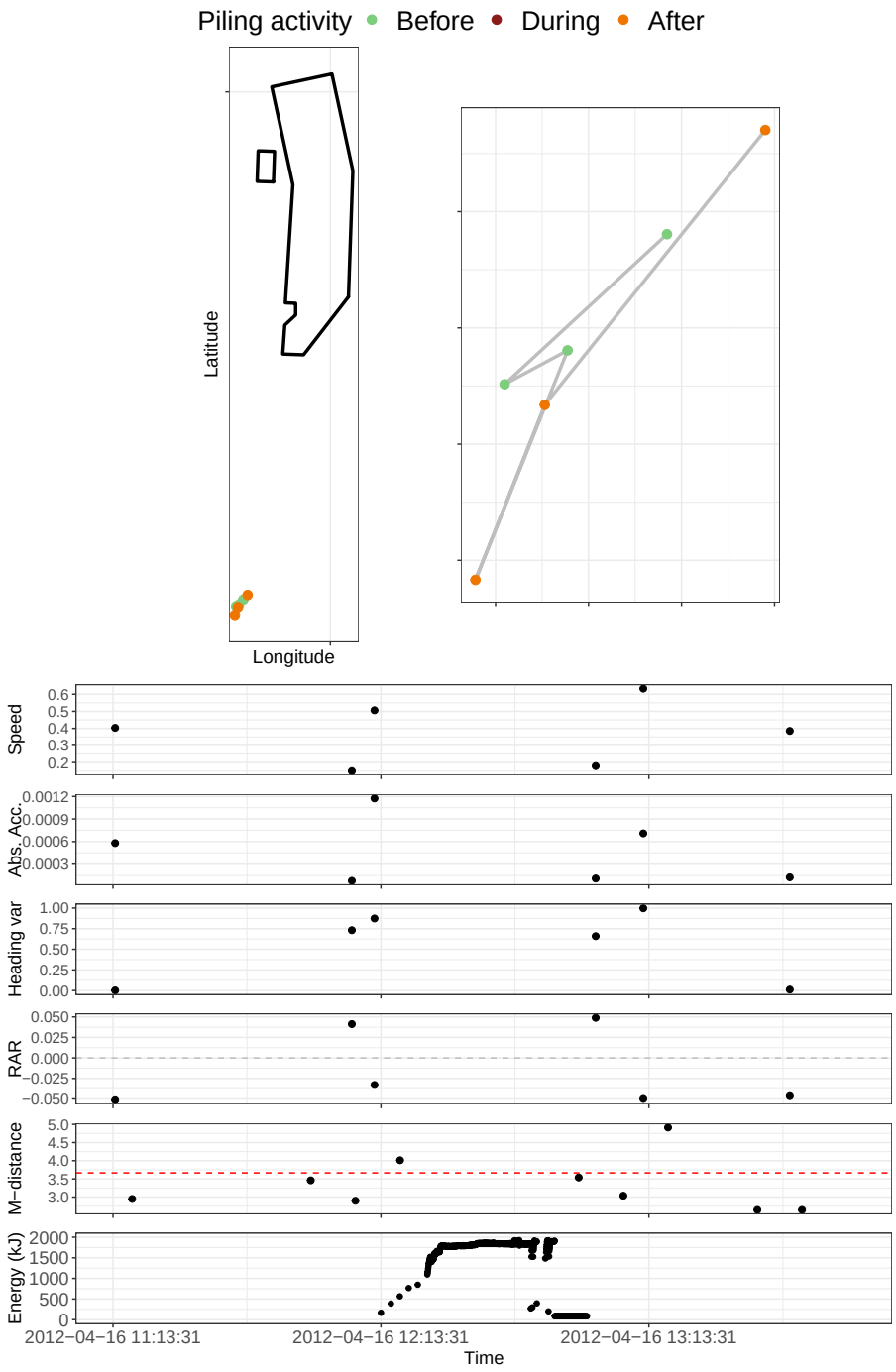


Figure S11 Detected horizontal response 3. Prior to pile driving, the seal is milling south of the wind farm. Near the start and end of the pile driving bout, there are sharp turns in heading. All panels show 1 hour prior to, and 1 hour post-pile driving.

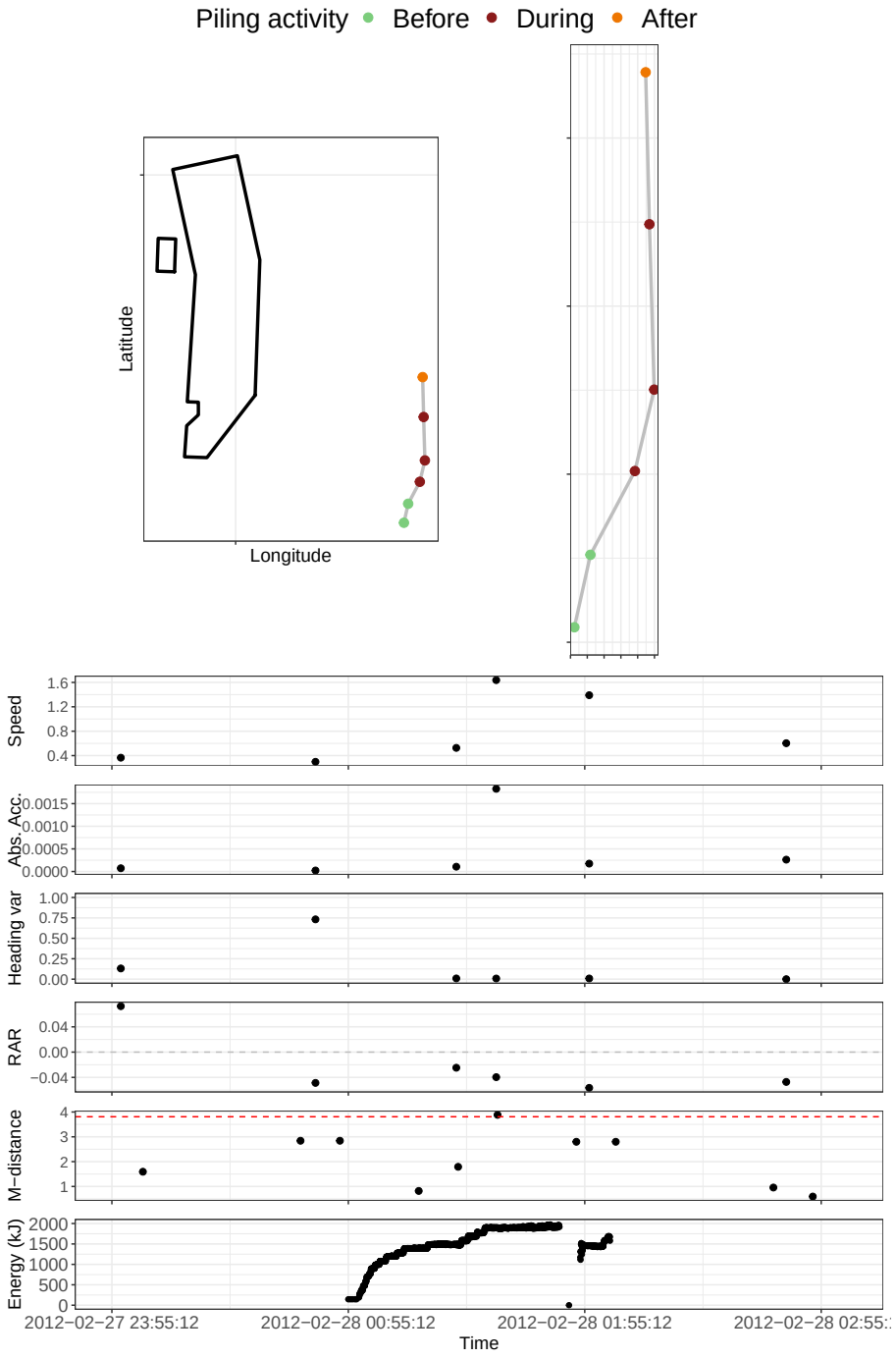


Figure S12 Detected horizontal response 4. Prior to pile driving, the seal is travelling northwards past the wind farm. During pile driving, there is a very high acceleration, from ∼0.6 m s^{-1} to ∼1.6 m s^{-1}, and an adjustment in heading. All panels show 1 hour prior to, and 1 hour post-pile driving.

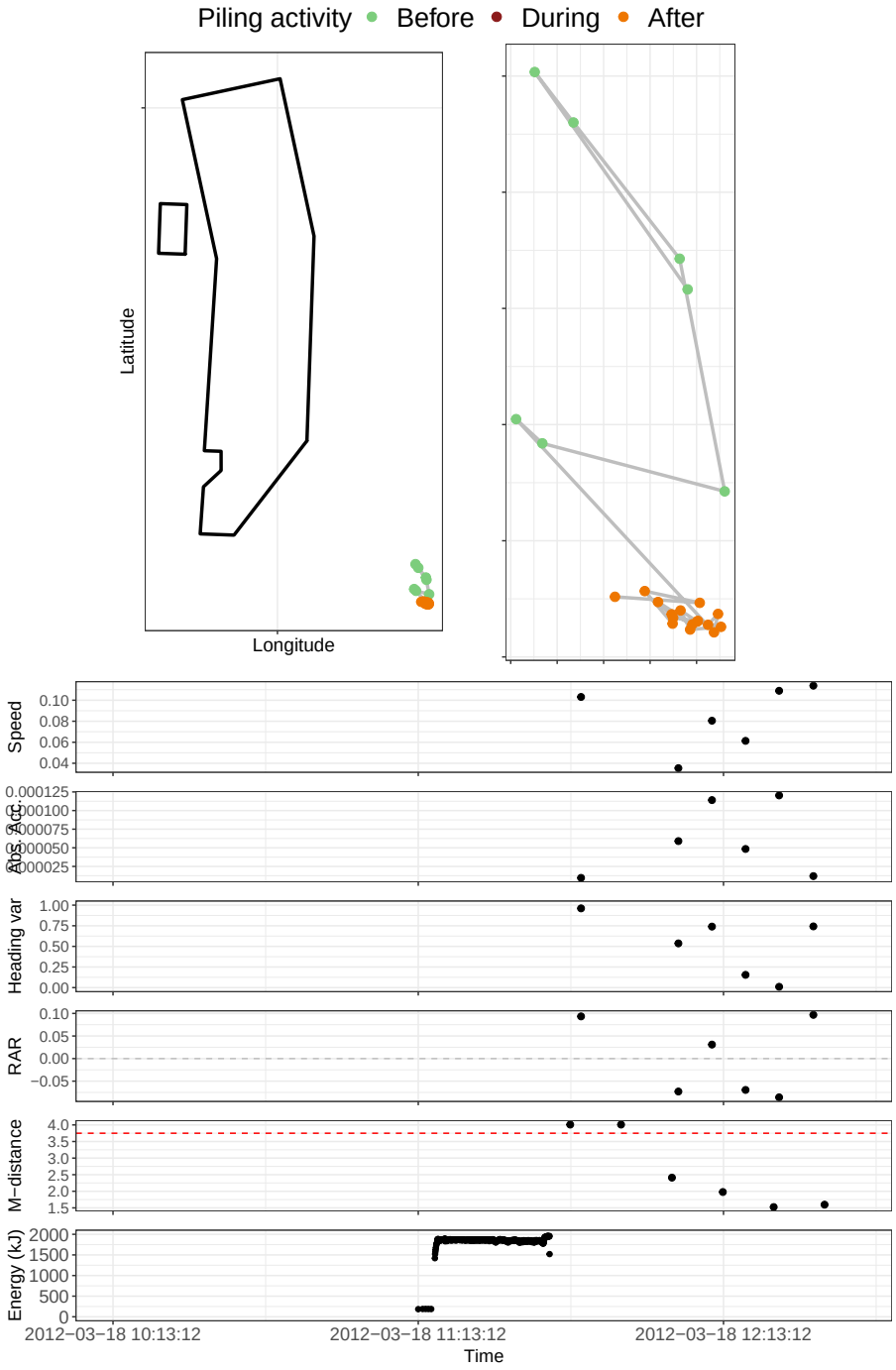


Figure S13 Detected horizontal response 5. Prior to pile driving the seal is meandering near the south-east end of the wind farm. There are no recorded GPS locations during the piling bout, but immediately after piling ends the seal is recorded making very sharp turns (high heading variability). The top map panels show 3 hours before and after piling; the bottom movement metric panels show 1 hour before and after piling.

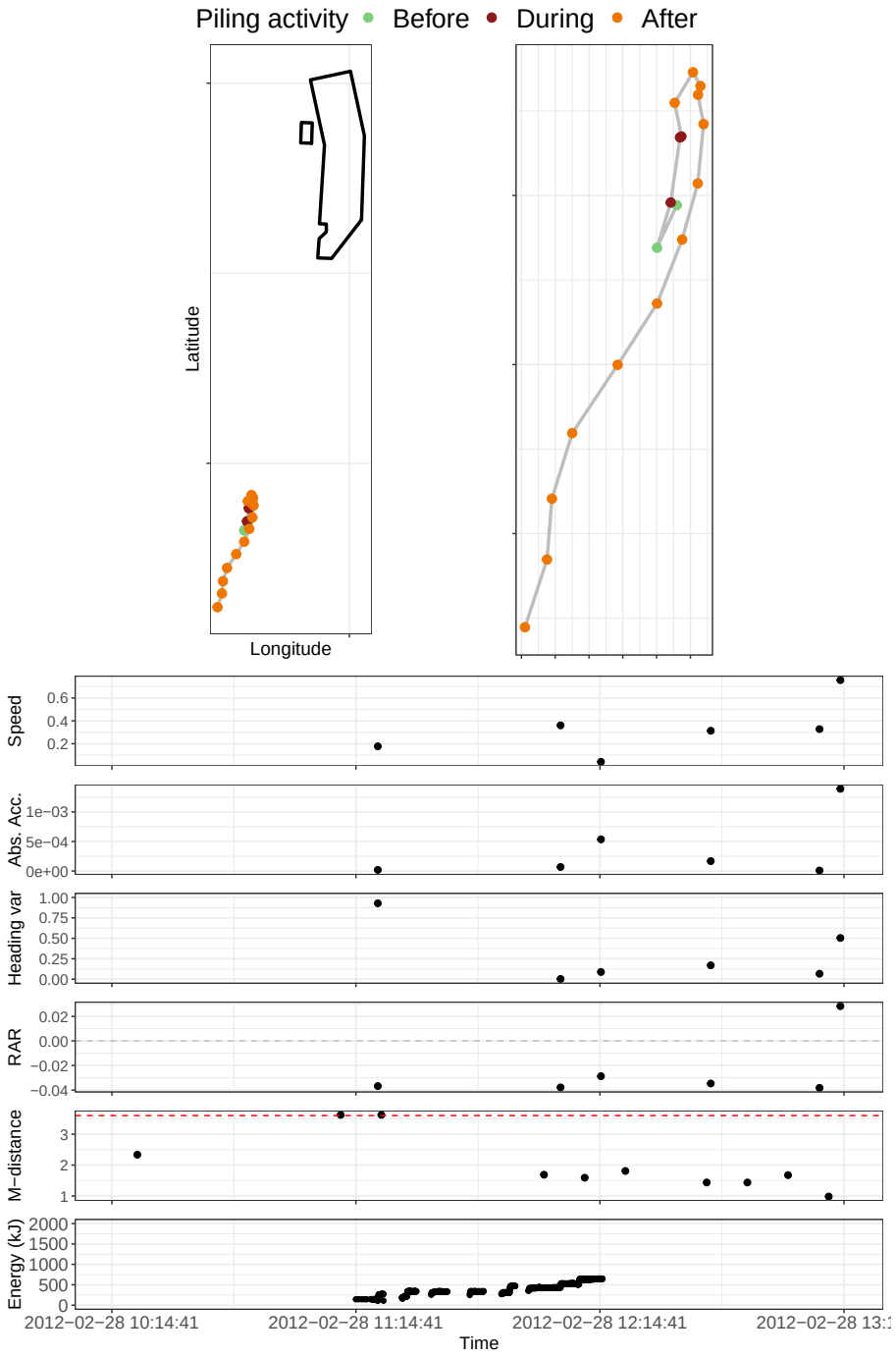


Figure S14 Detected horizontal response 6. Prior to pile driving, the seal is milling around south of the wind farm. During and after pile driving, the seal speeds up, turns around, and travels directly away from the wind farm. The top map panels show 3 hours before and after piling; the bottom movement metric panels show 1 hour before and after piling.

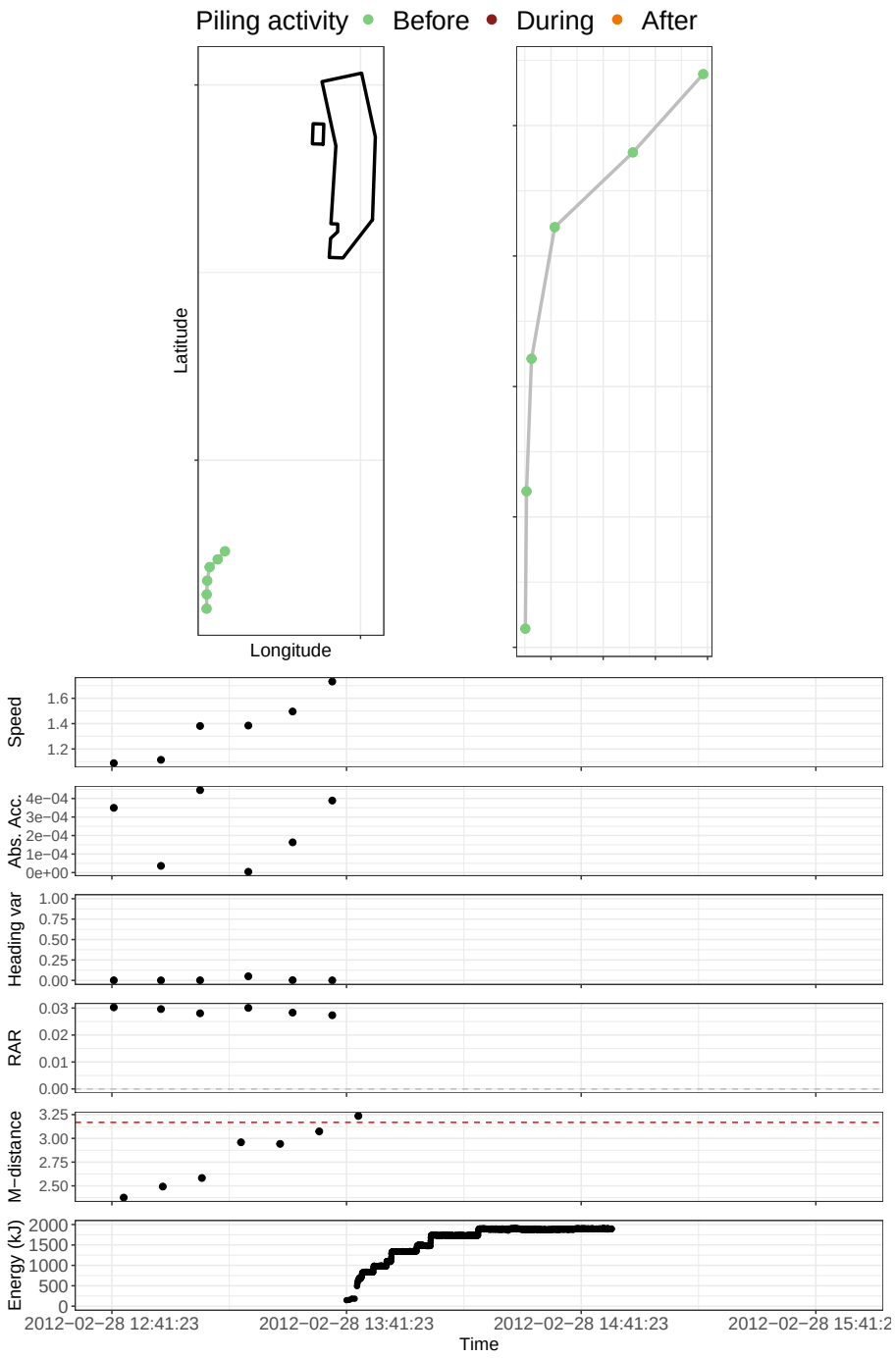


Figure S15 Detected horizontal response 7. Prior to pile driving, the seal is travelling south-west, away from the wind farm. As pile driving begins, the seal increases its speed to ∼1.8 m s^{-1} as it approaches the haulout site. All panels show 1 hour prior to, and 1 hour post-pile driving.

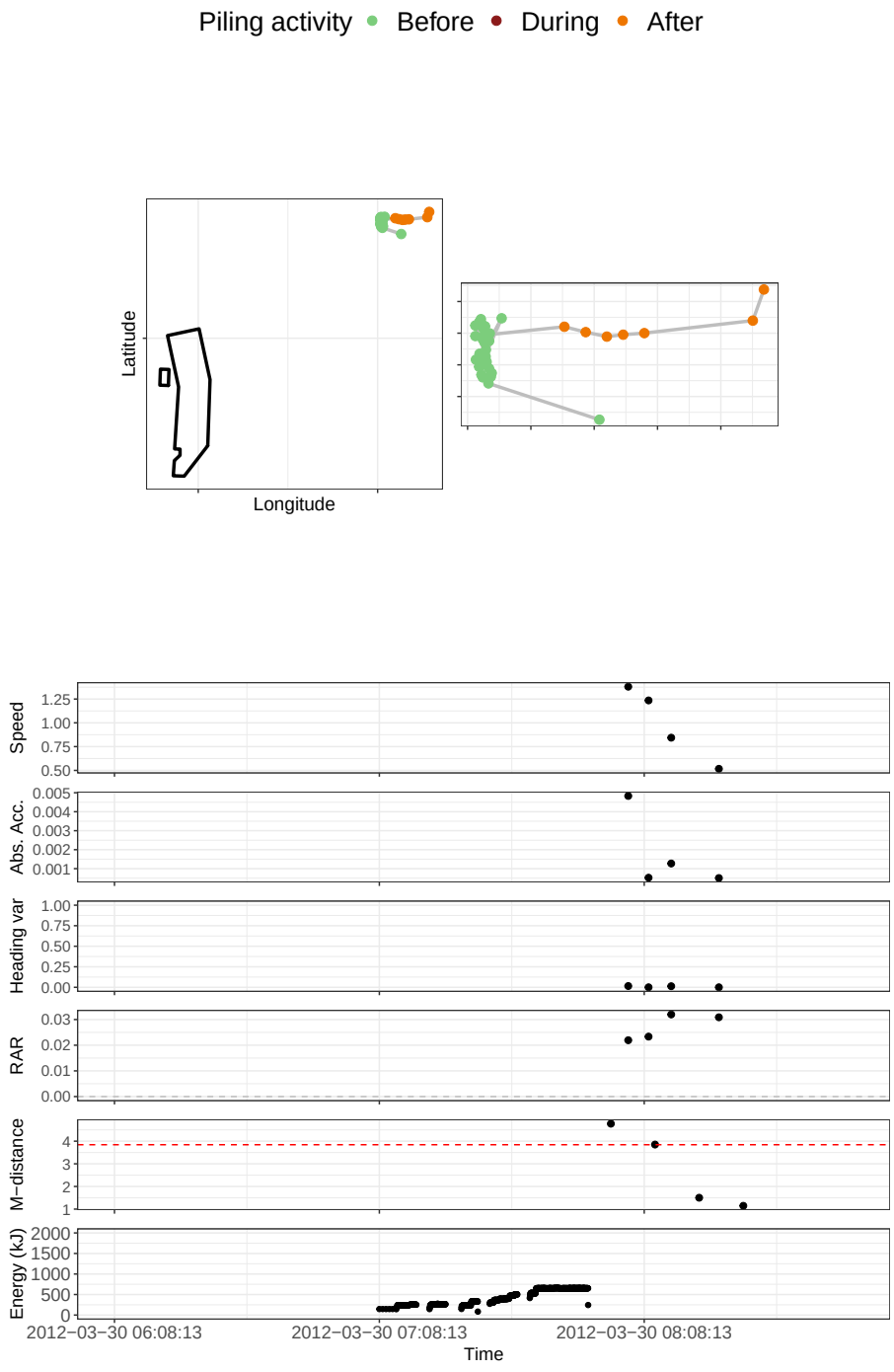


Figure S16 Detected horizontal response 8. For ∼12 hours the seal had been milling or floating (almost stationary) in approximately the same location. Immediately after pile driving ceases, the seal accelerates very quickly and moves at ∼1.3 m s^{-1} away from the wind farm. The top map panels show 12 hours before and 3 hours after piling; the bottom movement metric panels show 1 hour before and after piling.

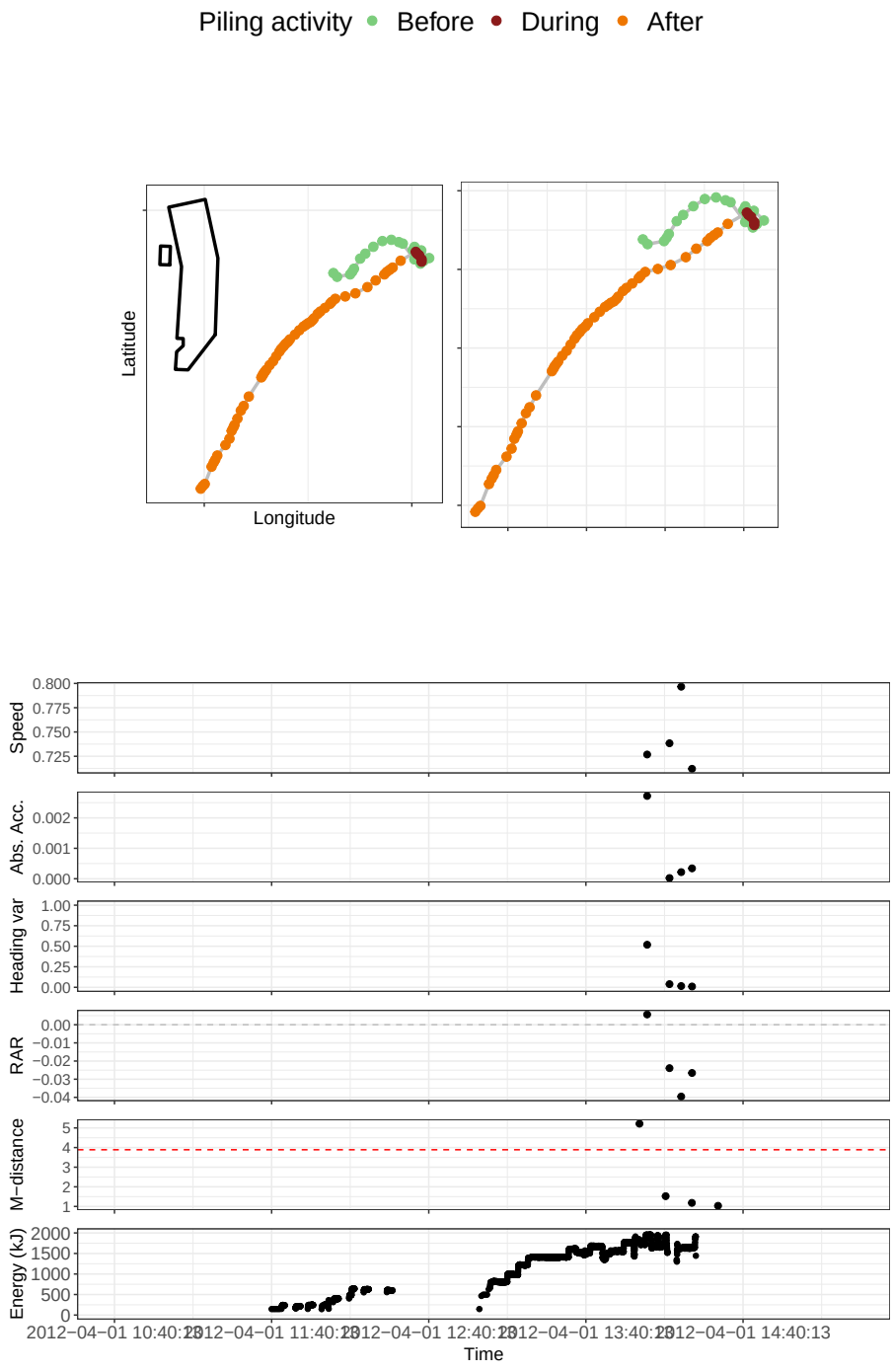


Figure S17 Detected horizontal response 9. For ∼12 hours the seal had been milling or floating (almost stationary) in approximately the same location. During high-energy pile driving, the seal accelerates, with moderate heading variability, and begins to travel back towards the haulout site. The top map panels show 12 hours before and 12 hours after piling; the bottom movement metric panels show 1 hour before and after piling.

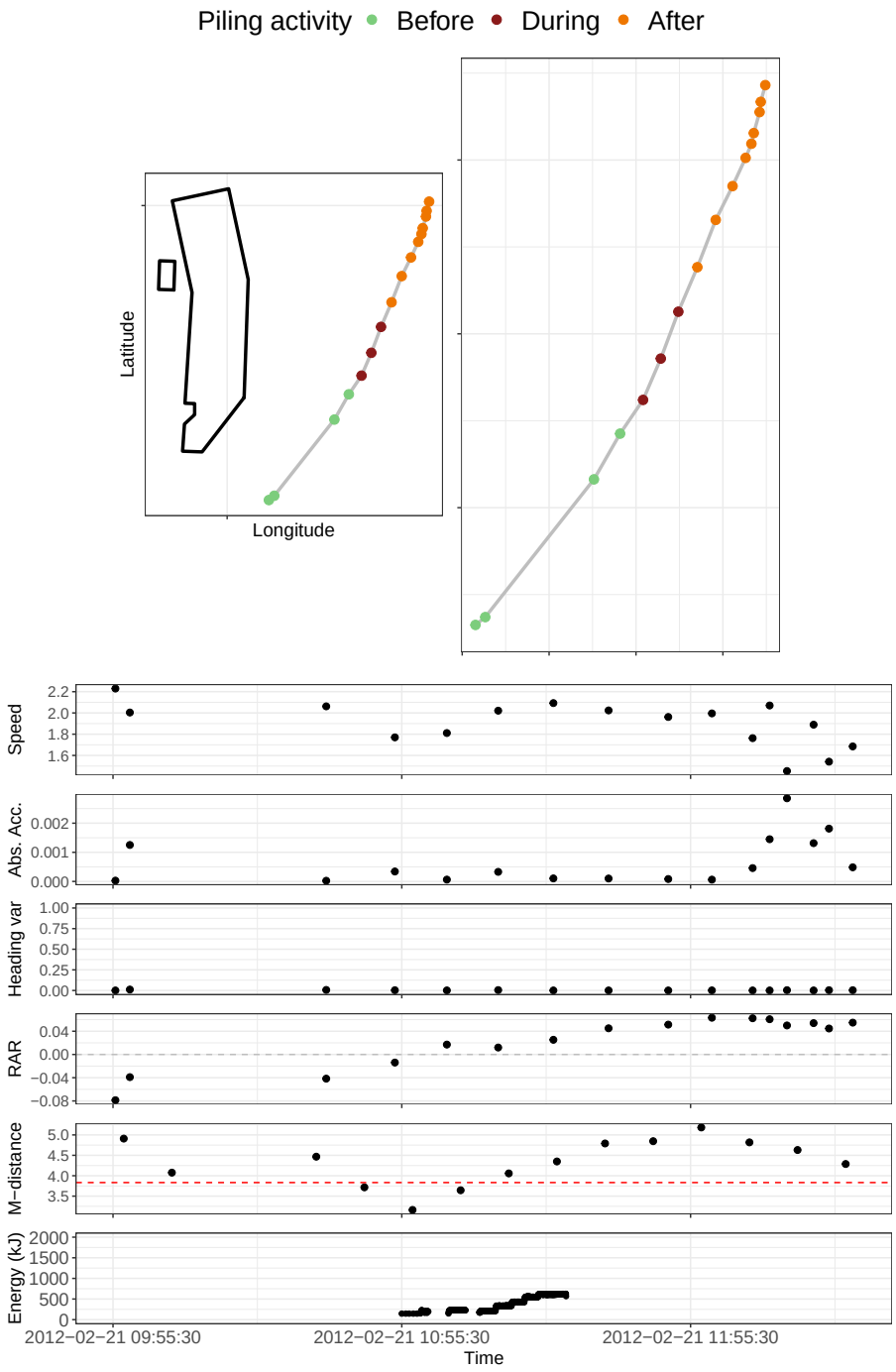


Figure S18 Detected horizontal response 10. During pile driving, the seal is travelling unusually fast directly past the wind farm in a north-east direction. All panels show 1 hour prior to, and 1 hour post-pile driving.

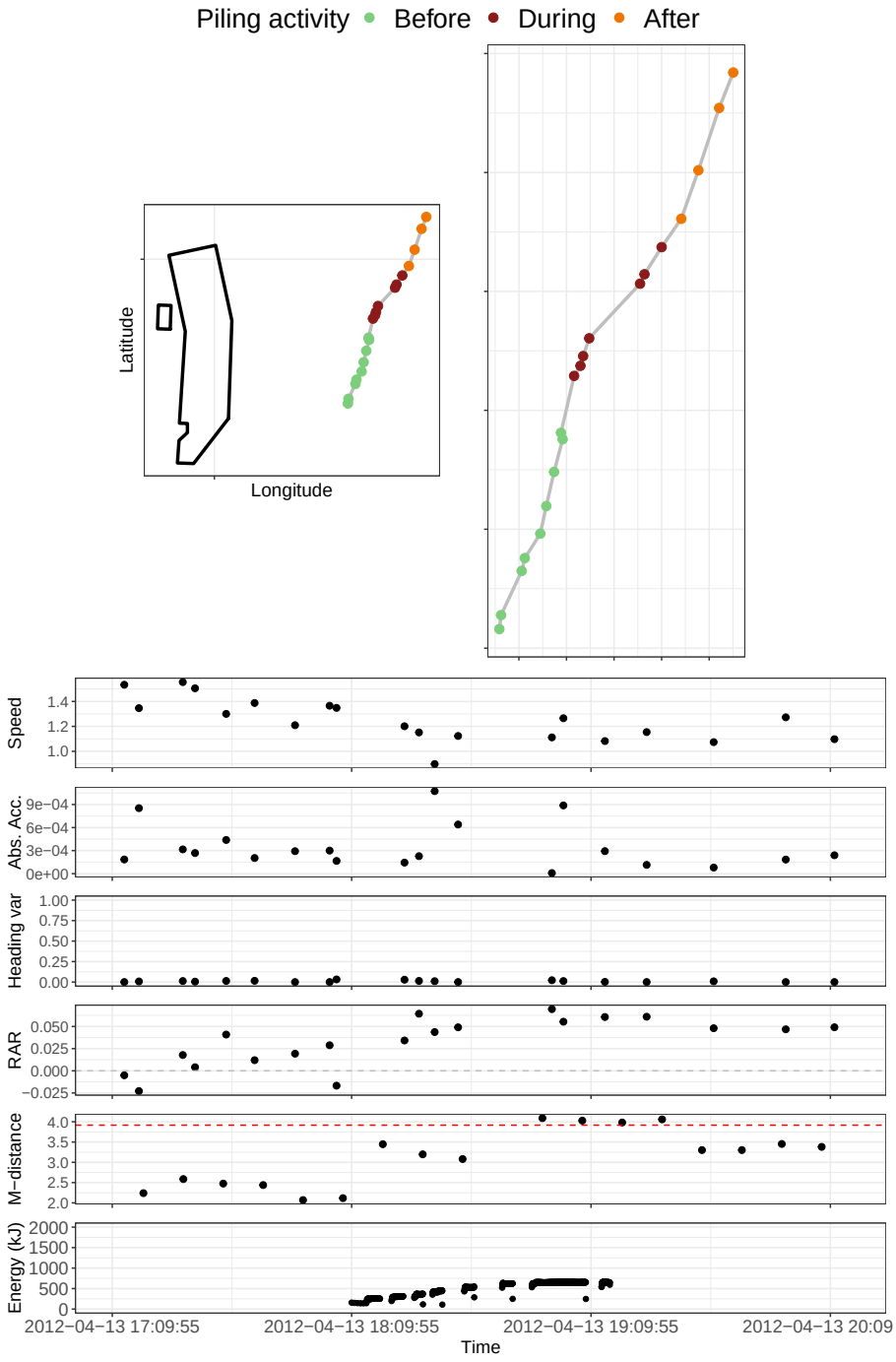


Figure S19 Detected horizontal response 11. Prior to pile driving the seal is travelling north-east past the wind farm. Halfway though the piling bout (approximately when the piling becomes continuous, rather than intermittent ramp-up), the seal has a short burst increase in speed and deviates from its original heading. All panels show 1 hour prior to, and 1 hour post-pile driving.

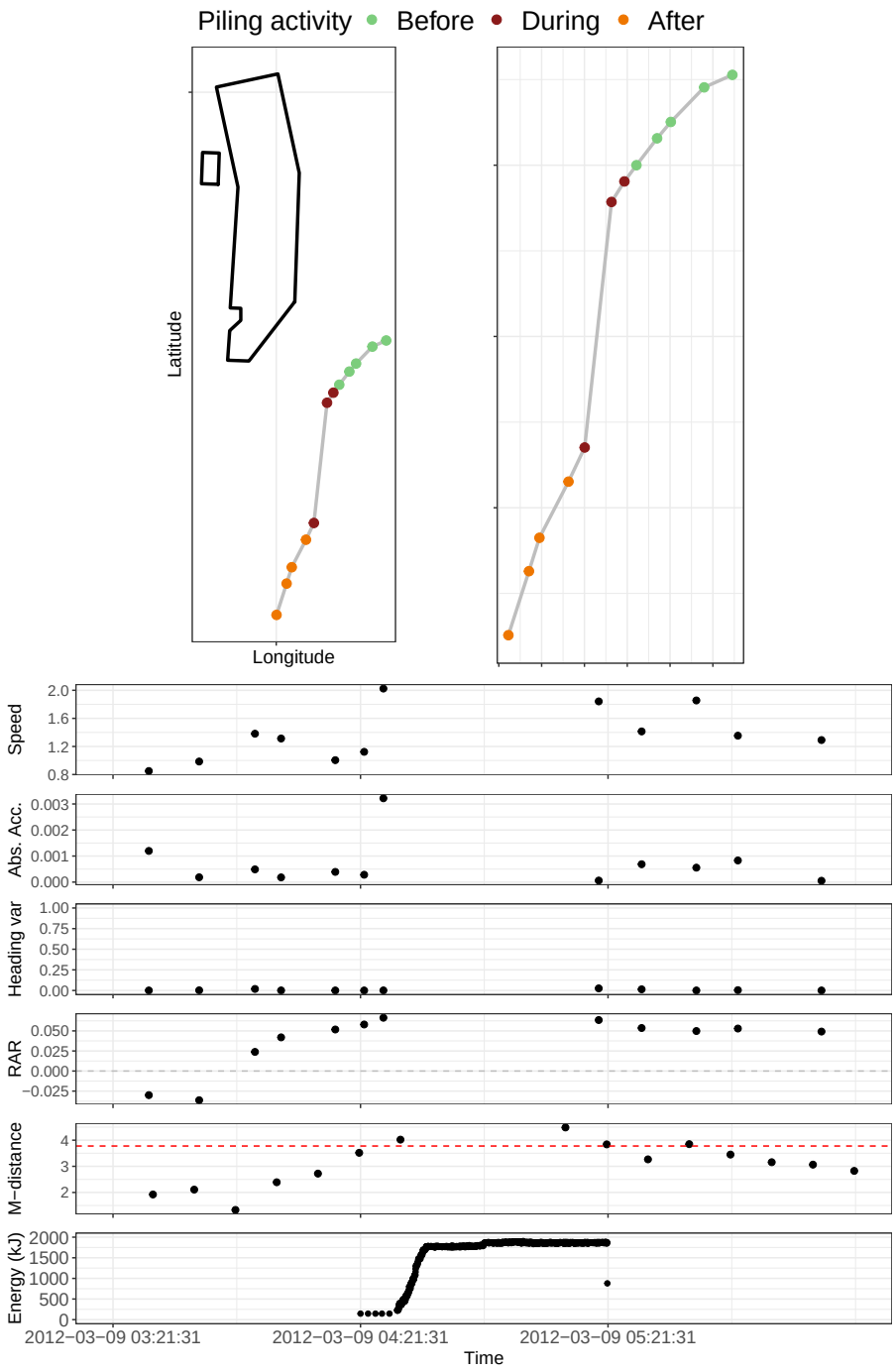


Figure S20 Detected horizontal response 12. Prior to pile driving the seal is moving at ∼0.8–1.4 m s^{-1} past the wind farm. During pile driving, the seal rapidly accelerates to ∼2.0 m s^{-1} and moves away from the wind farm. All panels show 1 hour prior to, and 1 hour post-pile driving.

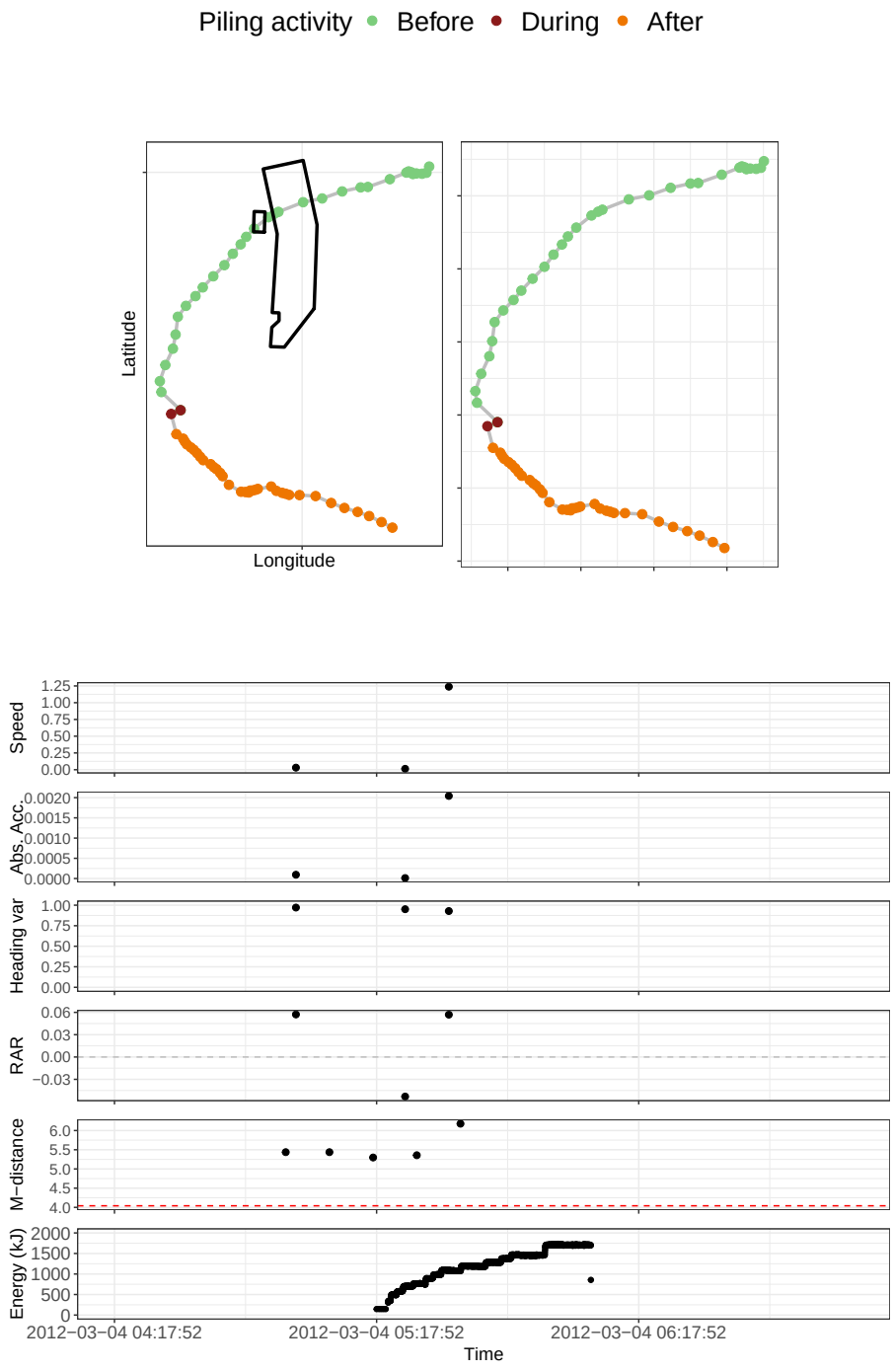


Figure S21 Detected horizontal response 13. Prior to pile driving the seal is travelling through and then past the wind farm. During pile driving, the seal stops moving, with a high heading variability recorded. After pile driving, the seal continues travelling. The top map panels show 12 hours before and 12 hours after piling; the bottom movement metric panels show 1 hour before and after piling.

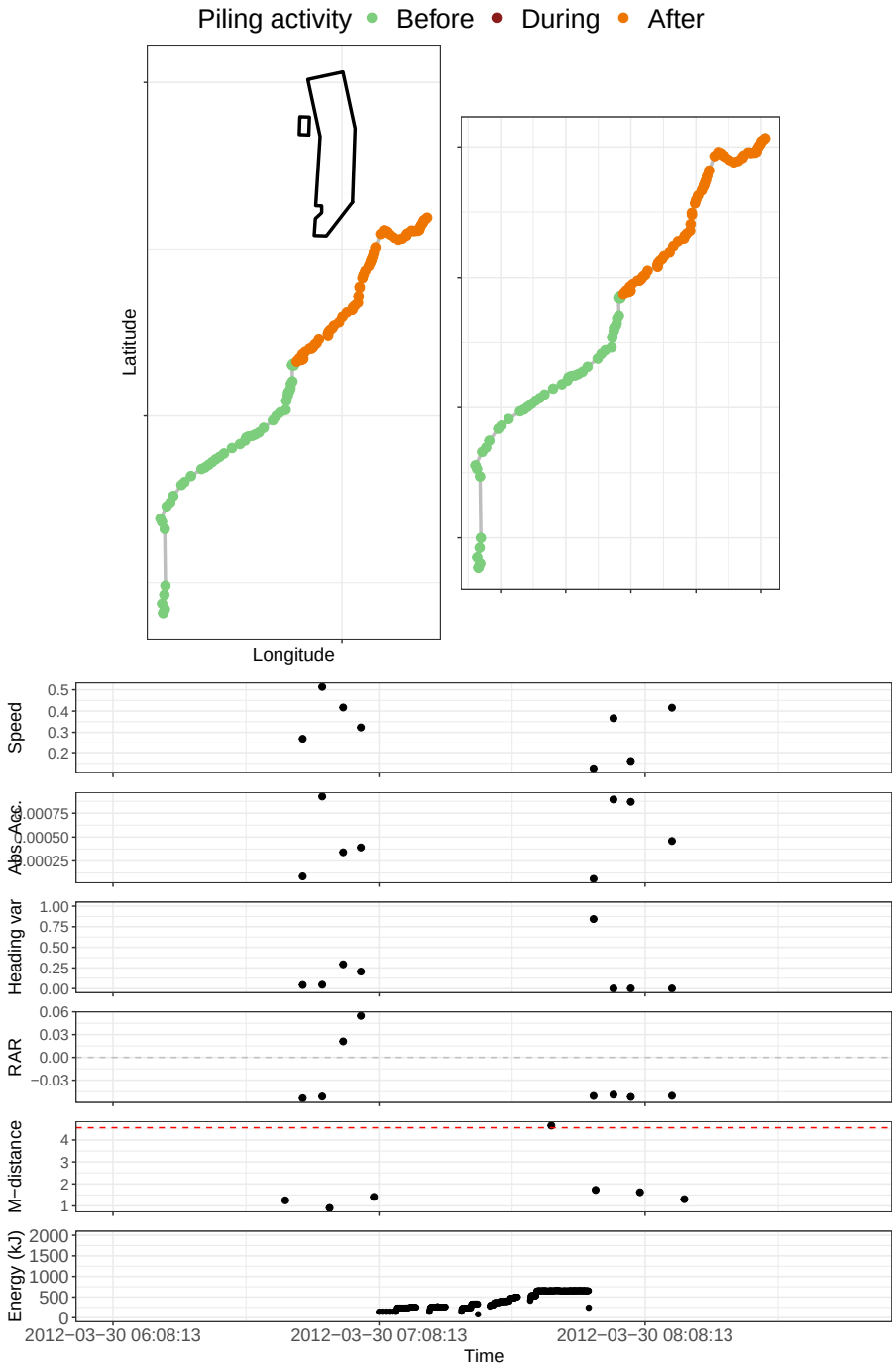


Figure S22 Detected horizontal response 14. Prior to pile driving, the seal is travelling north towards the wind farm. At the end of the pile driving bout, the seal makes a deviation in heading. The top map panels show 12 hours before and 12 hours after piling; the bottom movement metric panels show 1 hour before and after piling.

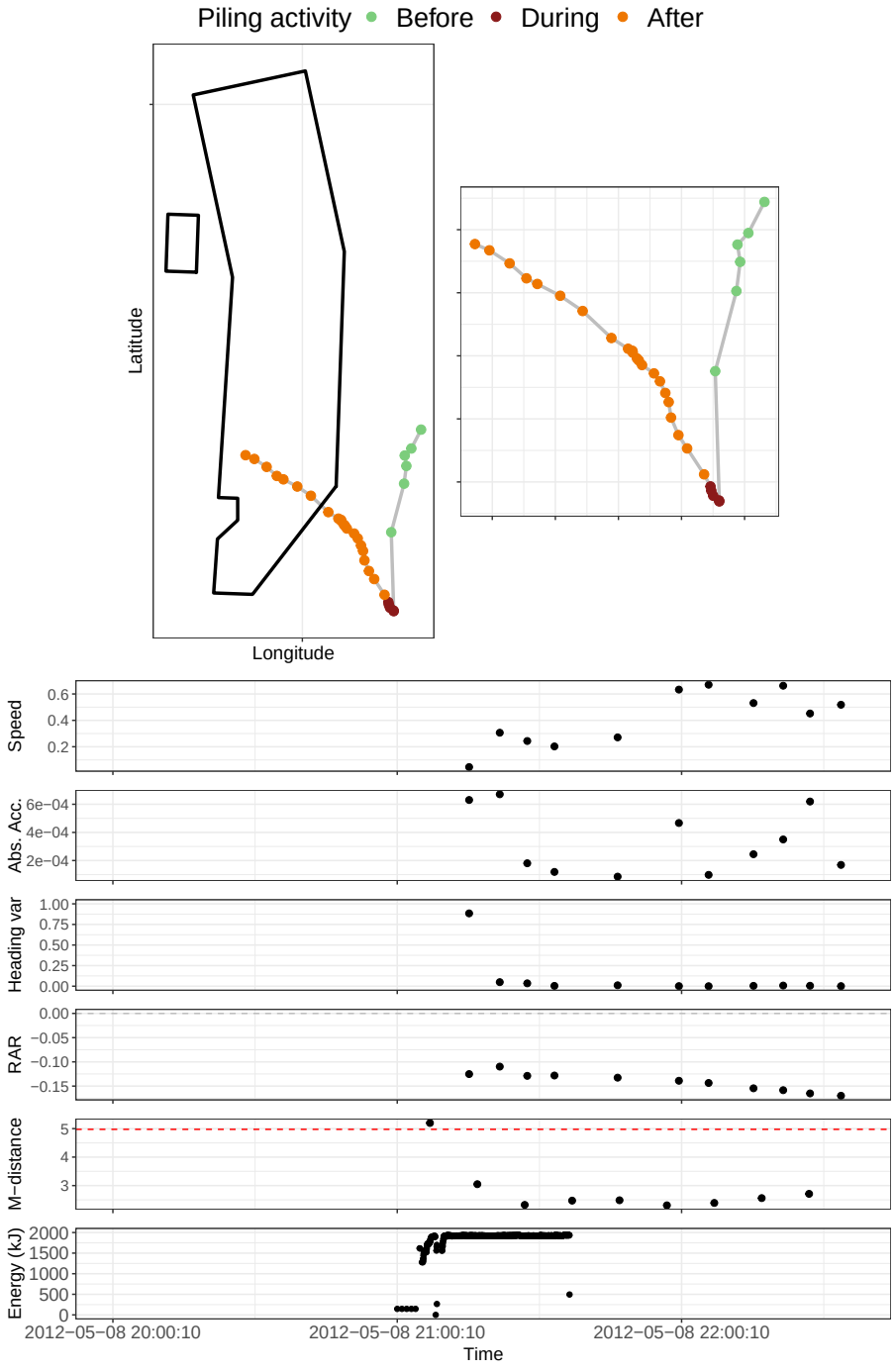


Figure S23 Detected horizontal response 15. Prior to pile driving the seal is heading south past the wind farm. During piling, the seal slows down suddenly and has an increase in heading variability. As pile driving ends, the seal heads north-west towards and into the wind farm. The top map panels show 3 hours before and 3 hours after piling; the bottom movement metric panels show 1 hour before and after piling.

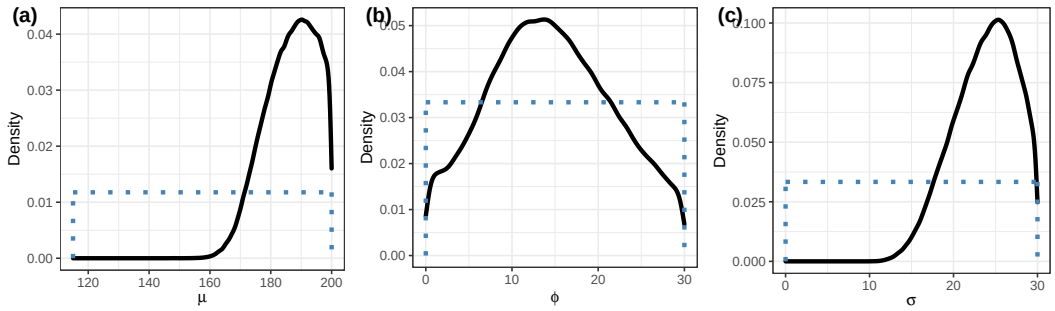


Figure S24 Posterior (black, solid line) and prior (blue, dotted line) densities for the parameters in the final dose-response model with no covariates.

Table S1 Estimated response probabilities against single-strike sound exposure level (SELss) of pile driving. Shown are the mean and 95% credible intervals in 5 dB increments, estimated from posterior samples of the dose-response model.

	Probability of response		
SELss (dB re $1\mu\text{Pa}^2 \cdot \text{s}$)	**Mean**	**2.5% quantile**	**97.5% quantile**
115	0.000	0.000	0.000
120	0.006	0.001	0.018
125	0.015	0.003	0.041
130	0.028	0.006	0.070
135	0.046	0.012	0.105
140	0.070	0.023	0.147
145	0.101	0.039	0.197
150	0.142	0.064	0.256
155	0.193	0.099	0.323
160	0.255	0.146	0.398
165	0.328	0.208	0.481
170	0.411	0.284	0.570
175	0.503	0.376	0.660
180	0.602	0.483	0.748
185	0.704	0.602	0.830
190	0.806	0.730	0.900
195	0.905	0.864	0.957
200	1.000	1.000	1.000

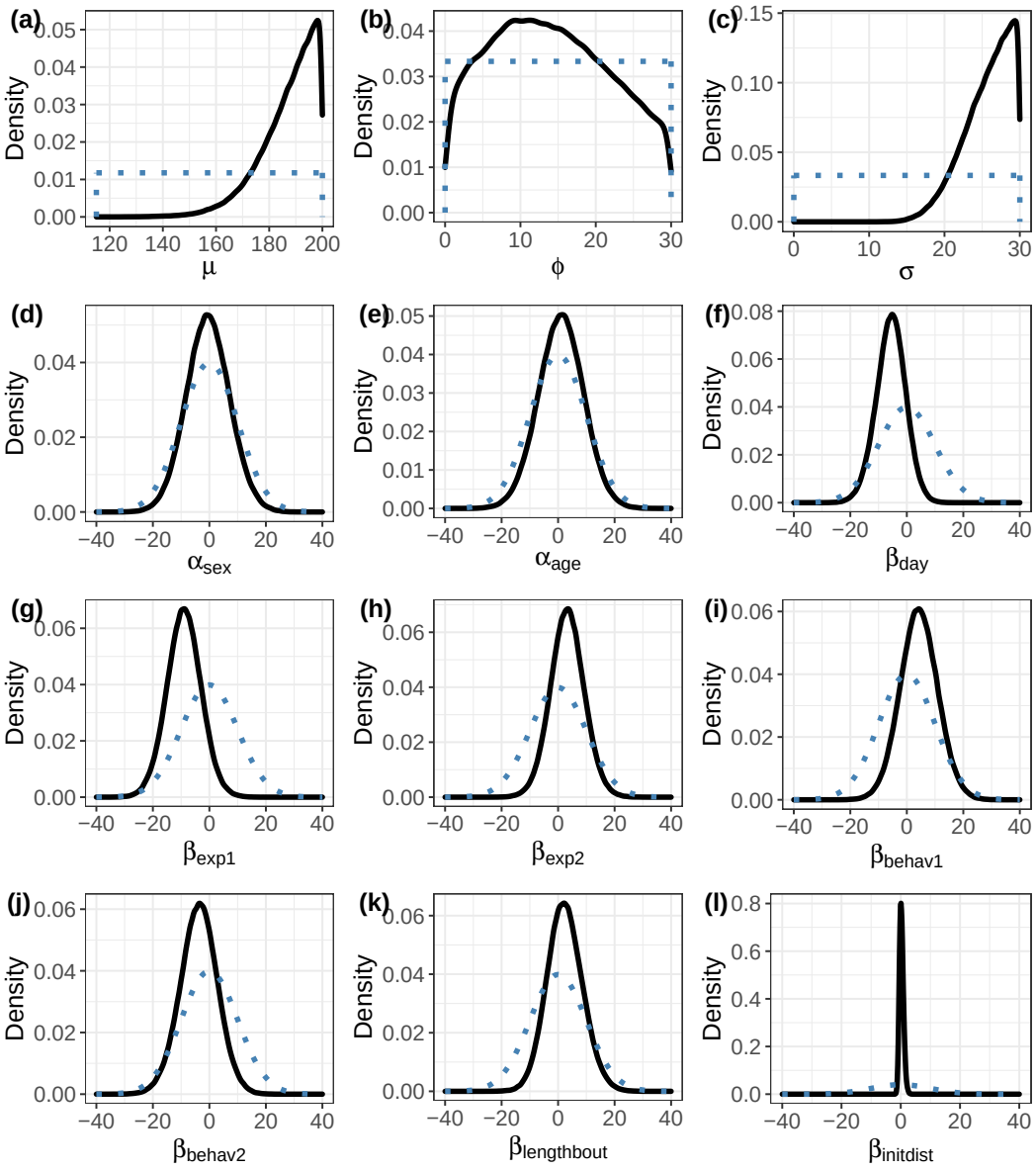


Figure S25 Posterior (black, solid line) and prior (blue, dotted line) densities for the parameters in the maximal model with all candidate covariates.

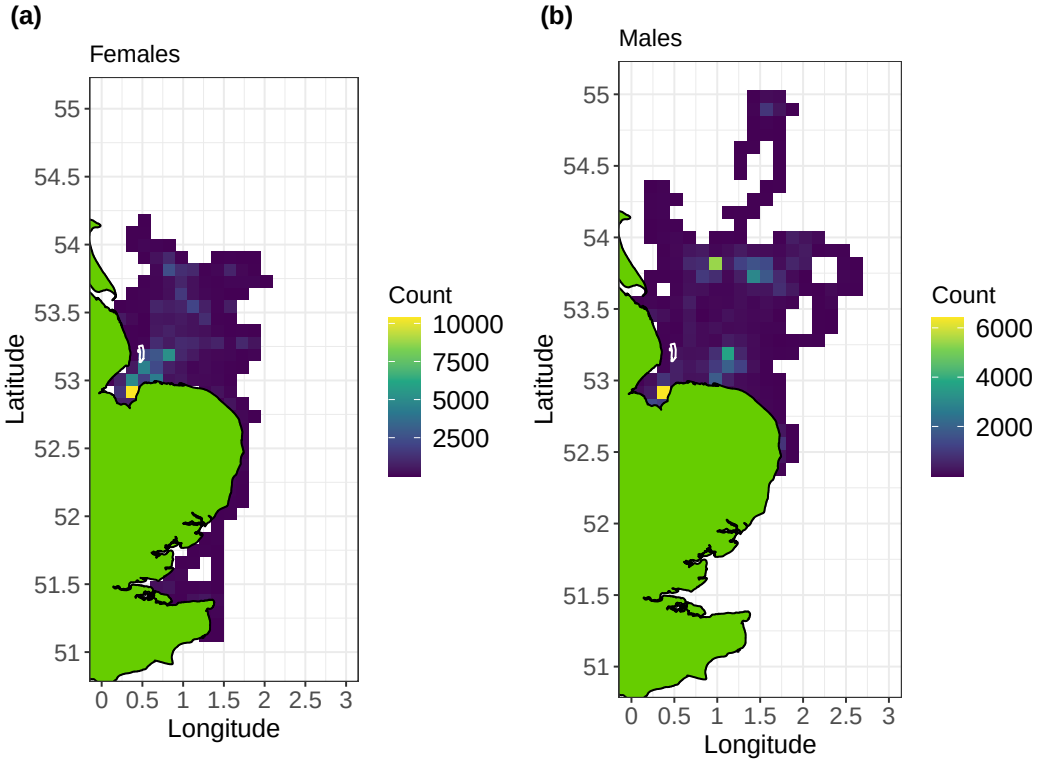


Figure S26 GPS locations by sex of 24 tagged harbour seals using The Wash in 2012 (January–May). Coloured grids denote the number of received GPS locations in each 10 x 10 km grid cell for females (a) and males (b).

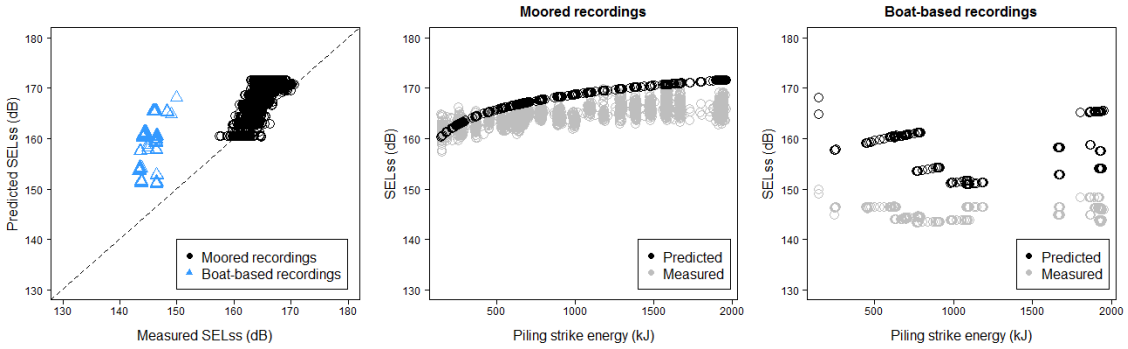


Figure S27 The modelled single-strike sound exposure levels (SELss, dB re $1\mu\text{Pa}^2 \cdot \text{s}$) were compared to measurements from recordings made in the study area of 2,902 piling strikes. Predicted SELss are shown against measured SELss. Recordings were made from a moored sound recorder 9 m below the water surface at a distance of 4,900 m from piling, and boat-based recordings at 1 m below the water surface and distances of 1,000–9,500 m from piling. Figure from supplementary material of Whyte et al. (2020).

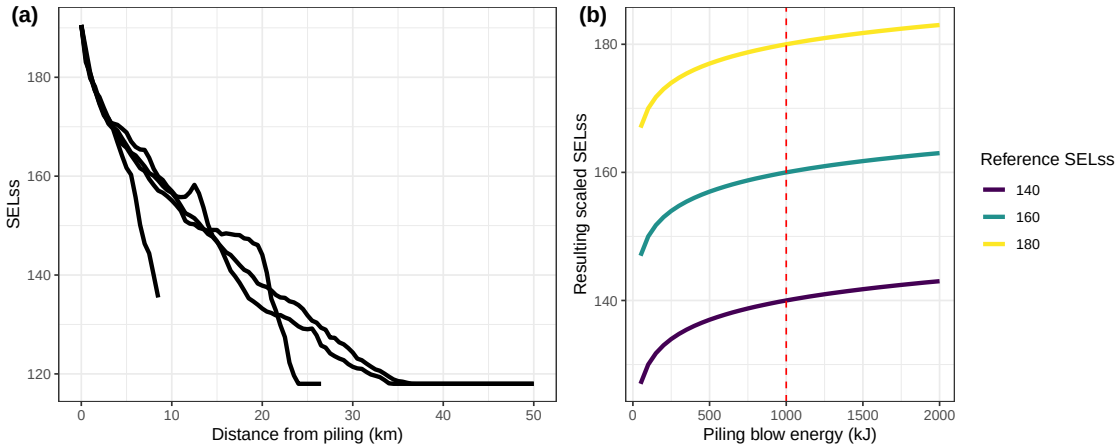


Figure S28 (a) Example of the relationship between distance and predicted single-strike sound exposure level (SELss, dB re $1\mu\text{Pa}^2 \cdot \text{s}$, median across depth). Shown is an example of the predicted SELss for four trajectories (north, south, east, west) spanning out from one of the piling locations (LS22). (b) Example of the relationship between piling blow energy, reference SELss, and the estimated scaled SELss. All piling strikes are modelled at a reference piling energy of 1000 kJ and then scaled by the blow energy to estimate the true SELss.

Appendix D: Supplementary material for Chapter 4

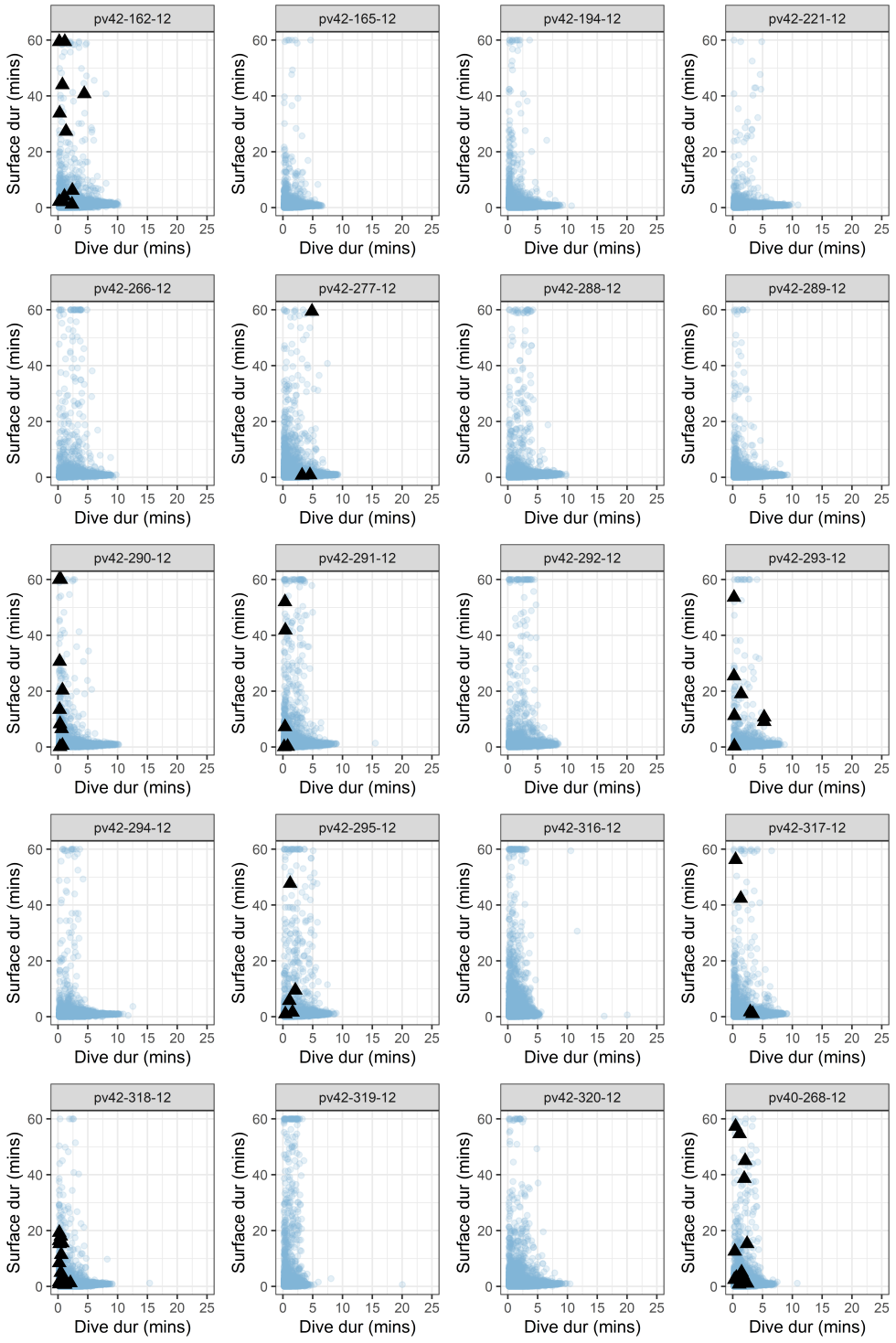


Figure S29 Scatterplot of dive duration and surface duration for each dive in baseline data (blue dots) and identified behavioural responses (black triangles). Analyses were conducted using groups of three sequential dives. Shown here are the 20 seals (panels) which had encounters with pile driving during the study period.

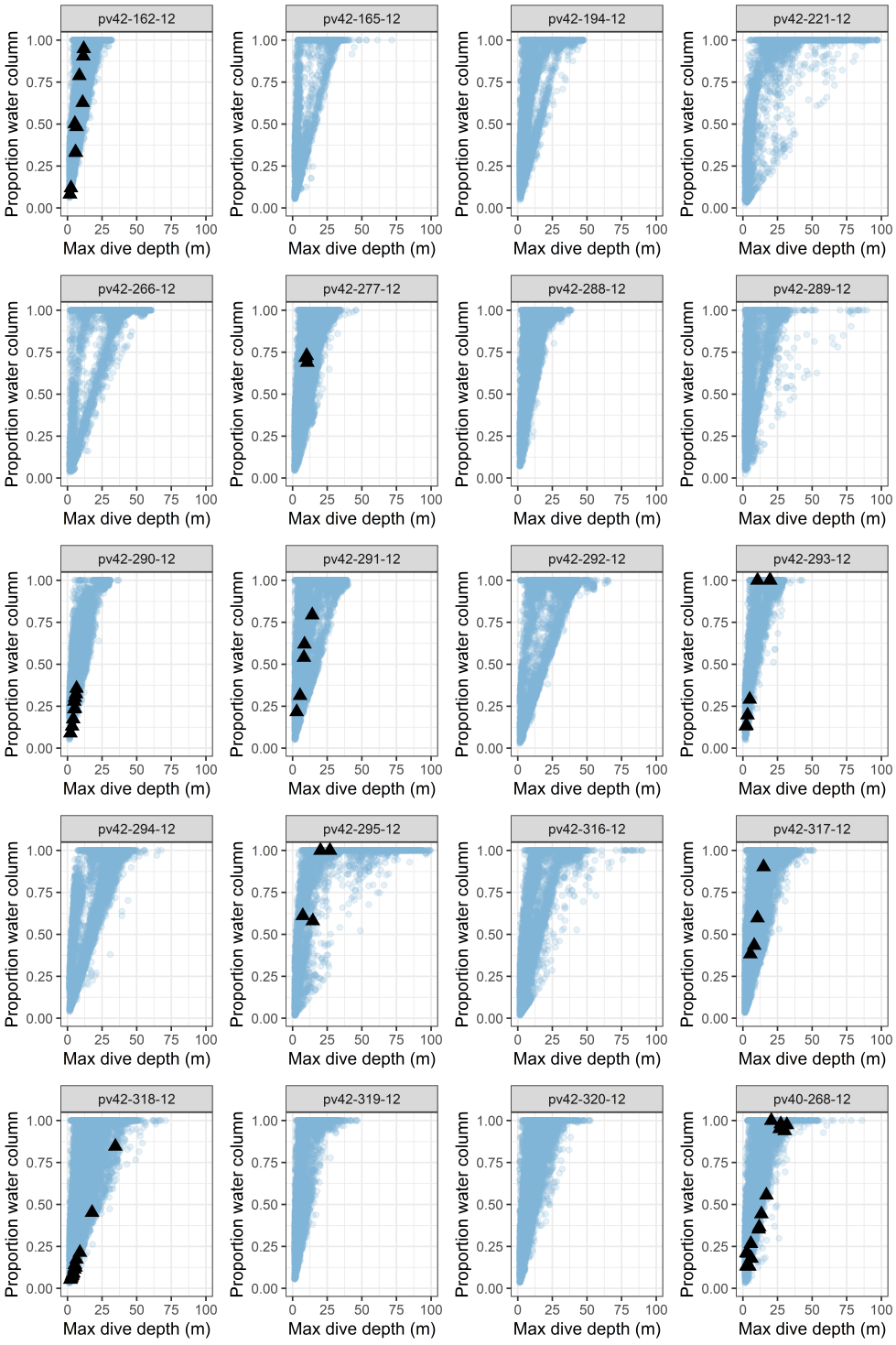


Figure S30 Scatterplot of proportion of the water column used and maximum dive depth (not used in analyses) for each dive in baseline data (blue dots) and identified behavioural responses (black triangles). Analyses were conducted using groups of three sequential dives. Shown here are the 20 seals (panels) which had encounters with pile driving during the study period.

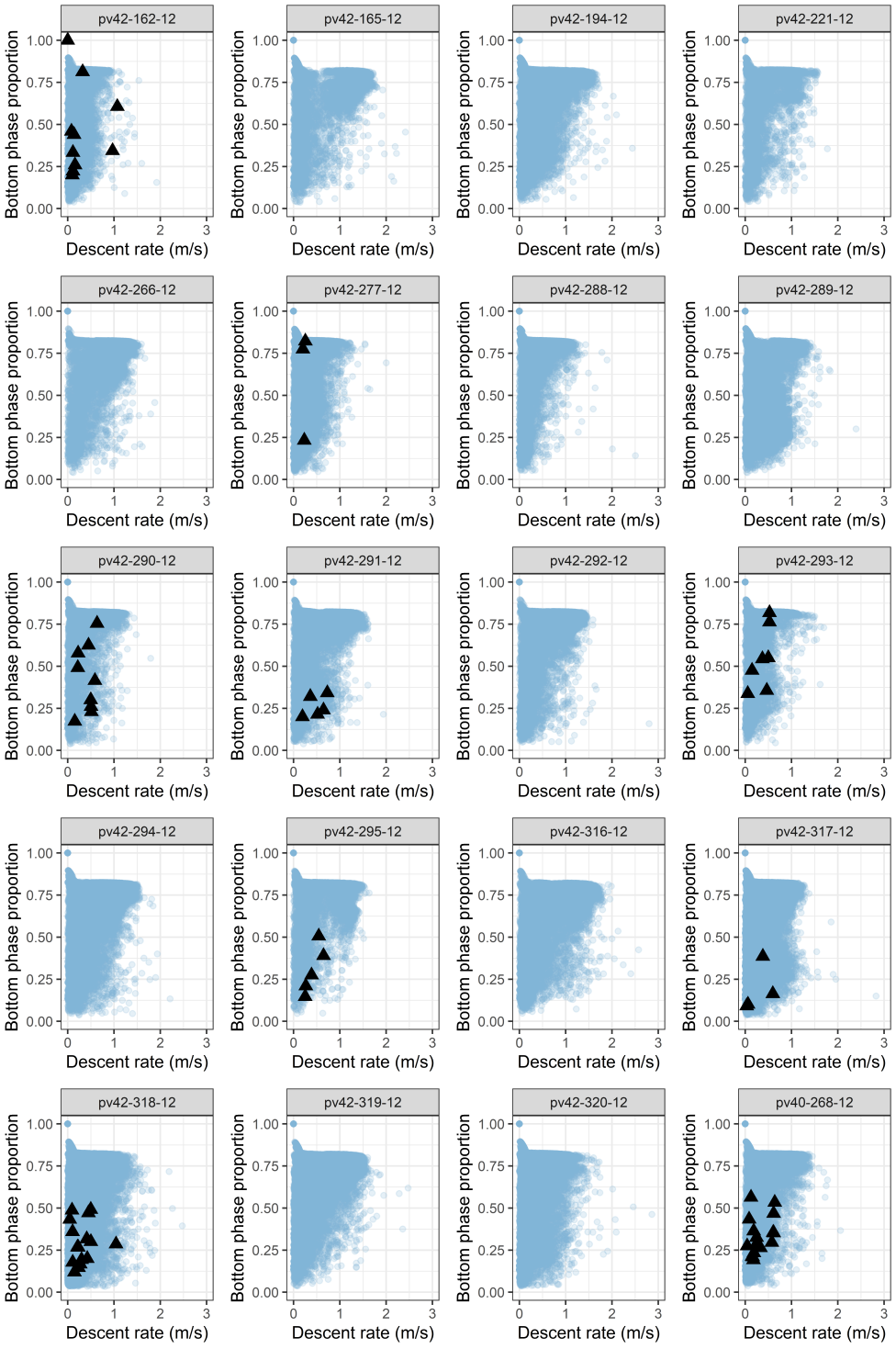


Figure S31 Scatterplot of proportion of the dive in the bottom phase and descent rate for each dive in baseline data (blue dots) and identified behavioural responses (black triangles). Analyses were conducted using groups of three sequential dives. Shown here are the 20 seals (panels) which had encounters with pile driving during the study period.

Table S2 Estimated dive response probabilities against single-strike sound exposure level (SELss) of pile driving. Shown are the mean and 95% credible intervals in 5 dB increments, estimated from posterior samples of the dose-response model.

	Probability of dive response		
SELss	**Mean**	**2.5%**	**97.5%**
(dB re $1\mu\text{Pa}^2 \cdot \text{s}$)		**quantile**	**quantile**
115	0.000	0.000	0.000
120	0.007	0.002	0.018
125	0.017	0.004	0.040
130	0.031	0.009	0.068
135	0.050	0.017	0.102
140	0.075	0.030	0.143
145	0.107	0.049	0.192
150	0.148	0.076	0.248
155	0.199	0.113	0.312
160	0.260	0.162	0.384
165	0.332	0.225	0.463
170	0.413	0.302	0.547
175	0.503	0.393	0.634
180	0.599	0.497	0.721
185	0.700	0.613	0.804
190	0.802	0.738	0.880
195	0.903	0.868	0.946
200	1.000	1.000	1.000

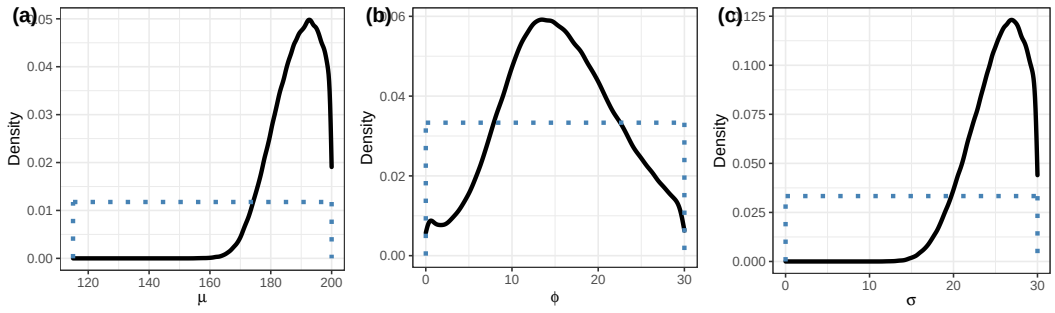


Figure S32 Posterior (black, solid line) and prior (blue, dotted line) densities for the parameters in the final vertical dose-response model with no covariates.

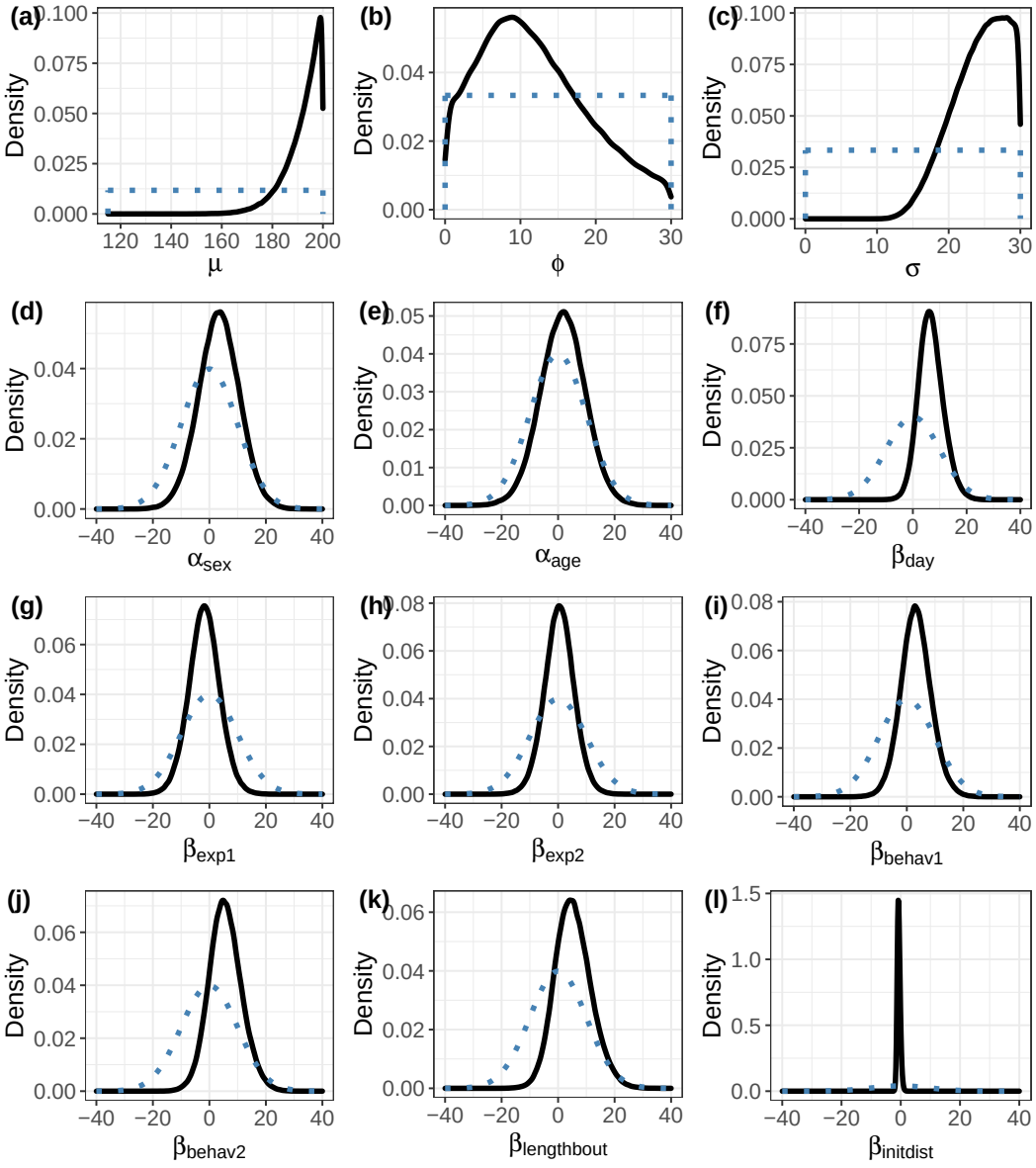


Figure S33 Posterior (black, solid line) and prior (blue, dotted line) densities for the parameters in the maximal model for vertical responses with all candidate covariates.

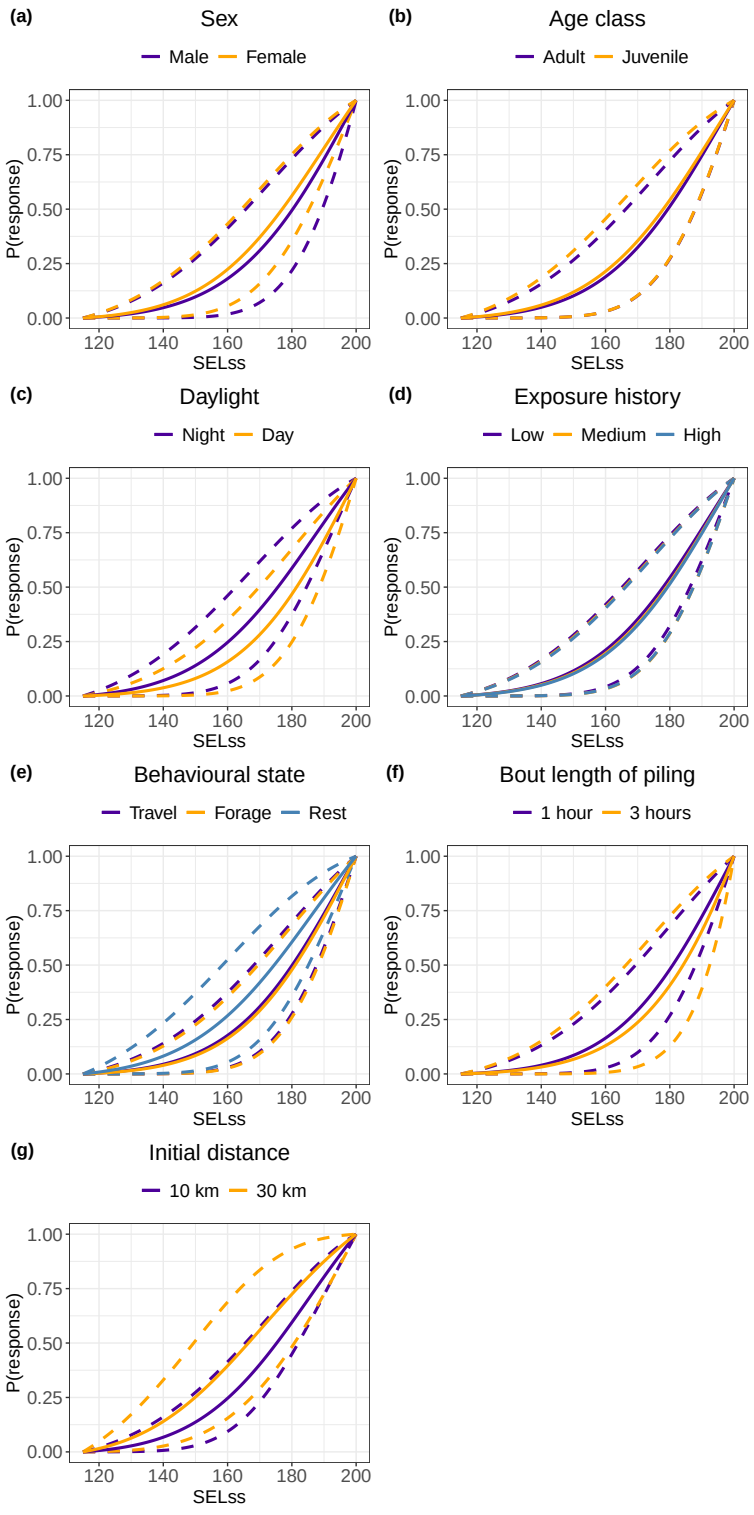


Figure S34 Estimated dose-response relationships for vertical responses in harbour seals as a function of single-strike sound exposure level (SELss in dB re $1\mu\text{Pa}^2\cdot\text{s}$, median across depth) of pile driving. Shown are results from the maximal model, with the predicted dose-response relationship from each covariate, assuming all were included in the model. For each covariate (panel), the mean (solid line) and 95% credible intervals (dashed lines) are illustrated.

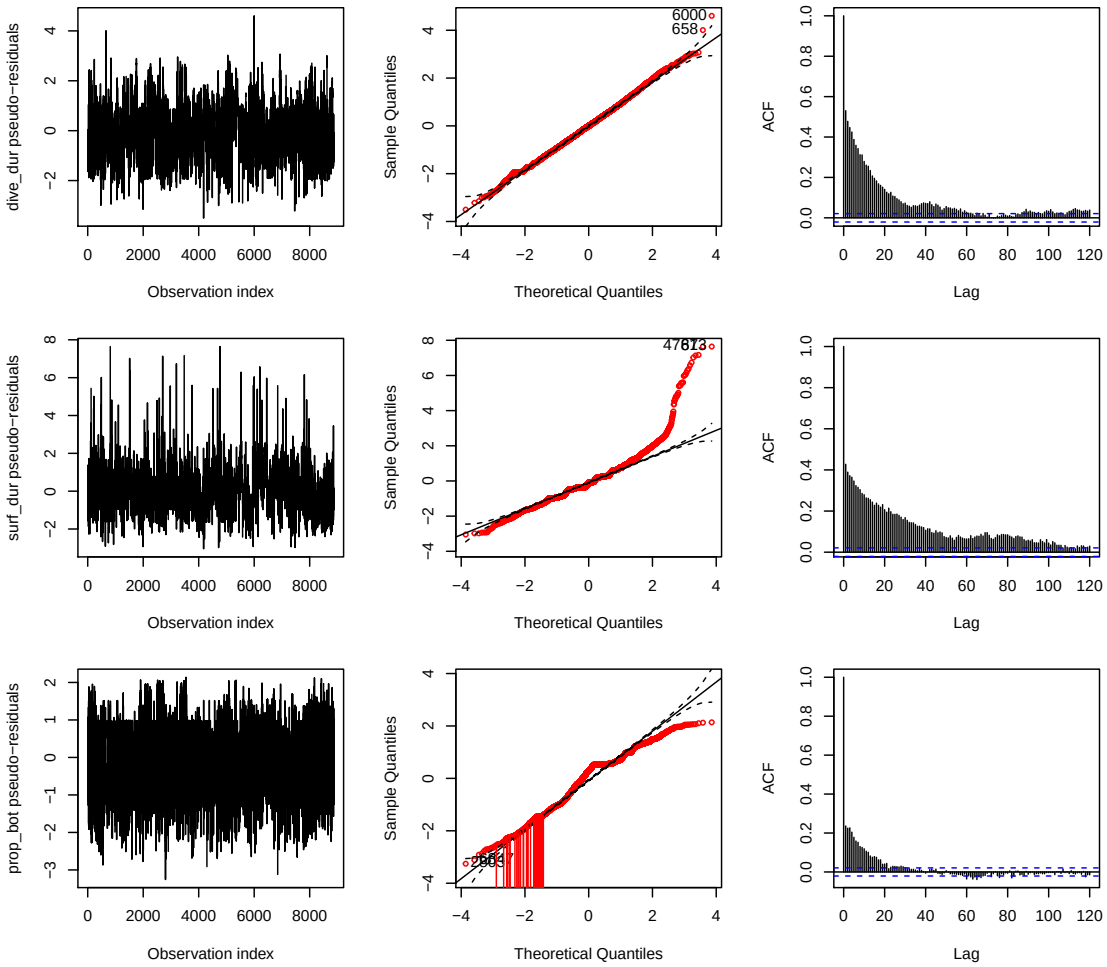


Figure S35 HMM pseudo-residuals, QQ-plots, and ACF plots for an example seal (pv42-221-12). Each row corresponds to the three observation metrics used in the full HMM (with piling and bathymetry covariates): (1) dive duration, (2) surface duration, and (3) proportion of the dive in the bottom phase.

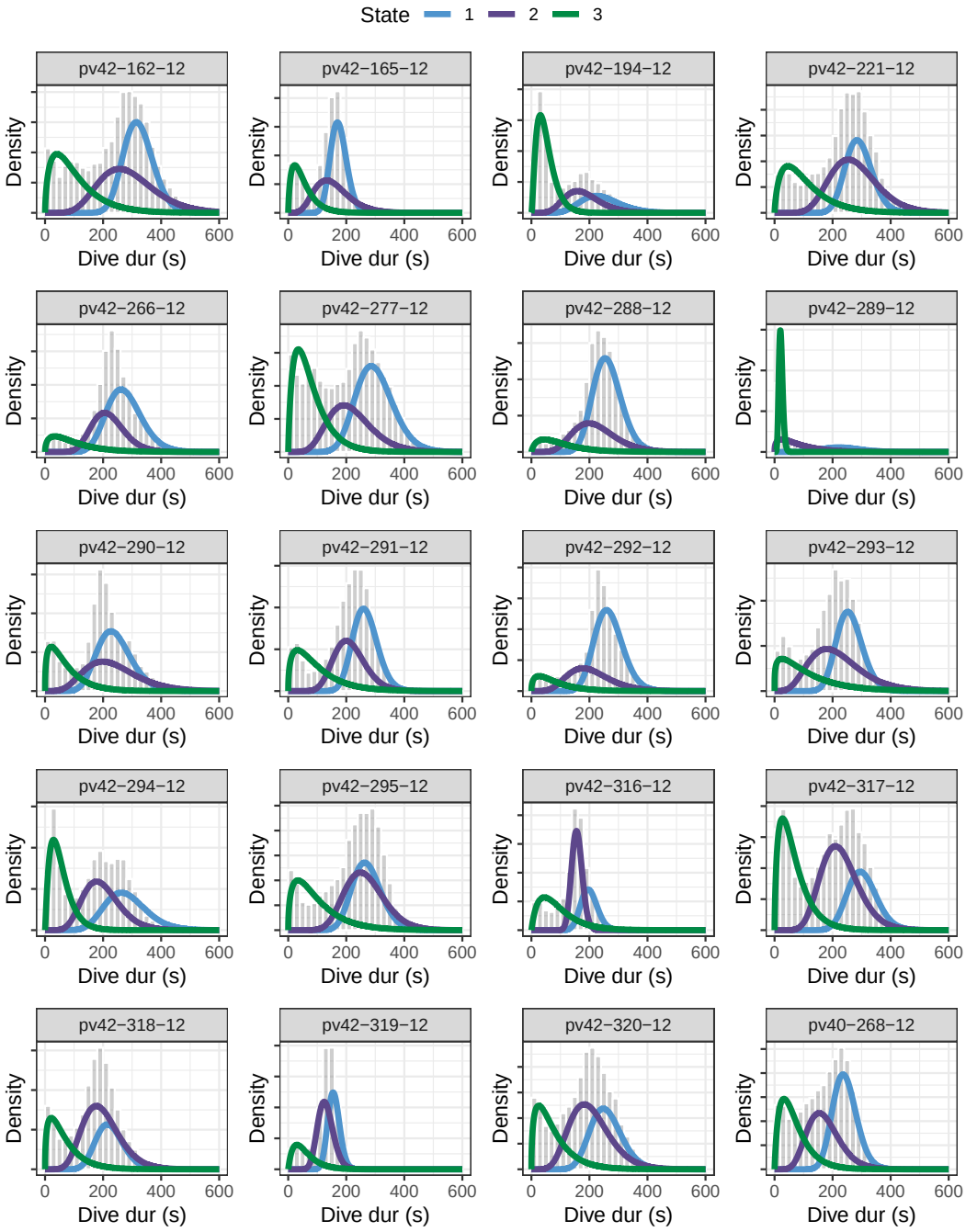


Figure S36 Fitted state-dependent distributions of dive duration (seconds) for all 20 seals. Shown in each panel are the estimates from a 3-state HMM with bathymetry and pile driving as covariates on the transition probabilities.

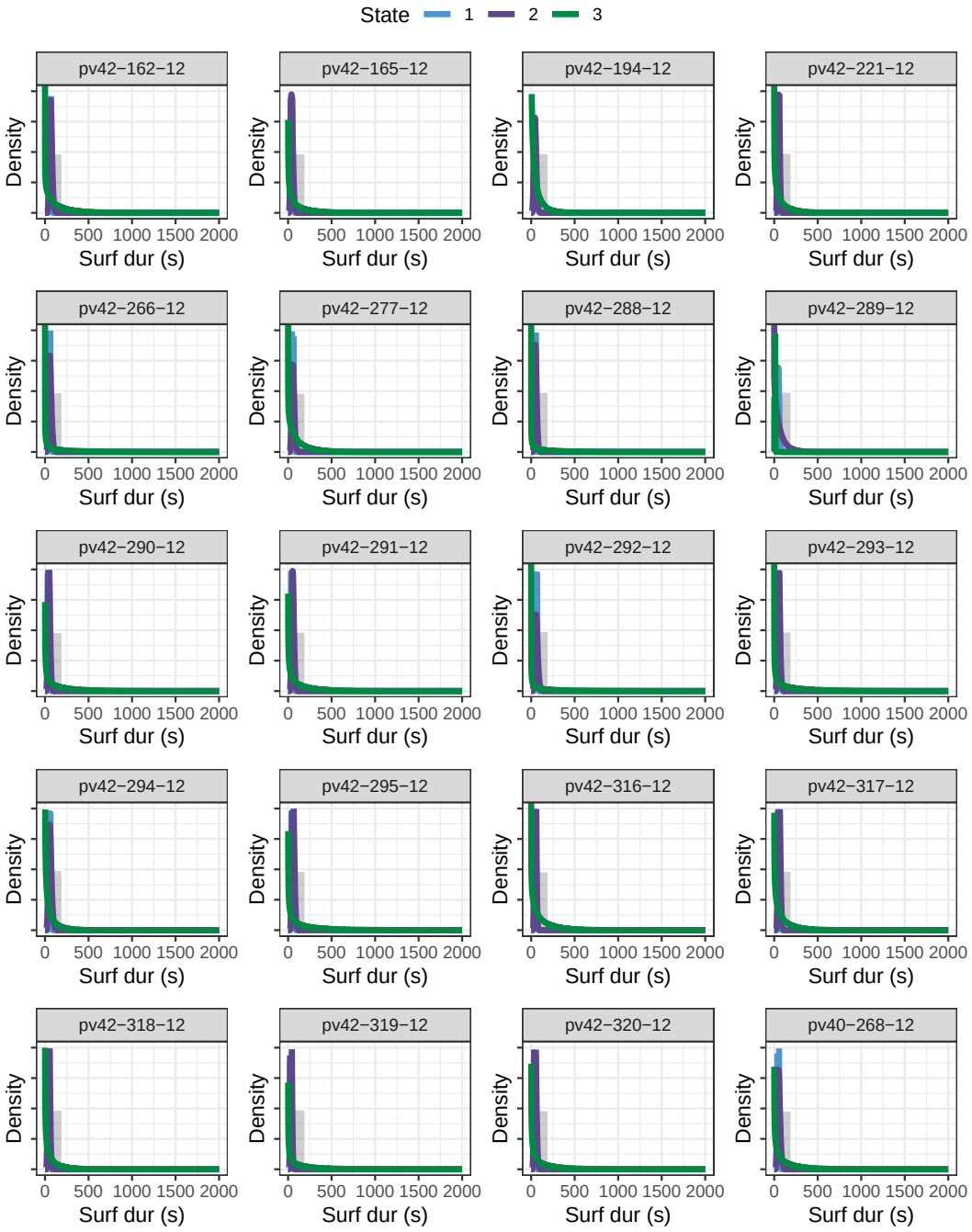


Figure S37 Fitted state-dependent distributions of surface duration (seconds) for all 20 seals. Shown in each panel are the estimates from a 3-state HMM with bathymetry and pile driving as covariates on the transition probabilities.

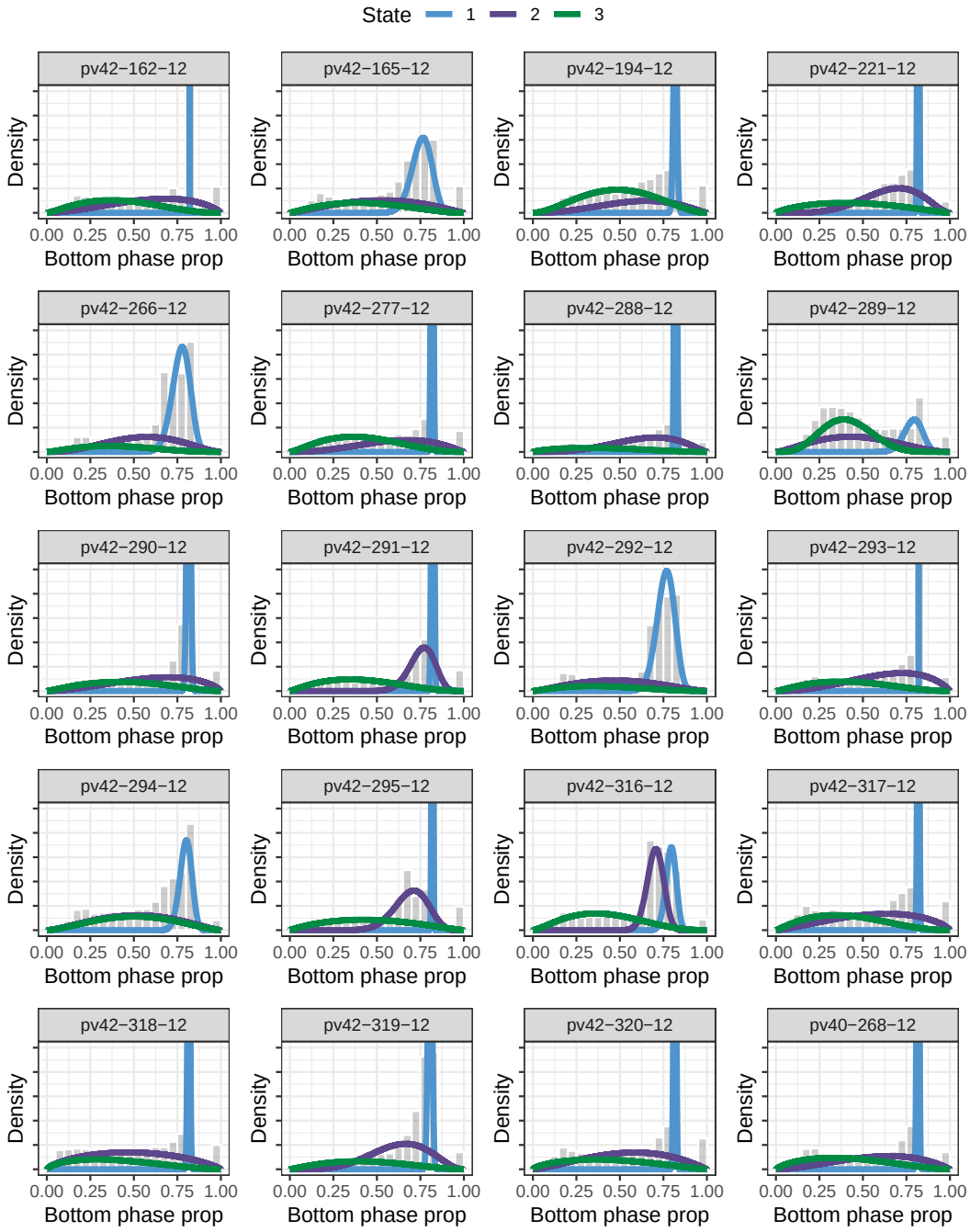


Figure S38 Fitted state-dependent distributions of proportion of dive in the bottom phase for all 20 seals. Shown in each panel are the estimates from a 3-state HMM with bathymetry and pile driving as covariates on the transition probabilities.

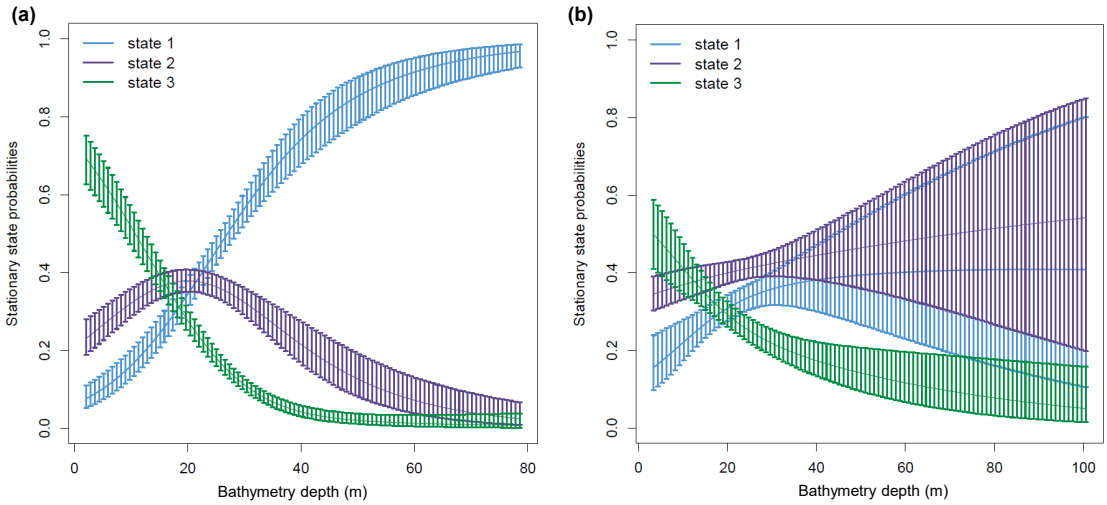


Figure S39 Probability of occupying each behavioural state (with 95% CI) as a function of available water depth (metres). Shown are the estimated stationary state probabilities (when piling=0) from two example seals from the full three-state HMM: (a) pv42-165-12 and (b) pv42-221-12.

Appendix E: Beaked whale DTAG simulations

Simulation study of Mahalanobis distance for beaked whale DTAG data

Katherine Whyte, Stacy DeRuiter, Catriona Harris, Len Thomas

As part of a 6-month leave-of-absence from my PhD, I worked at CREEM on the project described briefly below. This work was funded by the Centre for Research into Ecological and Environmental Modelling (CREEM, University of St Andrews) and the Atlantic Behavioral Response Study (Duke University).

Overall Aim

To assess the performance of Mahalanobis distance in detecting behavioural responses from high-resolution accelerometer tag data (e.g. DTAGs) on Cuvier's beaked whales (*Ziphius cavirostris*).

Simulating beaked whale DTAG data

We developed and used a simulator to generate data resembling that recorded from a DTAG, based on values observed in real data (Aguilar Soto et al., 2006; DeRuiter et al., 2013; Tyack et al., 2006). Data were simulated by:

1. Simulating a sequence of dive types (deep or shallow), via a first-order Markov chain.
2. For each dive of each dive type, simulating the durations of the dive phases (descent, bottom, ascent, surface), from a multivariate normal distribution.
3. For each dive phase, simulating the high-resolution tag time series data, via a multivariate autoregressive process.

We simulated data for baseline (normal) behaviour and behavioural responses to human disturbance.

Simulation study design

Data were simulated containing different types of response (avoidance, foraging disruption), different durations of response (short: 2 dives; long: 10 dives), and different strengths of response (weak, strong). An example of a simulated response is shown in Figure S40. Each of the eight simulated datasets contained 500 whales, and a control dataset with no behavioural response was also simulated. Various implementations of Mahalanobis distance based methods were then applied to the simulated datasets, to compare performance in detecting the simulated responses. Specifically, we modified the time window size, window overlap amount, location of the reference window, method of covariance matrix calculation, and the duration of baseline data. From this, we were able to quantify the false positive and detection rates of the method, and compare implementation options.

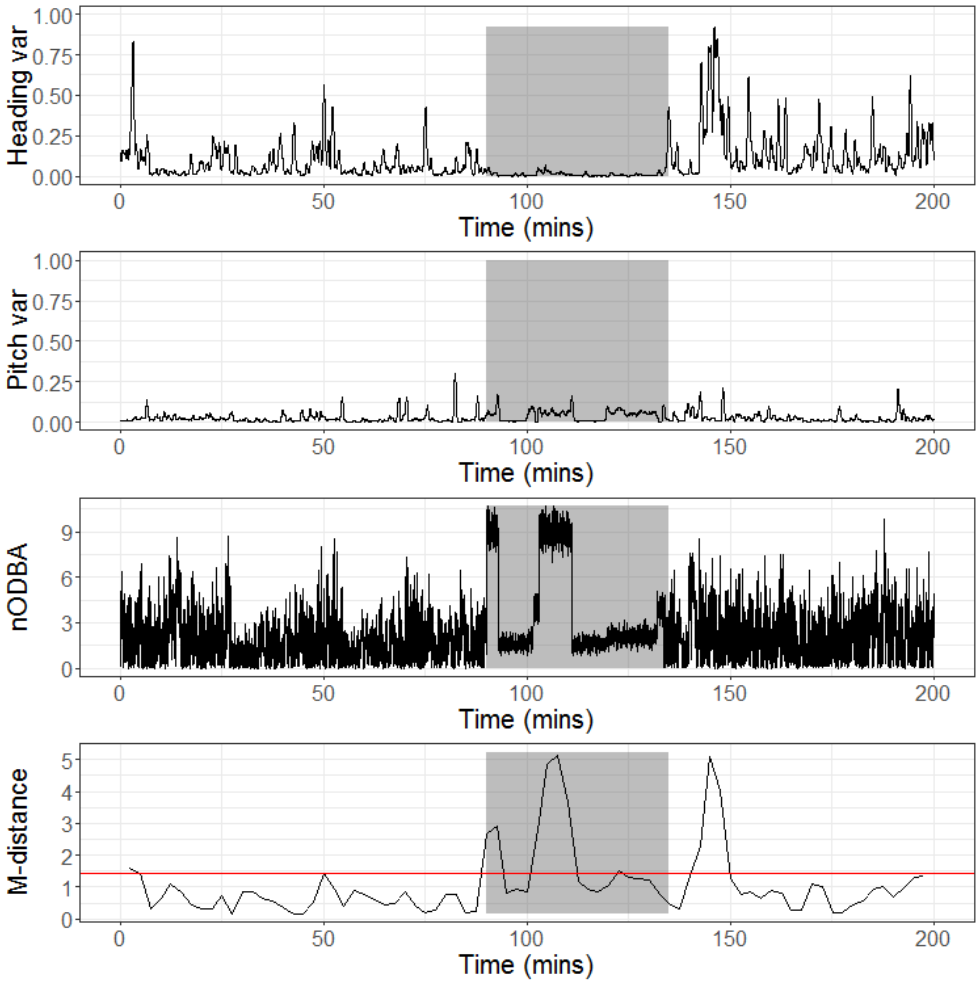


Figure S40 Example of a simulated beaked whale response. This example is a strong avoidance response, where the whale was simulated to have reduced variability in heading, reduced variability in pitch, and increased mean ODBA.

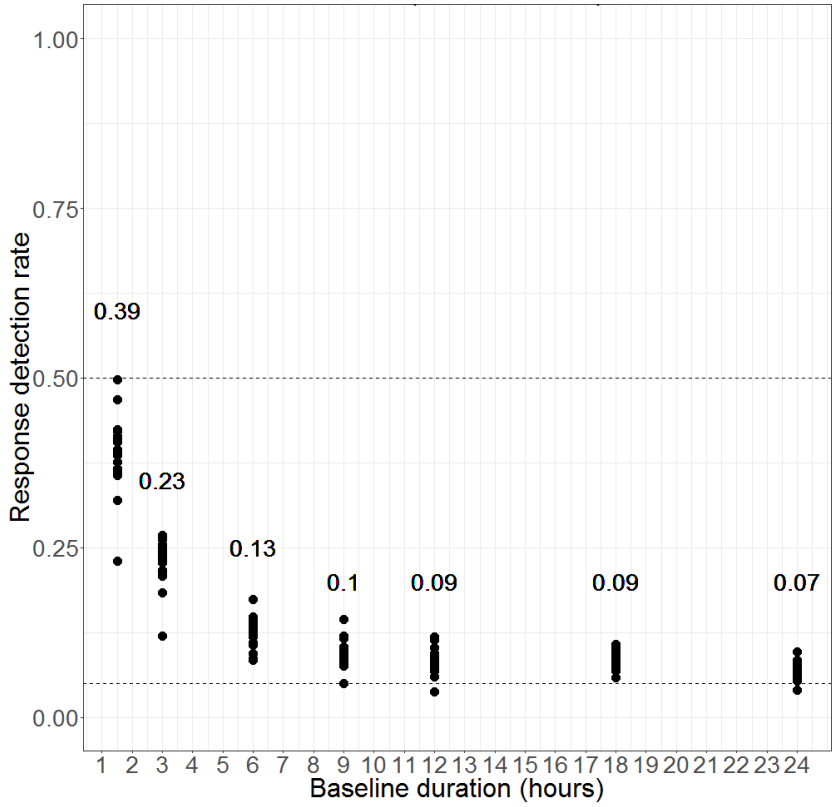


Figure S41 An example of the false positive rate (proportion of control simulations in which a response was detected) against baseline data duration. Each dot represents the results from one of the different Mahalanobis distance implementation options, and annotations denote the mean detection rate for each baseline duration. In this example, the covariance matrix was calculated using all of the simulated data for each animal.

Key findings

False positives

In general, the false positive rate (proportion of times a response was detected when none was present) was high when there was a short baseline duration; however, this declined with increasing baseline duration (Figure S41).

Different types of behavioural response

As may be expected, simulated responses which were stronger or longer in duration were generally detected more reliably. Overall, the simulated foraging disruption response had lower detection rates than the avoidance response. This method aims to detect extreme behaviours, and therefore it may be more challenging to detect reductions in behaviours, reduced behavioural variability, and the absence of behaviours.

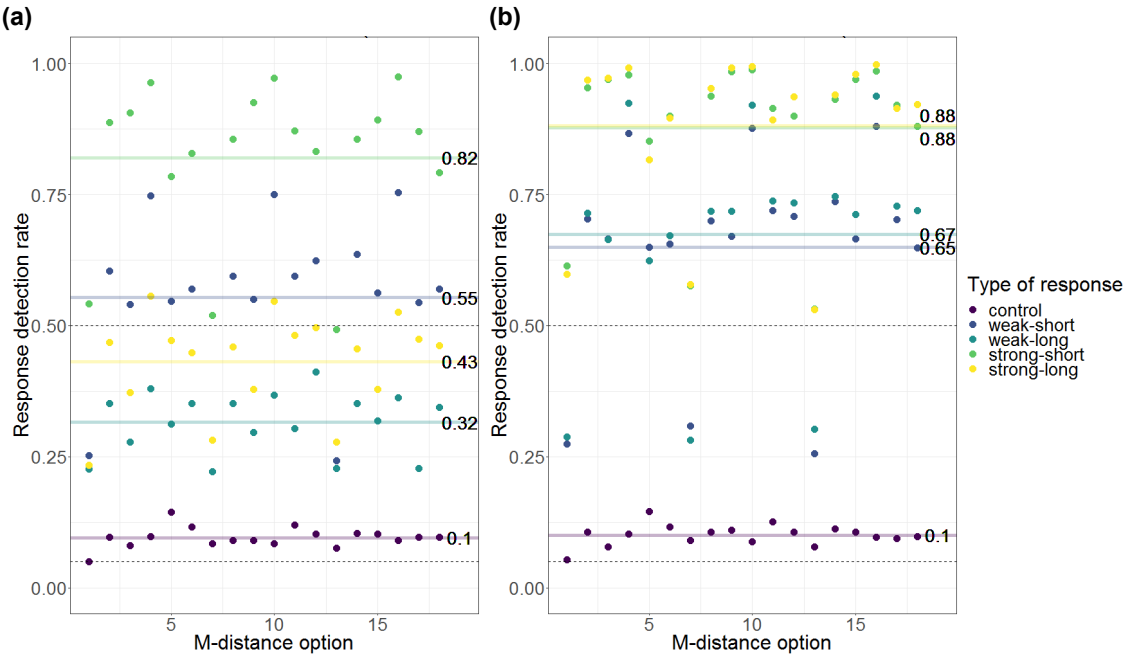


Figure S42 Example of response detection rates for an avoidance response when (a) the covariance matrix is based on the full dataset and (b) the covariance matrix is from the baseline data only. Shown are the detection rates for each Mahalanobis distance implementation option (x-axis) and different types of simulated response (colourscale). The average detection rate for each group is denoted by the horizontal coloured lines. This example is from simulations with 9 hours of baseline data.

Different Mahalanobis distance implementations

Using the covariance matrix on the baseline data only (compared to the entire dataset) generally increased the response detection rate, but had a limited effect on the false positive rate (Figure S42). Response detection rates were lower when using a pair of consecutive sliding windows, and the simulation study highlighted that there is a need to carefully choose an appropriate time window size (Figure S43). Large time windows may oversmooth data and lose information on the response, and small time windows may not contain enough information to detect a response. The amount of window overlap used appeared to have a negligible effect on the performance of the method (Figure S43). Overall, the patterns observed were consistent across different types, strengths and durations of simulated responses.

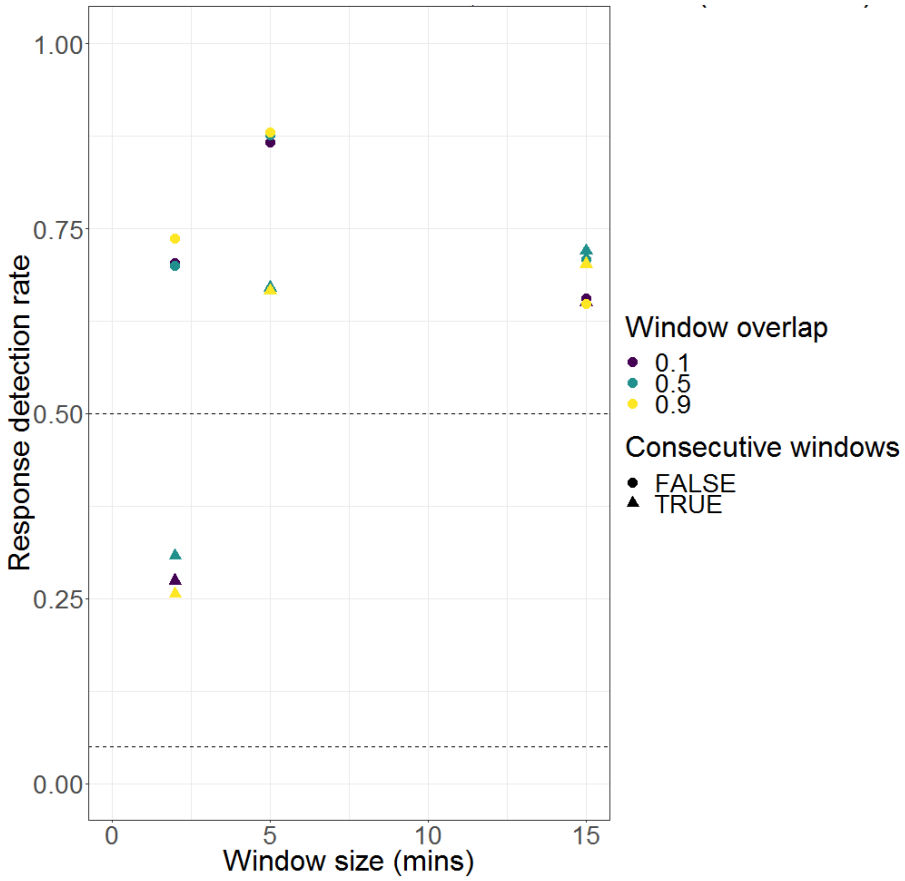


Figure S43 Example of response detection rates for different Mahalanobis distance implementation options: window size (x-axis), window overlap proportion (colourscale), and whether a pair of consecutive sliding windows was used (shape). This example is from simulations of a weak, short, avoidance response with 9 hours of baseline data.

Appendix F: JASA paper (Whyte et al. 2020)

Estimating the effects of pile driving sounds on seals: pitfalls and possibilities

Running title: Effects of pile driving on seals

Katherine F. Whyte [1,a),b)], Debbie J.F. Russell [1,b)], Carol E. Sparling [2], Bas Binnerts [3], and Gordon D. Hastie [1]

1. *Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, Fife, KY16 8LB, UK*
2. *SMRU Consulting, Scottish Oceans Institute, University of St Andrews, St Andrews, Fife, KY16 8LB, UK*
3. *TNO, Acoustics and Sonar expertise group, Oude Waalsdorperweg 63, 2597 AK The Hague, Netherlands*

---

a) *Electronic mail: kfw5@st-andrews.ac.uk*

b) *Also at: Centre for Research into Ecological and Environmental Modelling, The Observatory, University of St Andrews, St Andrews, Fife, KY16 9LZ, UK*

This paper is part of a special issue on The Effects of Noise on Aquatic Life.

Word count: 6654 words + 2246 for figures/tables = 8900

Date uploaded: 31st December 2019; revised 1st May 2020; revised 14th May 2020

ABSTRACT

Understanding the potential effects of pile driving sounds on marine wildlife is essential for regulating offshore wind developments. Here, tracking data from 24 harbour seals were used to quantify effects and investigate sensitivity to the methods used to predict these. The Aquarius pile driving model was used to model source characteristics and acoustic propagation loss (16 Hz–20 kHz). Predicted cumulative sound exposure levels (SELcums) experienced by each seal were compared to different auditory weighting functions and damage thresholds to estimate temporary (TTS) and permanent (PTS) threshold shift occurrence. Each approach produced markedly different results; however, the most recent criteria [Southall *et al.*, 2019] suggests that TTS occurrence was low (17% of seals). Predictions of seal density during pile driving (from [Russell *et al.*, 2016]) were compared to distance from the wind farm and predicted single-strike sound exposure levels (SELss), by multiple approaches. Predicted seal density significantly decreased within 25 km or above SELss (averaged across depths and pile installations) of 145 dB re 1 $\mu Pa^2 \cdot s$. However, there was substantial variation in SELss with depth and installation, and thus in the predicted relationship with seal density. These results highlight uncertainty in estimated effects, which should be considered in future assessments.

I. INTRODUCTION

In order to meet ambitious climate change targets, the demand for renewable energy is increasing and bringing substantial industrial activity to marine environments. In particular, the number, size, and capacity of offshore wind farms has been growing rapidly and is expected to continue to increase (Bailey et al., 2014; Breton et al., 2009). This expansion has been particularly prevalent in European waters, where there are currently more than 4,500 grid-connected offshore wind turbines across eleven countries, equivalent to a capacity of 18,499 MW (WindEurope, 2019).

In predicting and assessing the environmental impact of these offshore wind farms on the surrounding marine life, one of the key uncertainties is the potential effects of underwater construction noise. Of particular concern are the effects of high intensity sounds produced during pile driving, where brief impulsive sounds with source levels of up to 250 dB re 1 μPa @ 1 m (peak-peak) can be produced every 1-2 seconds (Bailey et al., 2010). The at-sea movements of harbour seals (*Phoca vitulina*) overlap with many areas of current and proposed development (Russell et al., 2014; Sharples et al., 2012), and so there are concerns that these sounds may damage hearing, elicit overt behavioural responses, and/or exclude seals from areas of their natural habitat (Hastie et al., 2015; Russell et al., 2016; Thompson et al., 2013). To accurately predict the effects of pile driving and determine how these could be mitigated, it is critical to understand the nature and severity of these potential effects and the sound levels at which they occur.

Estimating the effects of anthropogenic noise on marine mammal hearing can be challenging. Using available data on hearing sensitivities and hearing damage across species, Southall et al.

65 (2007) derived estimates of the minimum noise exposure required for the onset of temporary
66 (TTS) and permanent (PTS) threshold shifts in hearing sensitivity. They also generated a series
67 of frequency-weighted hearing sensitivity curves for different functional groups of marine
68 mammals (M-weightings). For pinnipeds underwater, TTS was predicted to occur at M-weighted
69 24-hour cumulative sound exposure levels (SELcum) of 171 dB re 1 $\mu Pa^2 \cdot s$ and PTS at 186 dB
70 re 1 $\mu Pa^2 \cdot s$ for impulsive sounds such as pile driving. These weighting functions and TTS/PTS
71 thresholds were subsequently updated in 2019, incorporating the most recent scientific
72 information on hearing abilities and auditory damage for each marine mammal species group
73 (Southall et al., 2019). In general for pinnipeds, these new weightings were slightly less
74 conservative. Pinnipeds were also subdivided into two groupings (phocids, otariids); for phocid
75 pinnipeds in water, the SELcum thresholds for impulsive sounds are now estimated to be 170
76 and 185 dB re 1 $\mu Pa^2 \cdot s$ for TTS and PTS respectively. It should be noted that for seals the
77 weighting functions and TTS/PTS thresholds for impulsive sounds described in Southall et al.
78 (2019) are the same as those provided by the US National Marine Fisheries Service (NMFS,
79 2016, 2018). Faulkner et al. (2019) simulated how these two different criteria may alter the
80 predicted effect zones from a variety of modelled noise sources, comparing the relative
81 differences between Southall et al. (2007) and Southall et al. (2019). For phocids, they concluded
82 that the more recent weighting functions are likely to substantially reduce the estimated range of
83 PTS risk, e.g. from approximately 10 km to 2 km for a theoretical scenario involving pile driving
84 24-hour SELcums at offshore wind farms in the North Sea (Faulkner et al., 2019).

85 A limited number of studies have investigated the effects of pile driving sounds on harbour seal
86 hearing and behaviour. Recent playbacks of broadband piling sounds (~500–800 Hz, single-
87 strike sound exposure level (SELss) of 152 dB re 1 $\mu Pa^2 \cdot s$ at 1 m depth, 2 m from the source)

88 were found to cause onset of TTS at unweighted SELcums of around 192 dB re 1 $\mu Pa^2 \cdot s$ in two
89 harbour seals in captivity (Kastelein et al., 2018). Small TTSs (2–4 dB) occurred in that
90 experiment and hearing recovered within 60 minutes. However, in the wild, animals may
91 encounter pile driving sounds at higher received levels than that tested therein. Two studies
92 (Hastie et al., 2015; Russell et al., 2016) investigated the predicted sound exposure and at-sea
93 behaviour of tagged harbour seals near pile driving activity at an offshore wind farm. Hastie et
94 al. (2015) calculated predictions of auditory injury in each tagged seal as a result of exposure to
95 piling sounds. The analysis showed that half of the tagged animals received predicted M-
96 weighted 24-hour SELcums that would cause PTS (based on Southall et al. (2007)). In addition,
97 there was a significant reduction in seal density up to 25 km from the wind farm during periods
98 of piling activity, relative to non-piling periods (Russell et al. (2016)). The magnitude of the
99 observed reduction decreased with increasing distance from the piling location, and recovery
100 time was relatively short, with seal density returning to pre-piling levels within two hours of the
101 cessation of piling.

102 Although both of these findings represented an important step forward in our understanding, the
103 direct application of these results in Environmental Impact Assessments (EIAs) may be
104 challenging. For example, since estimates of piling sound exposure (Hastie et al., 2015) were
105 completed, updated auditory weighting functions and thresholds for the onset of hearing damage
106 have been published (Southall et al., 2019). Further, to contextualise predicted changes in density
107 over space, Russell et al. (2016) illustrated how seal density changed in relation to distance from
108 the middle of the wind farm and in relation to predicted SELs (averaged across all pile
109 installations) for the loudest and quietest parts of the water column. However, although not
110 explicitly stated, both of the relationships (distance and SEL) presented in Russell et al. (2016)

111 represent the expected change in seal density for cumulatively increasing zones around pile
112 driving. For example, the presented change in seal density at 20 km represents the change for all
113 spatial cells within 20 km of the wind farm, and seal density at 40 km represents the change for
114 all cells within 40 km. These results could be misinterpreted and such cumulative predictions are
115 not particularly appropriate for the finer scale quantitative analyses often required to inform
116 EIAs. The predicted change in seal density for any given location also reflected a wide range of
117 predicted SELs (across depths and pile installations). To address these potential issues and make
118 the results more applicable to EIAs, we use the seal tag data from the previous studies to (1)
119 compare how estimates of SELcum and auditory damage may differ when different weighting
120 functions are applied to them; (2) quantify the relationship between predicted seal density change
121 and distance/SEL for both cumulative and annulus zones; (3) compare five different approaches
122 to combining SELs across pile installations and depths; and (4) investigate the robustness of
123 these relationships.

124 **II. METHODS**

125 **A. Seal tag data**

126 In January 2012, harbour seals were caught on or near haulout sites on intertidal sandbanks in
127 The Wash, south-east England, UK. To record the movements and dive behaviour of seals
128 around active pile driving, all animals were fitted with a SMRU Instrumentation GPS telemetry
129 tag (hereafter GPS/GSM tags; SMRU Instrumentation, University of St Andrews, Fife, UK).
130 Seals were first anesthetized using Zoletil® or Ketaset® in combination with Hypnovel®, and
131 GPS/GSM tags were attached to the fur at the back of the neck using a fast-setting two-part

epoxy adhesive or Loctite® 422 Instant Adhesive. All seal handling and procedures were carried out under Home Office Licence 60/4009.

Out of the 25 deployed tags in The Wash, three tags collected data for less than two days and so were excluded from further analyses. Two seals from a concurrent study approximately 200 km to the south (in the Thames) moved into The Wash during pile driving, and so were included in the dataset. This resulted in a total sample size of 24 individuals (11 males, 13 females) (details provided in Electronic Supporting Information[1]).

The tags provided GPS locations approximately every 15 minutes, as well as nine depth data points per dive and records of all haulout times. The data were cleaned and erroneous locations removed based on thresholds of residual error and the number of satellites. For more details of the data collection and study site, see Hastie et al. (2015) and Russell et al. (2016).

B. Pile driving

Operational data on pile driving at Lincs offshore wind farm were provided by Centrica plc. Throughout the period of the 2012 seal tag deployment, 27 monopiles were installed at Lincs by pile driving between 28th January and 11th May 2012 (Table I). A total of 77,968 piling strikes occurred during the study with a mean strike energy of 1202 (SD = 613) kJ. For further information on the pile driving, see Hastie et al. (2015).

Opportunistic recordings of pile driving were available from two sources: an autonomous moored sound recorder (DSG-Ocean Acoustic Datalogger; Loggerhead Instruments, Sarasota, FL, USA) at ~9 m depth and a range of 4,900 m from the pile driving location, and a series of boat-based recordings at ~1 m depth between 1,000 and 9,500 m from pile driving (using a Reson TC 4014 hydrophone with a Brüel and Kjaer amplifier (type 2635) and a calibrated Avisoft Ultrasound gate

154 416 digital acquisition system at a sample rate of 192 kHz)(for further information, see Hastie et
155 al. (2015)). These recordings covered the full range of pile driving blow energies and were
156 compared to the estimates made using the acoustic models.

157 **C. Acoustic modelling**

158 To estimate the sound levels resulting from the piling across the study area, a series of acoustic
159 modelling approaches were carried out. The Aquarius pile driving model (for detailed
160 description of the model and its validation, see de Jong et al. (2019)) was used to model source
161 characteristics and acoustic propagation loss. Note that this is a different sound propagation
162 modelling approach to the one used by Hastie et al. (2015) and Russell et al. (2016).

163 The Aquarius model uses information on the properties of the hammer and the pile (Table I) to
164 determine a source excitation spectrum using the model described by Deeks and Randolph
165 (1993). This source spectrum is integrated into a range dependent propagation model (normal
166 mode based adiabatic propagation model using the KrakenC (Porter, 2001) model to compute the
167 propagating modes) to predict acoustic propagation loss across the study area, incorporating
168 information on seabed characteristics and water depth. Here, the bathymetry was set to Mean Sea
169 Level (MSL) and the modelled receiver resolution was chosen equal to 1 m, which leads to a
170 smooth solution with depth such that linear interpolation can be used to obtain the received
171 levels at intermediate depth. The seabed was assumed to be homogenous, with properties
172 corresponding to medium sand (grain size parameter Φ=1.5) obtained from table 4.18 in Ainslie
173 (2010). This was the most common value in the considered modelling area, using data from the
174 EMODnet Bathymetry Data Portal. The properties of the water column were set at a
175 compressional sound speed of 1,500 ms^{-1} and a density of 1,024 kg/m^3, and the Thorp attenuation

model was used for volume attenuation (Ainslie, 2010; Sehgal et al., 2009). It should be noted that the effects of losses due to sea surface scattering and absorption were not considered for the purposes of the modelling.

Depth explicit model predictions were output as estimated single strike sound exposure levels ($SEL_{ss,ref}$, dB re 1 $\mu Pa^2 \cdot s$) at a reference strike energy of 1,000 kJ across a series of spatial grids within the study area at ~279 m resolution (Longitude: from -1 to 3 degrees with a 15 s resolution, Latitude: from 52 to 55 degrees with a 9 s resolution). Individual grids were produced for each 2.5 m depth bin (from 2.5 to 107.5 m depth); sound levels below the seabed were indicated by a "NaN" value. Model predictions included estimated SELss with three different frequency weightings applied to them; these were (i) unweighted, (ii) Pinnipeds-in-Water M-weighted (M_{pw}) (Southall et al., 2007), and (iii) Phocids-in-Water weighted (PCW) (Southall et al., 2019). Frequencies from 16 Hz to 20 kHz were modelled, using third octave centre frequency bands.

D. Acoustic exposure of the tagged seals

The tag data consisted of a series of time-stamped GPS locations when the seal was at the water surface. Further, during each dive, the tag provided dive depths at nine points distributed equally in time throughout each dive. As seal depths were derived from pressure sensor readings on-board the tag, they were measured relative to the water surface, leading to a potential mismatch with the original bathymetry data, which were relative to chart datum at the Lowest Astronomical Tide (LAT). Water depths relative to Mean Sea Level (MSL) were derived by applying the United Kingdom Hydrographic Office Vertical Offshore Reference Frame (VORF)

197 Lowest Astronomical Tide (LAT) correction (Iliffe et al., 2013) for the study area. These water
198 depths at MSL were used for the acoustic modelling and corresponding received levels for seals.

199 During periods of pile driving, tracks of seals were linearly interpolated between successive GPS
200 locations to provide estimated locations of seals at the estimated time-of-arrival of sound from
201 each pile driving strike (assuming a sound speed of 1,500 ms^{-1}). Similarly, dive depths at each of
202 these interpolated locations were estimated through linear interpolation between successive
203 measured dive depths. Together, these provided the estimated 3D locations of each seal at the
204 time it received the sound from all pile driving strikes for each piling location.

205 Each seal 3D location was matched to the corresponding spatial grid cell and the closest 2.5 m
206 depth bin (from 2.5 to 107.5 m depth) in the acoustic model, and the received $SEL_{ss,ref}$ was
207 identified based on propagation loss estimates at the associated location and depth for each
208 individual pile driving pulse. Information on the blow energy of each strike was then used to
209 scale the modelled reference $SEL_{ss,ref}$ (at 1,000 kJ strike energy) to obtain final estimates of
210 received SELss at each seal 3D location. This was carried out through energetic (broadband)
211 scaling of the SELss spectrum using Equation 1 to calculate the value that is added to the
212 modelled $SEL_{ss,ref}$:

213 Equation 1.

214 $$SEL_{ss} = SEL_{ss,ref} + 10\ log_{10}\frac{E}{E_{ref}}$$

215 where E is the energy (kJ) of the pile driving strike, E_{ref} is the reference strike energy (1,000 kJ),
216 $SEL_{ss,ref}$ is the modelled single strike sound exposure level at the reference strike energy, and
217 SELss is the resulting scaled single strike sound exposure level (dB re 1 $\mu Pa^2 \cdot s$).

E. Predictions of auditory damage

Auditory damage (in the form of hearing threshold shifts) was predicted for each tagged seal using three approaches. These were based on: 1) a threshold based on results from previous studies of TTS onset in harbour seals as a result of exposure to pile driving sounds (Kastelein et al., 2018) (unweighted); 2) the approach developed by Southall et al. (2007) for evaluating the likelihood of TTS and PTS in pinnipeds exposed to anthropogenic sound (M_{pw}); and 3) the updated approach described by Southall et al. (2019) for evaluating the likelihood of TTS and PTS in phocid seals exposed to anthropogenic sound (PCW). Previously, Hastie et al. (2015) used approach 2) to estimate the potential for auditory damage in tagged seals as a result of exposure to pile driving sounds during the installation of the Lincs offshore wind farm.

For each seal, estimated received SELss were summed over each 24-hour period (Julian day) containing pile driving to calculate the 24-hour SELcum under each method (unweighted, M_{pw}, and PCW):

Equation 2.

$$SELcum = 10\log_{10}\left\{\sum_{n=1}^{N} 10^{\frac{SEL_n}{10}}\right\}$$

where SELcum is the cumulative sound exposure level of all N piling strikes within the 24-hour period, and SEL_n is the received SELss for each piling strike n. For the purposes of estimating auditory threshold shifts, an 'effective quiet' value of 124 dB re 1 µPa (Finneran, 2015) was assumed (the highest SPL of a sound that will neither produce significant TTS nor retard recovery from TTS from prior exposure to a higher level). Each 24-hour SELcum was then compared to published TTS and PTS onset thresholds under each approach (Table II). It should

be noted that, although the M_{pw} and PCW weightings are based on exposure during a 24-hour period, the unweighted criteria (Kastelein et al., 2018) is based on the threshold at which TTS was observed at two hearing frequencies (4 and 8 kHz) in a 6-hour experimental setting.

F. Changes in seal density in relation to pile driving

Russell et al. (2016) generated population-level predictions of the at-sea density of seals during piling and breaks in piling. The movements of individual seals in response to piling were not modelled directly. These population-level predictions were based on analyses of 23 of the tagged harbour seals (individual pv42-194-12 was excluded as in one trip it travelled much further than the other individuals, leading to issues in specifying the accessible spatial area for all seals- see Russell et al. (2016) for details). The analyses were restricted to return trips from haulouts within The Wash and comprised a use-availability design within a generalised estimating equation (GEE) framework. This approach was used as it enabled the study to consider the entire accessible area for seals in The Wash, and model seal density in an area with a complex coastline. The GEE approach also enabled generation of uncertainty estimates robust to the presence of residual autocorrelation within individuals. Once the optimal models for seal density during piling and non-piling periods were fit, the differences in these two distributions on a 5 x 5 km resolution (867 spatial cells) were quantified, and predictions of percentage of the at-sea population in each cell were made. A parametric bootstrap from the GEE model was used to calculate the 95% confidence intervals (CIs) for both the predicted density (percentage of the at-sea population) and predicted percentage change in density (non-piling to piling).

Here, we compare how the predicted percentage change in seal density (between non-piling and piling) relates to both the distance from the centre of Lincs wind farm and the predicted received

SELss at each cell location. Predictions could not be made relative to the exact piling locations as, for the GEE model, seal location data were pooled across piling events and so contained several different piling locations. In Russell et al. (2016), the presented relationship was in cumulative zones of increasing distance: each increment represents all cells equal or less than that distance, e.g. the predicted change in seal density value at 40 km represents the change in all cells within a distance of ≤ 40 km from the wind farm. Here, we also quantify how this relationship changes in annulus zones with 5 km increments: each increment represents the previous 5 km, e.g. the predicted change in seal density value at 40 km represents the mean for all cells with distances of 35 to 40 km. We also quantify the relationship between predicted seal density and received SELss for both cumulative and annulus approaches. In cumulative zones, the predicted change in density at 135 dB re 1 $\mu Pa^2 \cdot s$ represents the change in all cells with a received level of ≥ 135 dB re 1 $\mu Pa^2 \cdot s$. In annulus zones, the predicted change in density at 135 dB re 1 $\mu Pa^2 \cdot s$ represents the change for all cells with estimated SELss of 135 to 140 dB re 1 $\mu Pa^2 \cdot s$. In both the distance and SELss relationships, the first zone (that closest to the wind farm) is the same between cumulative and annulus approaches (e.g. 0–5 km (annulus) is the same as ≤ 5km (cumulative), 175–180 dB re 1 $\mu Pa^2 \cdot s$ (annulus) is the same as ≥175 dB re 1 $\mu Pa^2 \cdot s$ (cumulative)). The approaches differ in how the subsequent estimates are calculated, with the annulus approach looking at seal density in each distance/SELss increment, and the cumulative approach increasing the zone size each time by adding in seal densities at larger distances/lower SELss. The cumulative predictions were repeated here for clear comparison with the annulus zones, as previous results in Russell et al. (2016) used a different acoustic propagation model. By both annulus and cumulative approaches, it was necessary to consider estimated received levels across piling events and depths. Therefore, the outputs of acoustic models for each of the 27

284 piling locations had to be combined. To investigate the relationship between percentage change
285 in density and estimated SELss, we used five approaches to combining SELss across piles and
286 depths:

287 1. Mean SELss (averaged across depths and the 27 piles) (Fig. 2);
288 2. Lower 95% CI of SELss across piles (averaged across depths) (Fig. S2)
289 3. Upper 95% CI of SELss across piles (averaged across depths) (Fig. S3)
290 4. SELss at the quietest depths (averaged across piles) (Fig. S4)
291 5. SELss at the loudest depths (averaged across piles) (Fig. S5)

292 For each of these approaches, we considered a single blow energy of 2,000 kJ (the maximum
293 energy reached in each piling event; Equation 1), and all SELss were averaged onto a 5 x 5 km
294 grid. Measurements by Nedwell et al. (2011) of ambient noise in The Wash during construction
295 of Lincs wind farm estimated a median ambient sound level of 118 dB re 1 $\mu Pa^2 \cdot s$, and so any
296 estimated SELss below this value were assigned to 118 dB re 1 $\mu Pa^2 \cdot s$. Following Russell et al.
297 (2016), a parametric bootstrap of the GEE model was used to calculate 95% confidence intervals
298 (CIs) for each zone; these CIs represent the uncertainty resulting from the distribution model (i.e.
299 they do not incorporate any uncertainty in received sound levels).

300 All additional analyses (to that conducted for Hastie et al. (2015) and Russell et al. (2016)) were
301 carried out using R (R Core Team, 2019) within packages maptools (Bivand et al., 2017), raster
302 (Hijmans, 2017), rgdal (Bivand et al., 2014) and sp (Pebesma et al., 2005).

303 **III. RESULTS**

304 **A. Acoustic exposure of the tagged seals**

Comparison of the measured SELss from the recordings of pile driving showed that median absolute error in SELss across all measured piling blows was 4 dB re 1 $\mu Pa^2 \cdot s$ (Fig. S6). In general, errors were higher for the boat-based measurements made close to the surface (median error = 14 dB re 1 $\mu Pa^2 \cdot s$), compared to those from the moored recorder (median error = 3 dB re 1 $\mu Pa^2 \cdot s$).

During the seal tag deployment, the maximum estimated unweighted SELss at individual seals varied from 113 to 173 dB re 1 $\mu Pa^2 \cdot s$. The maximum SELss (173 dB re 1 $\mu Pa^2 \cdot s$) occurred for seal 'pv40-268-12' (Fig. 1) at a range of 4.7 km and a dive depth of 23.6 m. For further details of acoustic exposure of each tagged seal, see Electronic Supporting Information (Fig. S1).

B. Predictions of auditory damage

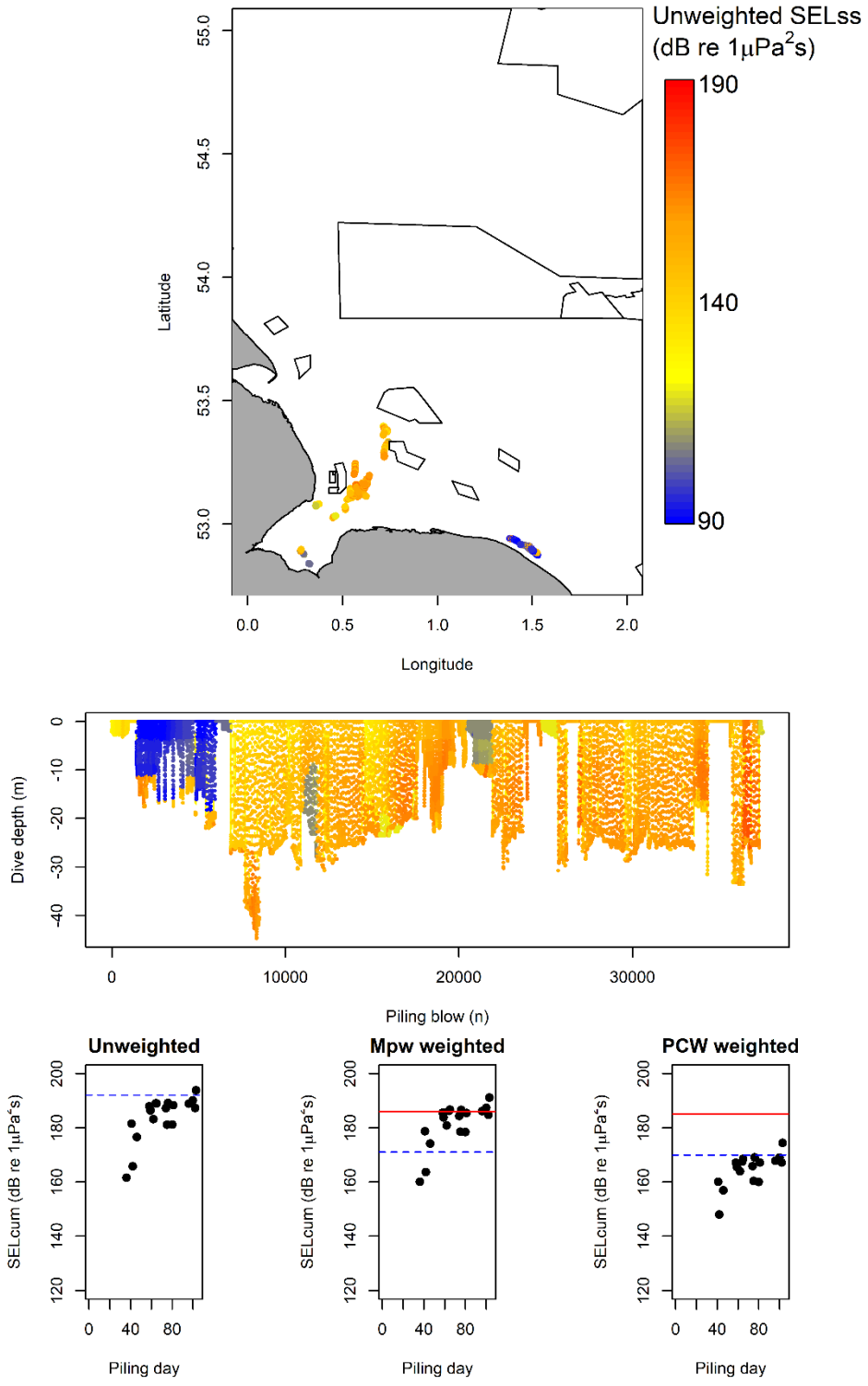
The use of each weighting function resulted in markedly different SELcum estimates from pile driving (Table III). In general, unweighted SELcum were highest (as it is unweighted none of the sound is filtered) and PCW-weighted SELcum (Southall et al., 2019) were lowest.

Predicted unweighted SELcum from pile driving varied between tagged seals (Table III) with maximum SELcum for each seal ranging from 153 to 200 dB re 1 $\mu Pa^2 \cdot s$. It was predicted that five (21%) of the seals did not receive any SELss above the assumed level of effective quiet (124 dB re 1 μPa). Three (13%) of the seals exceeded unweighted sound levels (192 dB re 1 $\mu Pa^2 \cdot s$) previously shown to result in TTS in harbour seals exposed to pile driving sounds (Table II). The closest approach distance to pile driving for each of these three seals was between 3.9 and 5.0 km (Table III).

Predicted M_{pw}-weighted SELcum (Southall et al., 2007) varied between individual seals (Table III) with maximum SELcum (M_{pw}) ranging from 150 to 197 dB re 1 $\mu Pa^2 \cdot s$. Five (21%) of the

seals did not receive any SELss (M_{pw}) above the assumed level of effective quiet (124 dB re 1 μPa). In total, four (17%) of the tagged seals were predicted to receive SELcum (M_{pw}) that exceeded the estimated PTS onset threshold of 186 dB re 1 $\mu Pa^2 \cdot s$ for pinnipeds in water exposed to pulsed sounds, and twelve (50%) were predicted to exceed the TTS onset threshold of 171 dB re 1 $\mu Pa^2 \cdot s$ (Table II). For the individuals estimated to exceed PTS thresholds, closest approach distances ranged from 3.9 to 6.9 km, and for TTS from 3.9 to 17.0 km (Table III).

Predicted maximum PCW-weighted SELcum (Southall et al., 2019) ranged from 134 to 179 dB re 1 $\mu Pa^2 \cdot s$ (Table III). Ten (42%) of the seals did not receive SELss (PCW) above the assumed level of effective quiet (124 dB re 1 μPa). None of the tagged seals were predicted to receive SELcum (PCW) that exceeded the estimated PTS onset threshold (185 dB re 1 $\mu Pa^2 \cdot s$), and four (17%) were predicted to exceed the TTS onset threshold (170 dB re 1 $\mu Pa^2 \cdot s$) for phocids in water exposed to impulsive sounds (Table II). For each of these four seals estimated to exceed TTS thresholds, closest approach distances to piling ranged from 3.9 to 6.9 km (Table III).


Unweighted SELss
(dB re 1μPa²s)
190
140
90
Latitude
55.0
54.5
54.0
53.5
53.0
0.0
0.5
1.0
1.5
2.0
Longitude
Dive depth (m)
0
-10
-20
-30
-40
0
10000
20000
30000
Piling blow (n)
Unweighted
Mpw weighted
PCW weighted
SELcum (dB re 1μPa²s)
120
140
160
180
200
0
40
80
Piling day

341

FIG. 1. (Color online) Example of the estimated acoustic exposure from pile driving at one of the tagged harbour seals (ID#: pv40.268.12). The figure shows the estimated locations of the seal (top panel) and the dive depth (middle panel) of the seal at the times it received the sound from each piling strike. The points in both panels have been colour coded by estimated unweighted single strike Sound Exposure Levels (SELss; dB re 1 $\mu Pa^2 \cdot s$). The lower panels show the estimated cumulative sound exposure levels (SELcum; dB re 1 $\mu Pa^2 \cdot s$) to the tagged seal for each 24 hr period, including the unweighted SELcums, M-weighted (M_{pw}) SELcums (Southall et al., 2007), and PCW-weighted SELcums (Southall et al., 2019). The estimated onset thresholds for TTS (dashed line) and PTS (solid line) are shown for each weighting.

C. Changes in seal density in relation to pile driving

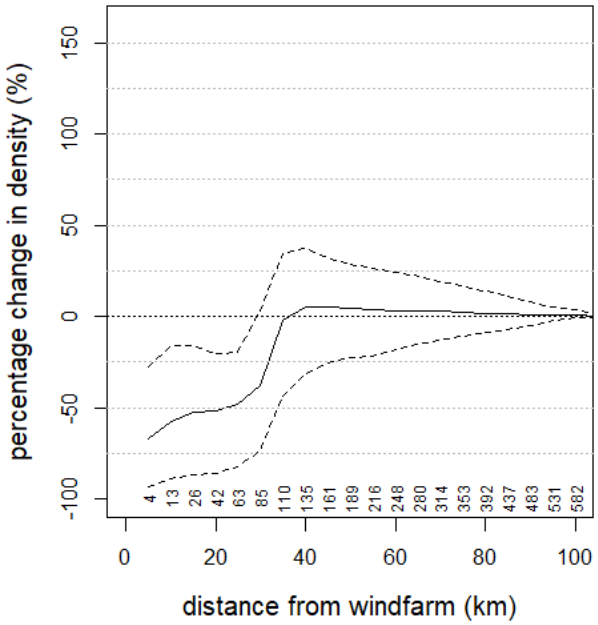
During piling, seal density was predicted to significantly decrease (defined as when the upper CI is a negative percentage change in density) within 25 km of the wind farm site by both cumulative (Fig. 2(a)) and annulus (Fig. 2(b)) approaches. This decrease was detected in all 5 km distance bands (annulus) out to 25 km (Fig. 2(b)). There was no significant change in density detected beyond this distance, considering either cumulative or annulus zones. The predicted change in density (and confidence intervals) of the cumulative approach (Fig. 2(a)) converged towards zero as the largest zone considered encompassed almost the entire study area (all cells within 100 km of the wind farm) and so there would be no overall change in density (percentage of seals).

Seal density was also predicted to decline with increased received sound levels (Fig. 2(c)(d)). Using the first metric (mean SELss across depths and piles), the cumulative approach revealed significant declines when all cells $\geq$ 140 dB re 1 $\mu Pa^2 \cdot s$ are considered (Fig. 2(c)); however, when each received level zone is considered separately (annulus), declines are only detected in each 5

364 dB zone above 145 dB re 1 $\mu Pa^2 \cdot s$ (Fig. 2(d), Table V). There was substantial variation in the
365 predicted SELss (Figs. S2-5(c)) with depth and pile considered, and thus in the resulting
366 percentage change in density-SELss relationship (Fig. S2-5(a)(b)). Due to the variation in these
367 relationships, there was also variation in the SELss threshold above which a significant decline in
368 seal density would be predicted (Table IV). Indeed, considering the lower 95% CI across piles
369 (averaged across depths) revealed no clear relationship with seal density for annulus zones (Fig.
370 S2). In contrast, the upper 95% across piles revealed a significant decrease in density for all
371 annulus zones from 160 dB re 1 $\mu Pa^2 \cdot s$ (Fig. S3). Considering the quietest (Fig. S4) or loudest
372 (Fig. S5) depths separately, there was a significant decrease in density in all annulus zones from
373 145 and 150 dB re 1 $\mu Pa^2 \cdot s$ respectively. In general, the annulus approaches did not predict
374 significant declines in seal density until higher received SELss levels than the cumulative approach
375 (Table IV). For further information on the variation in predicted density between different piling
376 events and water depths, see Electronic Supporting Information.

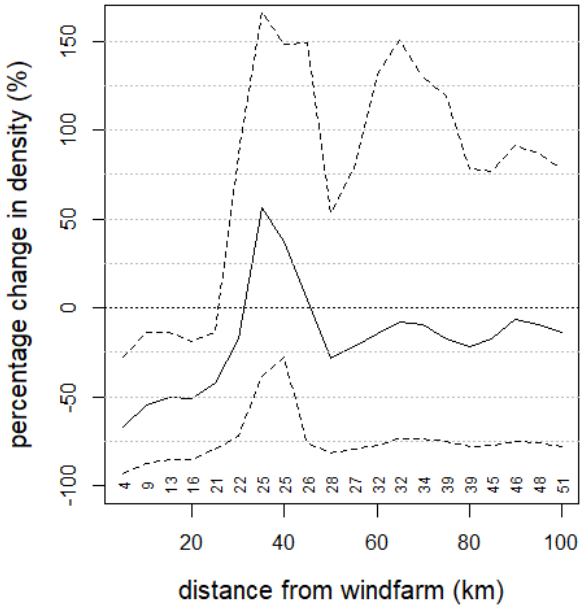
CUMULATIVE

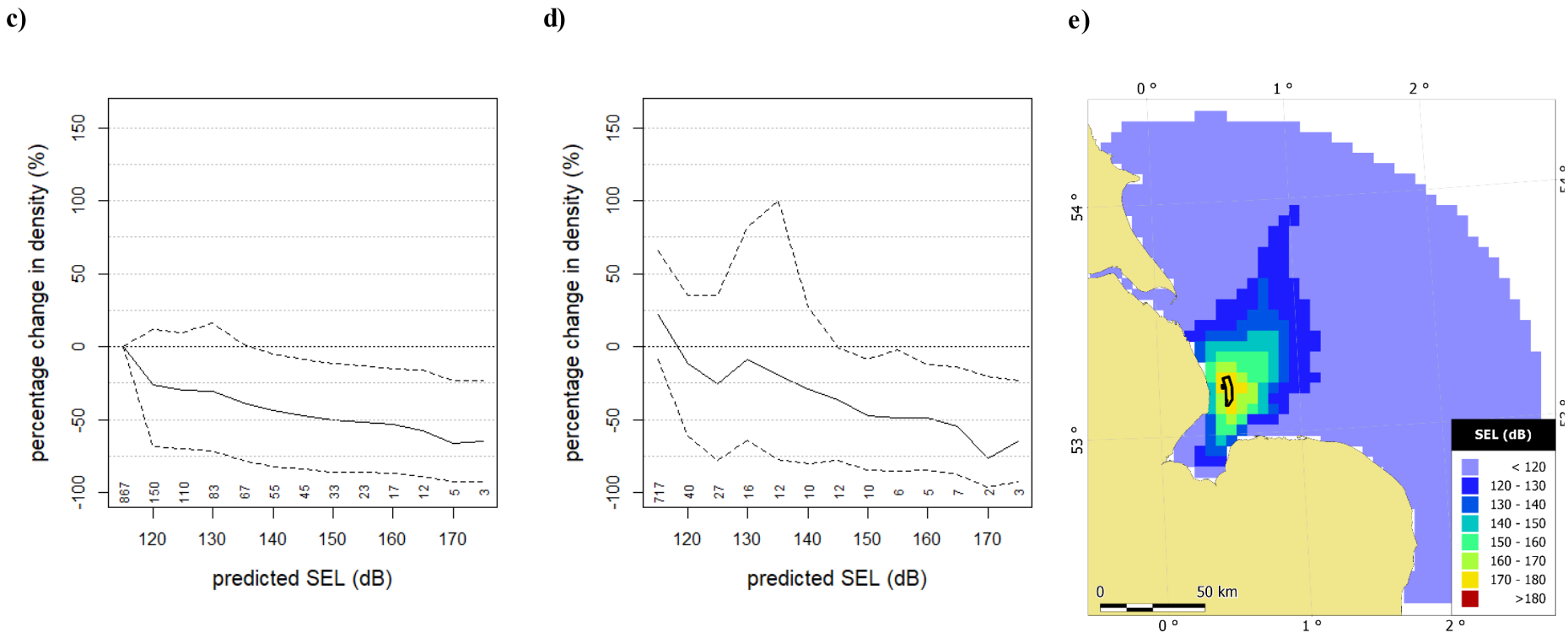
a)



ANNULUS

b)





377 **FIG. 2.** (Color online) Predicted changes in seal density as a function of distance from the centre of the wind farm (a-b) and estimated sound
378 exposure level (c-d) (SEL, dB re 1 $\mu Pa^2 \cdot s$), with SEL averaged across all water depths and piles. (a) Seal density in cumulative zones of
379 increasing distance: plotted density change at distance $\boldsymbol{x}$ is the change in all spatial cells $\leq \boldsymbol{x}$ km (as presented in Russell et al. (2016)). (b) Seal
380 density in annulus 5 km increments: plotted density change at distance $\boldsymbol{x}$ is the change in all spatial cells between $\boldsymbol{x} - \mathbf{5}$ and $\boldsymbol{x}$ km. (c) Seal

density in cumulative zones of received sound level: plotted density change at SEL $\boldsymbol{x}$ is the change in all spatial cells $\geq \boldsymbol{x}$ dB. (d) Seal density in annulus 5 dB increments: plotted density change at SEL $\boldsymbol{x}$ is the change in all spatial cells between $\boldsymbol{x}$ and $\boldsymbol{x} + \mathbf{5}$ dB. Annotations denote the number of spatial grid cells in each distance/SEL category. The dashed lines represent 95% confidence intervals. The corresponding predicted SELs across the study area (averaged across depths and piles) are shown in (e).

IV. DISCUSSION AND CONCLUSIONS

This study used tracking data from 24 harbour seals near a wind farm construction site (Hastie et al., 2015; Russell et al., 2016) to explore four questions in relation to the sensitivity of predicted sound exposures, auditory damage, and changes in seal density to a range of commonly used techniques and assumptions. (1) We found marked differences in the numbers of seals predicted to suffer auditory damage depending upon the choice of weighting functions and thresholds (between 13 – 50%, and 0 – 17% of seals were predicted to exceed TTS and PTS thresholds, respectively). (2) Predictions of seal density during pile driving, as a function of both distance and predicted received levels, differed between the use of cumulative versus annulus zones. We recommend that future studies use annulus zones, and impact assessments use the results from the annulus predictions (Table V). (3 & 4) The relationship between changes in seal density and predicted received level varied markedly depending on how variations in pile installation and water depth were accounted for. These findings have implications for the use of results from such studies (Hastie et al., 2015; Russell et al., 2016) by policy makers and regulators. In particular, we have found that the choice of method can lead to different estimates of effects and therefore different recommendations for future regulation.

The use of each weighting function resulted in marked differences in estimated SELcum on harbour seals. Specifically, unweighted SELcums from pile driving were highest (up to 200 dB re 1 $\mu Pa^2 \cdot s$), M_{pw}-weighted SELcums (Southall et al., 2007) were intermediate (up to 197 dB re 1 $\mu Pa^2 \cdot s$), and PCW-weighted SELcums (Southall et al., 2019) were lowest (up to 179 dB re 1 $\mu Pa^2 \cdot s$). This is to be expected given the differences in the each of the weighting functions. The approach developed by Southall et al. (2007) was designed as relatively conservative initial guidance and the M_{pw} weighting function was therefore flat across the hearing range frequencies of each functional species group. For seals exposed to pile driving

412 sounds in the current study, this resulted in SELcums (M_{pw}) that are only ~1-3 dB lower than
413 unweighted values. More recent guidance uses information from new auditory damage
414 studies to develop a series of updated weighting functions for each functional species group
415 (Southall et al., 2019). This resulted in SELcum (PCW) levels that were ~20-35 dB lower
416 than unweighted values.

417 Correspondingly, the differences in acoustic exposures between the M_{pw} and PCW weighting
418 functions led to variation in the percentage of seals predicted to receive SELcums exceeding
419 published TTS (50 vs. 17 %) and PTS (17 vs. 0 %) thresholds. Using an unweighted
420 threshold, a predicted 13% of individuals exceeded values associated with TTS; no PTS
421 thresholds are available for unweighted pulsed sounds. These results from individual seals
422 exposed to sound broadly reflect the conditions simulated in Faulkner et al. (2019), with the
423 Southall et al. (2019) criteria resulting in markedly lower effects ranges for auditory damage
424 from pile driving sounds.

425 These weighting-function specific percentages are lower than previous predictions of
426 auditory damage from pile driving sound exposure. Specifically, using the same seal tag data
427 to that analysed here, Hastie et al. (2015) predicted maximum 24-hour SELcum (M_{pw}) values
428 ranging from 171 to 195 dB re 1 $\mu Pa^2 \cdot s$ for individual seals; 50% of seals were predicted to
429 exceed the PTS onset threshold (compared to 17% here) and all (100%) exceeded the TTS
430 threshold (compared to 50% here). The difference between these results was due to the
431 different sound propagation approaches used, highlighting the clear sensitivity of predicted
432 acoustic exposure and the associated threshold shifts, to the assumptions of commonly used
433 propagation models.

434 The acoustic modelling approach used here predicts the effects of strike energy and
435 bathymetry, and takes into account more information on the environment and pile driving
436 source (compared to Hastie et al., (2015)). Whilst this reduces uncertainty, there are still

potential sources of variation which are not taken into account. Pile penetration depth can affect the dynamic behaviour of the pile and so could affect the sound produced (de Jong et al., 2019). Here, we assume a homogenous medium sand seabed and a constant water depth at mean sea level. Although the majority of the study area is of this sediment type, variation in this could increase uncertainty in predictions of received level and associated effects on animals, especially for the lower frequencies modelled (< 1 kHz). Assuming a constant water depth is a common approach for acoustic modelling. However, for areas with a strong tidal cycle, it is possible that variation in propagation conditions over the tidal cycle (and associated water depths) could be considerable. Investigation into this variation across tidal cycles would be a useful avenue for future research, although whether it would be computationally feasible to integrate this into individual impact studies is unclear. The uncertainties associated with the Aquarius modelling approach are discussed further in de Jong et al. (2019). Comparisons of the model estimates with a series of measurements from opportunistic boat-based hydrophones and a moored recorder suggests that the error in model estimates is approximately 4 dB re 1 $\mu Pa^2 \cdot s$. The boat-based recordings made near the water surface (~1 m) all measured lower SELss than the model predictions for the shallowest depth bin (2.5 m). Whilst not a formal validation, this comparison highlights the potential uncertainty of received levels near the surface, and the performance of the model for estimating near surface piling noise (although the comparison is only made above the modelled depths). Received levels near the surface are highly variable due to interference patterns, sound speed profile ducts and waves, and measurements are likely to be sensitive to environmental conditions such as wind and wave activity. The measurements from the moored recorder at ~9 m below the surface provided a close match to the model predictions. The conditions at these depths are more representative of the majority of the water column, as variability in propagation conditions is much less.

462 Here we extended the potential utility of the results from Russell et al. (2016) by presenting
463 changes in seal density as a function of annulus zones of distance and five metrics of
464 predicted received level. Using annulus distance zones confirmed significant decreased
465 density up to 25 km from the centre of the windfarm (as found using cumulative distance
466 zones; Russell et al. (2016)). To compare overall seal distribution between piling and non-
467 piling (a binary comparison), it was necessary to generate one received level per cell (across
468 all 27 pile installations and water depths). Russell et al. (2016), using cumulative zones,
469 predicted a significant decrease in seal density from received levels (averaged across all
470 installations) above 140 – 155 dB re 1 $\mu Pa^2 \cdot s$, based on the quietest and loudest part of the
471 water column. Here, we show these levels are affected both by the sound propagation model
472 used (130 – 140 dB re 1 $\mu Pa^2 \cdot s$ for quietest-loudest depths, cumulative), and the use of
473 annulus rather than cumulative zones (145 – 150 dB re 1 $\mu Pa^2 \cdot s$ for quietest-loudest depths,
474 annulus) (Table IV). Examining the variation in SELss across pile installations (95 % CIs)
475 revealed substantial variation in the level of significant decrease in density; indeed only for
476 the upper 95% CI could a significance level be quantified (Table IV). Annulus zones
477 (especially at larger distances/lower received levels) show wider confidence intervals for
478 changes in density than for cumulative zones. This is due to the increasing sample size
479 associated with the increasing size of cumulative zone; the change will converge on zero
480 change in percentage density as the cumulative zones encompass an increasing proportion of
481 the study area. Additionally, the received levels at which there is a significant predicted effect
482 on seal density are lower for the cumulative approach. The cumulative method always
483 includes the zones of highest exposure (and potential effect on behaviour) and so this likely
484 enables the overall density change to be detected further from the wind farm.

485 Here, we illustrated how the predicted SELss associated with significant decreases in seal
486 density varies across pile installations and depth. However, there are other sources of

487 variation that we did not account for. For instance, we only considered the average maximum
488 piling energy reached over all piling events (2,000 kJ) and not the received levels from each
489 piling strike with potentially different sequences of piling energies. There may also be
490 changes in the seafloor between piling sites and potentially equipment changing the source
491 spectrum of different piling strikes. Linking population level responses to a particular sound
492 level necessitates averaging over a wide range of possible situations, including different
493 external conditions (e.g. piling ramp-up sequence, time of day), and differences between and
494 within individual animals (e.g. behavioural state, previous exposure history). These
495 differences may increase variability in predicted responses. Considering only the average
496 (population-level) response makes it challenging to identify factors which might make
497 animals more or less responsive to sound, information which could be used in future
498 assessments of noise impacts.

499 In summary, we use tracking data on wild harbour seals exposed to pile driving sounds to
500 update quantitative estimates of effects on seal hearing and behaviour. The findings of Hastie
501 et al. (2015) and Russell et al. (2016) remain amongst the few studies quantifying the effects
502 of pile driving on seals; as such, they are widely used in EIAs. While we recognise the
503 contribution these findings make, it is important that researchers, regulators, policy makers,
504 and industry recognise the inherent limitations associated with studies predicting auditory
505 damage and population level redistribution. Auditory damage in marine mammals is a rapidly
506 evolving field of research (Kastelein et al., 2018; Southall et al., 2019), and this current study
507 demonstrates the importance of updating the predictions as new information becomes
508 available. It also illustrates the sensitivity and limitations of predictions made with commonly
509 used acoustic propagation models. We recommend future studies, where possible, carry out a
510 spatially diverse set of acoustic measurements to calibrate and hence reduce the uncertainties
511 associated with the acoustic source and propagation modelling. These acoustic measurements
512 should be used to monitor noise levels during construction and help characterise the variation

513 in sound produced from different strike energies. Efforts should be made to validate sound
514 propagation models in the environment and conditions they are proposed to be used in, for
515 both impact assessments and scientific studies. In particular, these measurements should
516 focus on the expected location and conditions (depth, habitat) of the study population.
517 Underwater noise monitoring is often a requirement of consent for offshore wind farm
518 projects and as such, should enable model verification across a large range of environments
519 and pile types. Researchers should also endeavour to publish updated predictions of auditory
520 damage following Southall et al. (2019). A clear avenue for future work would be to validate
521 these types of predictions through the collection of auditory threshold information pre- and
522 post- exposure to pile driving; this could potentially be carried out on wild seals using
523 auditory evoked potential measurements (Wolski et al., 2013) or in a captive environment
524 using controlled exposures and psychophysical methods (e.g. Kastak et al. (2005); Kastelein
525 et al. (2012)).

526 Population-level redistribution studies are a key first step in determining the presence and
527 magnitude of potential effects, and the time to recovery (to pre-disturbance distribution).
528 Researchers should make their findings as applicable as possible for use by stakeholders (e.g.
529 using annulus rather than cumulative zones in quantitative EIA analyses). In particular,
530 relating changes in density to distance from a source can improve understanding of the
531 potential implications of avoidance (in terms of collision risk (e.g. tidal turbines), barrier
532 effects and loss of habitats or resources). However, there are a number of important caveats
533 associated with population level redistribution studies. For example, it is not clear whether
534 these changes in density are a result of more animals leaving the area, less new animals
535 entering the area, or a combination of both. Such studies necessarily combine multiple
536 potential disturbance events and animal responses, and here we showed that these also
537 encompass a wide range of potential received levels.

With developments of tracking technology and on-animal long-term sound recordings (e.g. Mikkelsen et al. (2019)), information on individual behaviour and sound exposure is rapidly improving. Analytical tools (e.g. DeRuiter et al. (2013); Quick et al. (2017)) to model such data mean that studies of responses to sound are no longer restricted to considering population-level distribution patterns. A useful avenue for future research would be to investigate how individual seals respond to sound exposure. Studying behaviour of individuals may provide greater insight into the mechanisms behind the population-level patterns seen and enable us to quantify dose-response relationships taking into account the variability between individuals. This will ultimately improve efforts to extrapolate and model effects at the population level.

ACKNOWLEDGEMENTS

This paper was part of work presented at the fifth International Meeting on The Effects of Noise on Aquatic Life held in Den Haag (The Netherlands), July 2019, and we are grateful for the Rodney Coates award given to KFW for the presentation of this project at the conference. We also thank the editor and two anonymous reviewers whose insightful comments greatly improved the paper. Data collection was funded as part of the Department of Energy and Climate Change's (now Department of Business, Energy and Industrial Strategy) Offshore Energy Strategic Environmental Assessment programme, with additional resources from National Capability funding from the Natural Environment Research Council to the Sea Mammal Research Unit (grant no. SMRU1001). Sound propagation modelling and subsequent analyses were funded by Race Bank Offshore Wind Farm Ltd. KFW was funded by University of St Andrews and the Department of Business, Energy and Industrial Strategy, as part of a PhD studentship.

ENDNOTES

---

[1] See supplementary material at [URL will be inserted by AIP] for further plots of estimated sound exposure, seal density, acoustic measurements, and details of tagged individuals.

REFERENCES

Ainslie, M. A. (**2010**). *Principles of sonar performance modelling*, Springer-Praxis, Chichester, UK.

Bailey, H., Brookes, K. L., and Thompson, P. M. (**2014**). "Assessing environmental impacts of offshore wind farms: Lessons learned and recommendations for the future," Aquat. Biosyst., **10**, 1–13. doi:10.1186/2046-9063-10-8

Bailey, H., Senior, B., Simmons, D., Rusin, J., Picken, G., and Thompson, P. M. (**2010**). "Assessing underwater noise levels during pile-driving at an offshore windfarm and its potential effects on marine mammals.," Mar. Pollut. Bull., **60**, 888–897. doi:10.1016/j.marpolbul.2010.01.003

Bivand, R., Keitt, T., Rowlingson, B., and Pebesma, E. (**2014**). *rgdal: Bindings for the geospatial data abstraction library*, R package version 0.8-16.

Bivand, R., and Lewin-Koh, N. (**2017**). *maptools: Tools for reading and handling spatial objects*, R package version 0.9-2. Retrieved from https://cran.r-project.org/package=maptools

Breton, S. P., and Moe, G. (**2009**). "Status, plans and technologies for offshore wind turbines in Europe and North America," Renew. Energy, **34**, 646–654. doi:10.1016/j.renene.2008.05.040

587 Deeks, A. J., and Randolph, M. F. (**1993**). "Analytical modelling of hammer impact for pile
588 driving," Int. J. Numer. Anal. Methods Geomech., **17**, 279–302.

589 DeRuiter, S. L., Southall, B. L., Calambokidis, J., Zimmer, W. M. X., Sadykova, D., Falcone,
590 E. A., Friedlaender, A. S., Joseph, J. E., Moretti, D., Schorr, G. S., Thomas, L., and Tyack,
591 P. L. (**2013**). "First direct measurements of behavioural responses by Cuvier's beaked
592 whales to mid-frequency active sonar," Biol. Lett., **9**, 20130223–20130223.
593 doi:10.1098/rsbl.2013.0223

594 Faulkner, R. C., Farcas, A., and Merchant, N. D. (**2019**). *Risk assessment of permanent auditory*
595 *injury in marine mammals : Differences arising from the application of the Southall and*
596 *NOAA criteria*, Scottish Marine and Freshwater Science Report Vol 10 No 1. Published
597 by Marine Scotland Science. doi:10.7489/12205-1

598 Finneran, J. J. (**2015**). "Noise-induced hearing loss in marine mammals: A review of temporary
599 threshold shift studies from 1996 to 2015," J. Acoust. Soc. Am., **138**, 1702–1726.

600 Hastie, G. D., Russell, D. J. F., Mcconnell, B., Moss, S., Thompson, D., and Janik, V. M.
601 (**2015**). "Sound exposure in harbour seals during the installation of an offshore wind farm:
602 Predictions of auditory damage," J. Appl. Ecol., **52**, 631–640. doi:10.1111/1365-
603 2664.12403

604 Hijmans, R. J. (**2017**). *raster: Geographic data analysis and modeling*, R package version 2.6-
605 7. Retrieved from https://cran.r-project.org/package=raster

606 Iliffe, J. C., Ziebart, M. K., Turner, J. F., Talbot, A. J., and Lessnoff, A. P. (**2013**). "Accuracy
607 of vertical datum surfaces in coastal and offshore zones," Surv. Rev., **45**, 254–262.

608 de Jong, C., Binnerts, B., Prior, M., Colin, M., Ainslie, M., Mulder, I., and Hartstra, I. (**2019**).
609 "Wozep – WP2: update of the Aquarius models for marine pile driving sound
610 predictions," TNO Rep. (2018), number R11671, The Hague, Netherlands, p. 94.

611 Retrieved from
612 https://www.noordzeeloket.nl/publish/pages/160801/update_aquarius_models_pile_drivi
613 ng_sound_predeictions_tno_2019.pdf

614 Kastak, D., Southall, B. L., Schusterman, R. J., and Kastak, C. R. (**2005**). "Underwater
615 temporary threshold shift in pinnipeds: Effects of noise level and duration," J. Acoust.
616 Soc. Am., **118**, 3154–3163. doi:10.1121/1.2047128

617 Kastelein, R. A., Gransier, R., Hoek, L., Macleod, A., and Terhune, J. M. (**2012**). "Hearing
618 threshold shifts and recovery in harbor seals (Phoca vitulina) after octave-band noise
619 exposure at 4 kHz," J. Acoust. Soc. Am., **132**, 3525–3537. doi:10.1121/1.4757641

620 Kastelein, R. A., Helder-Hoek, L., Kommeren, A., Covi, J., and Gransier, R. (**2018**). "Effect
621 of pile-driving sounds on harbor seal (*Phoca vitulina*) hearing," J. Acoust. Soc. Am., **143**,
622 3583–3594. doi:10.1121/1.5040493

623 Mikkelsen, L., Johnson, M., Wisniewska, D. M., van Neer, A., Siebert, U., Madsen, P. T., and
624 Teilmann, J. (**2019**). "Long-term sound and movement recording tags to study natural
625 behavior and reaction to ship noise of seals," Ecol. Evol., **9**, 2588–2601.
626 doi:10.1002/ece3.4923

627 Nedwell, J. R., Brooker, A. G., and Barham, R. J. (**2011**). *Measurement and assessment of*
628 *underwater noise during impact piling operations at the Lincs offshore wind farm*,
629 Subacoustech Environmental Report No. E273R0203.

630 NMFS (**2016**). *Technical guidance for assessing the effects of anthropogenic sound on marine*
631 *mammal hearing: Underwater acoustic thresholds for onset of permanent and temporary*
632 *threshold shifts*, U.S. Dept. of Commer., NOAA. NOAA. Technical Memorandum
633 NMFS-OPR-55, 178 pages.

634 NMFS (**2018**). *2018 Revisions to: Technical guidance for assessing the effects of*

635 *anthropogenic sound on marine mammal hearing (Version 2.0): Underwater thresholds*
636 *for onset of permanent and temporary threshold shifts*, U.S. Dept. of Commer., NOAA.
637 NOAA Technical Memorandum NMFS-OPR-59, 167 pages.

638 Pebesma, E., and Bivand, R. S. (**2005**). *S classes and methods for spatial data: the sp package*,
639 R package version 2.01-40.

640 Porter, M. B. (**2001**). *The KRAKEN normal mode program*, SACLANT Undersea Research
641 Centre, 207 pages. Retrieved from http://oalib.hlsresearch.com/Modes/kraken.pdf

642 Quick, N., Scott-Hayward, L., Sadykova, D., Nowacek, D., and Read, A. (**2017**). "Effects of a
643 scientific echo sounder on the behavior of short-finned pilot whales (*Globicephala*
644 *macrorhynchus*)," Can. J. Fish. Aquat. Sci., **74**, 716–726. doi:10.1139/cjfas-2016-0293

645 R Core Team (**2019**). *R: A language and environment for statistical computing*, Vienna,
646 Austria. Retrieved from https://www.r-project.org

647 Russell, D. J. F., Brasseur, S. M. J. M., Thompson, D., Hastie, G. D., Janik, V. M., Aarts, G.,
648 McClintock, B. T., Matthiopoulos, J., Moss, S. E. W., and McConnell, B. (**2014**). "Marine
649 mammals trace anthropogenic structures at sea," Curr. Biol., **24**, R638–R639.
650 doi:10.1016/j.cub.2014.06.033

651 Russell, D. J. F., Hastie, G. D., Thompson, D., Janik, V. M., Hammond, P. S., Scott-Hayward,
652 L. A. S., Matthiopoulos, J., Jones, E. L., and McConnell, B. J. (**2016**). "Avoidance of wind
653 farms by harbour seals is limited to pile driving activities," J. Appl. Ecol., **53**, 1642–1652.
654 doi:10.1111/1365-2664.12678

655 Sehgal, A., Tumar, I., and Schönwälder, J. (**2009**). *Variability of available capacity due to the*
656 *effects of depth and temperature in the underwater acoustic communication channel* IEEE
657 Ocean. 2009- Eur., 1–6 pages. doi:10.1109/OCEANSE.2009.5278268

658 Sharples, R. J., Moss, S. E., Patterson, T. A., and Hammond, P. S. (**2012**). "Spatial variation in

659 foraging behaviour of a marine top predator (*Phoca vitulina*) determined by a large-scale
660 satellite tagging program," PLoS One, **7**, e37216. doi:10.1371/journal.pone.0037216

661 Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Jr., C. R. G.,
662 Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E., Richardson, W., Thomas, J., and
663 Tyack, P. L. (**2007**). "Marine Mammal Noise Exposure Criteria: Initial Scientific
664 Recommendations," Aquat. Mamm., **33**, 411–521. doi:10.1578/AM.33.4.2007.411

665 Southall, B. L., Finneran, J. J., Reichmuth, C., Nachtigall, P. E., Ketten, D. R., Bowles, A. E.,
666 Ellison, W. T., Nowacek, D. P., and Tyack, P. L. (**2019**). "Marine mammal noise exposure
667 criteria: Updated scientific recommendations for residual hearing effects," Aquat.
668 Mamm., **45**, 125–232. doi:10.1578/AM.45.2.2019.125

669 Thompson, P. M., Hastie, G. D., Nedwell, J., Barham, R., Brookes, K. L., Cordes, L. S., Bailey,
670 H., and McLean, N. (**2013**). "Framework for assessing impacts of pile-driving noise from
671 offshore wind farm construction on a harbour seal population," Environ. Impact Assess.
672 Rev., **43**, 73–85. doi:10.1016/j.eiar.2013.06.005

673 WindEurope (**2019**). *Offshore wind in Europe: Key trends and statistics 2018*, Offshore wind
674 in Europe: Key trends and statistics 2018. Brussels, Belgium, 1–37 pages.
675 doi:10.1016/S1471-0846(02)80021-X

676 Wolski, L. F., Anderson, R. C., Bowles, A. E., and Yochem, P. K. (**2013**). "Measuring hearing
677 in the harbor seal (Phoca vitulina): Comparison of behavioral and auditory brainstem
678 response techniques," J. Acoust. Soc. Am., **113**, 629–637.

679

680

681 **TABLES**

682 **TABLE I.** List of parameters used for the percussive pile driving source modelling.

Parameter	Value
Pile diameter	5.2 metres
Wall thickness	58.35 mm* *estimated using API equation: D (diameter) = 5,200 mm t (thickness) = 6.35 + D/100 = 58.35 mm
Pile material properties	Material: Steel Density ρ: 7,850 kg/m^3 Elasticity E: 210 GPa Compressional sound speed c_p: 5,172 m/s Poisson ratio ν: 0.3
Range of strike energies	54 to 2,035 kJ
Hammer type	MHU 1900S
Ram mass	95 ton
Anvil mass	31 ton
Contact stiffness	20 GPa
Frequency range modelled	16 Hz to 20 kHz

683

684

685

686

687

688

689

690

691

692

693

694

695 **TABLE II.** List of thresholds used to estimate auditory damage in harbour seals exposed to
696 pile driving sounds. Shown are the cumulative sound exposure levels (SELcum, dB re 1
697 µPa2·s) estimated to cause temporary (TTS) or permanent (PTS) threshold shifts in hearing,
698 using three different methods of weighting sound frequencies. The M_{pw} and PCW weightings
699 are based on exposure during a 24-hour period, whereas the unweighted threshold is based on
700 observed TTS in a 6-hour experimental setting.

701

Frequency weighting method	**TTS threshold**	**PTS threshold**	**Reference**
Unweighted	192	-	Kastelein et al. (2018)
M_{pw} weighted (M-weighted, Pinnipeds in Water)	171	186	Southall et al. (2007)
PCW weighted (Phocids in Water)	170	185	Southall et al. (2019)

702

703

704

705

706

707

708

709

710

711

712

713

TABLE III. Summary of the closest distance to pile driving (km) and the maximum estimated 24-hour cumulative sound exposure level (SELcum; dB re 1 $\mu Pa^2 \cdot s$) for each tagged seal, including the unweighted SELcum, M_{pw} weighted SELcum (Southall et al., 2007), and PCW weighted SELcum (Southall et al., 2019). The asterisk (*) highlights SELcums exceeding onset thresholds for TTS and double-asterisk (**) for those exceeding onset thresholds for PTS (please note there are no PTS thresholds for the unweighted SELcums).

Seal reference number	Closest distance to piling (km)	Unweighted	M_{pw} weighted	PCW weighted
pv40-268-12	3.9	194 *	191 **	174 *
pv40-270-12	40.4	-	-	-
pv42-162-12	9.3	184	182 *	165
pv42-165-12	6.9	191	189 **	170 *
pv42-194-12	26.9	172	170	-
pv42-198-12	29.9	-	-	-
pv42-220-12	34.2	-	-	-
pv42-221-12	25.3	166	163	134
pv42-266-12	24.9	154	152	-
pv42-277-12	4.7	200 *	197 **	179 *
pv42-287-12	38.7	-	-	-
pv42-288-12	15.7	170	169	148
pv42-289-12	27.5	-	-	-
pv42-290-12	16.9	176	174 *	155
pv42-291-12	14.0	177	175 *	158
pv42-292-12	34.8	153	150	-
pv42-293-12	17.0	176	174 *	156
pv42-294-12	30.7	159	157	-
pv42-295-12	11.3	187	185 *	167
pv42-316-12	5.8	186	184 *	165
pv42-317-12	17.0	185	183 *	164
pv42-318-12	13.8	184	182 *	164
pv42-319-12	21.7	166	164	-
pv42-320-12	5.0	194 *	192 **	176 *

724 **TABLE IV.** Summary of estimated single strike sound exposure levels (SELss, dB re 1
725 $\mu Pa^2 \cdot s$) of pile driving, above which a significant decline in seal density is predicted. Five
726 approaches to combining SELss across piling events and depths are shown, alongside
727 previously published results. Two approaches of summarising corresponding seal density
728 estimates over space are calculated: annulus or cumulative zones (both in 5 dB increments).

729
730

	Approach	
	Annulus	**Cumulative**
Mean (averaged across depths and piles)		
(1) Mean	145	140
Comparison across piles (averaged across depths)		
(2) Lower 95% CI of piles	No clear relationship	145
(3) Upper 95% CI of piles	160	150
Comparison across depths (averaged across piles)		
(4) Quietest depth	145	130
(5) Loudest depth	150	140
Russell et al. (2016) (averaged across piles)		
Quietest depth	-	140
Loudest depth	-	155

731
732

733 **TABLE V.** Predictions of seal density (and changes in seal density) during piling and breaks in piling. Seal densities are presented for each predicted
734 sound exposure level (SELss) category (annulus), along with the number of spatial grid cells corresponding to each SELss category. SELss were
735 averaged across all water depths and piling events. Values in bold denote significant changes (confidence intervals not containing 0 % change in
736 density).

737

SELss (dB re 1 μPa²·s)	**Number of spatial cells**	**Mean density (% of at-sea population)**			**Percentage change in density**			
		Non-piling	Piling	Difference	Mean	Median	Lower 95% CI	Upper 95% CI
115 – 120	717	53.91	65.94	12.03	22.31	20.26	-8.96	65.95
120 – 125	40	8.77	7.79	-0.98	-11.21	-12.54	-61.26	35.14
125 – 130	27	5.53	4.11	-1.42	-25.60	-29.44	-78.16	35.27
130 – 135	16	8.82	8.08	-0.74	-8.43	-13.08	-64.36	82.19
135 – 140	12	4.62	3.71	-0.91	-19.65	-22.19	-77.47	100.36
140 – 145	10	3.83	2.70	-1.13	-29.40	-36.17	-80.43	27.10
145 – 150	**12**	**3.28**	**2.09**	**-1.19**	**-36.37**	**-40.52**	**-77.97**	**-0.34**
150 – 155	**10**	**3.59**	**1.89**	**-1.70**	**-47.31**	**-51.46**	**-84.70**	**-8.56**
155 – 160	**6**	**2.05**	**1.05**	**-1.00**	**-48.72**	**-52.46**	**-85.48**	**-1.90**
160 – 165	**5**	**2.80**	**1.44**	**-1.36**	**-48.52**	**-54.35**	**-84.63**	**-12.20**
165 – 170	**7**	**2.10**	**0.96**	**-1.14**	**-54.38**	**-58.67**	**-87.73**	**-13.64**
170 – 175	**2**	**0.08**	**0.02**	**-0.06**	**-76.26**	**-79.27**	**-96.04**	**-20.32**
175 – 180	**3**	**0.62**	**0.22**	**-0.40**	**-64.80**	**-68.41**	**-92.17**	**-22.93**

738

739