*The Annals of Applied Statistics*
2023, Vol. 17, No. 4, 3570–3588
https://doi.org/10.1214/23-AOAS1776


# CONTINUOUS-TIME MODELLING OF BEHAVIOURAL RESPONSES IN ANIMAL MOVEMENT

BY THÉO MICHELOT[1,a], RICHARD GLENNIE[2,b], LEN THOMAS[2,c], NICOLA QUICK[3,e] AND CATRIONA M. HARRIS[2,d]

[1]*Department of Mathematics and Statistics, Dalhousie University,* [a]*theo.michelot@dal.ca*
[2]*Centre for Research into Ecological and Environmental Modelling, University of St Andrews,* [b]*research@glennies.co.uk,* [c]*len.thomas@st-andrews.ac.uk,* [d]*catriona.harris@st-andrews.ac.uk*
[3]*Nicholas School of the Environment, Duke University,* [e]*nicola.quick@plymouth.ac.uk*

There is great interest in ecology to understand how wild animals are affected by anthropogenic disturbances, such as sounds. For example, behavioural response studies are an important approach to quantify the impact of naval activity on marine mammals. Controlled exposure experiments are undertaken where the behaviour of animals is quantified before, during, and after exposure to a controlled sound source, often using telemetry tags (e.g., accelerometers or satellite trackers). Statistical modelling is required to formally compare patterns before and after exposure, to quantify deviations from baseline behaviour. We propose varying-coefficient stochastic differential equations (SDEs) as a flexible framework to model such data with two components: (1) time-varying baseline dynamics, modelled with nonparametric or random effects of time-varying covariates, and (2) a nonparametric response model, which captures deviations from baseline. SDEs are specified in continuous time, which makes it straightforward to analyse data collected at irregular time intervals, a common situation for animal tracking studies. We describe how the model can be embedded into a state-space modelling framework to account for measurement error. We present inferential methods for model fitting, model checking, and uncertainty quantification (including on the response model). We apply this approach to two behavioural response study data sets on beaked whales: a satellite track and high-resolution depth data. Our results suggest that the whales' horizontal movement and vertical diving behaviour changed after exposure to the sound source, and future work should evaluate the severity and possible consequences of these responses. These two very different examples showcase the versatility of varying-coefficient SDEs to measure changes in behaviour, and we discuss implications of disturbances for the whales' energetic balance.

**1. Introduction.** There has been a lot of effort in conservation biology to understand how human activity affects wildlife. One particular focus has been to investigate the effect of ship sonars and other anthropogenic sounds on marine mammals (Tyack et al. (2011), Southall et al. (2019)). Controlled exposure experiments (CEEs) consist of monitoring the movement or behaviour of animals, typically using telemetry tags, before and after sound exposure, to determine if individuals respond behaviourally to the stimuli. A continuously increasing quantity of tag data now exists to address this issue from multiple ongoing studies. This has created a need for adequate statistical methods to describe baseline behaviour and, crucially, quantify deviations from it following disturbance. A particular focus of these studies has been on identifying energetically-costly behaviours, such as sudden avoidance or interruption of foraging, as effects from these changes can accumulate to decrease animals'

Received December 2022; revised May 2023.
*Key words and phrases.* Stochastic differential equation, diffusion process, behavioural response study, beaked whale.

survival and reproductive rates. These studies have used various types of tags, with different observed variables and sampling designs, and model formulations have, therefore, often depended on the data type and the goal of inference.

In marine mammal studies, the most common approach to identify interruptions of foraging behaviour has been to summarise high-frequency data at the scale of dives and compare baseline dives to exposed dives, for example, in terms of dive duration, maximum depth, or average acceleration (e.g., using DTags; Johnson and Tyack (2003)). Baseline and exposed dives can, for example, be compared using the Mahalanobis distance, calculated using selected multivariate data streams, to identify unusual behaviour (DeRuiter et al. (2013)). An alternative has been to use these dive summary variables to identify latent behavioural states of animals in a hidden Markov model (DeRuiter et al. (2017)). In that context the effect of disturbance on the probabilities of switching between the behavioural states can be estimated to quantify the response. These approaches have usually required summarising data to the dive level and have, therefore, not focused on small-scale changes in an animal's behaviour during an exposed dive. In cases where within-dive movement was analysed, the aim was to quantify whether and when a change had occurred rather than provide a mechanistic description of the impact of disturbance on the animals' movement activity (e.g., Stimpert et al. (2014)).

In some studies, animals are equipped with satellite tags that record two-dimensional locations, to detect horizontal movement away from the source of disturbance (Cioffi et al. (2022)). Due to satellite transmission limitations, these data typically have high measurement error and irregular intervals corresponding to times the animal came to the surface. Statistical analysis of such data is challenging, and visual assessment is typically used to measure avoidance. Recently, continuous-time discrete-space models have been proposed to analyse such noisy irregular trajectories, which require modelling animal movement on a discrete spatial grid (Jones-Todd et al. (2022)).

We propose varying-coefficient stochastic differential equations (SDEs) as a versatile method to estimate behavioural responses from different types of CEE tag data. Multiple SDE formulations have been proposed for the analysis of animal movement data, including Brownian motion (Pozdnyakov et al. (2014)), Ornstein–Uhlenbeck processes (Dunn and Gipson (1977)), and the integrated Ornstein–Uhlenbeck process (Johnson et al. (2008)). In those models the animal's movement dynamics are specified in terms of a few parameters, for example, representing mean speed or autocorrelation. The varying-coefficient approach we propose here provides great flexibility to express these parameters as functions of time-varying covariates (Michelot et al. (2021)). We demonstrate the utility of this approach for behavioural response studies, based on several extensions to the approach of Michelot et al. (2021): (1) estimation of deviations from baseline model using difference smooths, (2) uncertainty quantification using simultaneous intervals, (3) measurement error using state-space models, and (4) model checking using posterior predictive checks. We illustrate the utility of these models with two common types of CEE data: high-resolution data on diving behaviour from archival tags and low-resolution position data from satellite tags. This statistical framework is widely applicable beyond these examples.

**2. Beaked whale movement data.** Beaked whales have been the focus of multiple behavioural response studies (BRS) due to their apparent vulnerability to the effects of military sonar systems (DeRuiter et al. (2013), Southall et al. (2016), Tyack et al. (2011)). For this purpose, CEEs have been conducted with different types of animal-borne tags, including movement and acoustic sensors at different resolutions, to detect individual behavioural response (Southall et al. (2016)). In this paper we focus on CEEs for Cuvier's beaked whales (*Ziphius cavirostris*) and analyse two types of data with different variables and resolutions:

coarse two-dimensional location data from a satellite tag and fine-resolution depth data from DTags. Plots of the data are shown in Appendix A of the Supplementary Material (Michelot et al. (2023b)).

2.1. *Satellite tag data*. The satellite tag analysed here was deployed as part of the Atlantic BRS, a study on the effects of midfrequency active sonar on deep diving whales. The tag was a SPLASH10-292, Argos satellite-linked location-depth tag (produced by Wildlife Computers, Redmond, Washington) remotely deployed using a DAN-INJECT JM 25 pneumatic projector (DanWild LLC, Austin, Texas) in the LIMPET configuration (Andrews, Pitman and Ballance (2008)) from a nine m rigid-hulled aluminium boat.

The tag was deployed on an adult male Cuvier's beaked whale off Cape Hatteras, North Carolina, on 24 May 2018, and it transmitted for 38 days. Location estimates were derived from Service Argos receivers on polar-orbiting satellites and were assigned an accuracy class based on the timing and number of transmissions received during a satellite pass (see Foley et al. (2021), for details). Only the higher accuracy positions were used. Reliable locations can only be recorded when the whale is at the surface and when satellites are available, which severely limits data collection. Specifically, the latitude of the study site provides only 9% temporal satellite coverage (Cioffi et al. (2022)), and the whales spend most of their time deep underwater, with average 2.2 minute surface ventilation periods (Shearer et al. (2019)). As a result, the locations were sparse in time (average of two locations per day) and included measurement error. The measurement error was available in the form of error ellipses, each corresponding to the $\sqrt{2}$-sigma contour of a bivariate normal distribution (McClintock et al. (2015)).

To increase data resolution, an Argos goniometer (Woods Hole Group Inc., Bourne, MA, U.S.A.) was deployed from the research vessel to collect further data from the tagged whale's transmitter (Cioffi et al. (2022)). Locations from the goniometer had high spatiotemporal resolution, but they only covered short time periods when the vessel was within range of the whale. We added these data to the satellite trajectory to increase the information available for this analysis. We assumed that the goniometer locations had isotropic error ellipses (as defined above) with radius that depended on the strength of the signal received (a proxy for distance between the vessel and the whale). Specifically, we set the radius to 100 m when the signal was stronger than $-50$ dB, 500 m between $-51$ and $-70$ dB, 1 km between $-71$ and $-80$ dB, two km between $-81$ and 90 dB, and 10 km for signals weaker than $-91$ dB.

On 3 June 3 2018, at 16:00:04 UTC, the whale was exposed to an hour-long CEE of midfrequency active sonar, similar to the tactical sonars used by the U.S. and other navies (Southall et al. (2016)). Data visualisation suggests that the whale moved away from the sonar source (Southall et al. (2020)), but this has not been confirmed by statistical analysis.

2.2. *DTag data*. DTAGs are multisensor archival tags that are attached to animals via suction cups for up to tens of hours and record various acoustic and movement variables, including depth at 50 Hz resolution (Johnson and Tyack (2003)). Our analysis included data from two separate studies: the SOCAL BRS (four tags; for full tag details, see DeRuiter et al. (2013), Southall et al. (2016)) and the Atlantic BRS (one tag; for details, see Southall et al. (2020)). The tags were programmed to release after a predetermined period, if they had not already detached from the animal, and were recovered to download recorded data. Pressure recordings were converted to depths and orientation offset from tag position were performed using calibration information for each tag (Johnson and Tyack (2003)). Data were downsampled to 15-second resolution for analysis to reduce computational effort.

Beaked whales typically perform two types of dives: deep dives (up to several kilometers of depth), during which their foraging activity occurs, and shallow dives (Shearer et al. (2019)). For this analysis we only retained dives with a maximum depth greater than 700 m to investigate changes in foraging behaviour (DeRuiter et al. (2013)). The processed data set included 13 dives, each about one hour in length.

Two of the whales were exposed to midfrequency active sonar during a deep dive, each for a period of 30 minutes. DeRuiter et al. (2013) used dive summaries, such as duration and maximum depth, to investigate behavioural changes following sonar exposure from these data. In this paper we propose a different approach based directly on the high-resolution data, which focuses on response at a short temporal scale after the start of exposure.

**3. Behavioural response model.** One approach to describing behavioural responses of animals consists in specifying two components: a model of baseline behaviour and a model for deviations from that baseline ("response" behaviour). The mathematical formulation of both components will depend on the specific application and should be informed by the research question. One key point is that the definition of the baseline model partially (and implicitly) determines what constitutes a behavioural response, regardless of the model used for the deviations. In this section we propose varying-coefficient SDEs as a flexible model of baseline behaviour, applicable to various data types. We describe how responses can be modelled in that framework, in particular, using difference smooths, and we discuss underlying assumptions.

3.1. *Varying-coefficient stochastic differential equations.* Varying-coefficient SDEs are a versatile class of time series models with time-varying dynamics (Michelot et al. (2021)). We consider the Itô SDE for the continuous-time process $(Z_t)$,

$$dZ_t = \mu(Z_t, \boldsymbol{\theta}_t)\,dt + \sigma(Z_t, \boldsymbol{\theta}_t)\,dW_t,$$

where $\mu$ is the drift function and $\sigma$ the diffusion function, $W_t$ is a standard Wiener process, and $\boldsymbol{\theta}_t$ is a vector of time-varying parameters. The drift $\mu$ measures the expected change over infinitesimal time increments, and the diffusion $\sigma$ captures stochastic variability around this expected change. In the animal movement context, $Z_t$ can, for example, be the location, velocity, or depth of the animal at time $t$. The drift function might, therefore, represent a preferred direction of movement or attraction toward a point in space. The diffusion captures random deviations, that is, aspects of movements that cannot be described directly in the drift.

The functions $\mu$ and $\sigma$ are often chosen to have a simple parametric form to help with implementation and interpretation. In the following we use Brownian motion and the Ornstein–Uhlenbeck process for illustration, as these are the models we use in the case study, but the methodology generalises to other SDEs. In the case of Brownian motion, we have $\mu(Z_t, \boldsymbol{\theta}_t) = a$ and $\sigma(Z_t, \boldsymbol{\theta}_t) = \sigma$, where $a \in \mathbb{R}$ and $\sigma > 0$ are constant drift and diffusion parameters, respectively. Similarly, the Ornstein–Uhlenbeck process is defined by the SDE with $\mu(Z_t, \boldsymbol{\theta}_t) = b(a - Z_t)$ and $\sigma(Z_t, \boldsymbol{\theta}_t) = \sigma$, where $a \in \mathbb{R}$ is the long-term mean of the process, $b > 0$ is the strength of the attraction to the mean, and $\sigma > 0$ measures the volatility. In the varying-coefficient approach, the parameters of the SDE are specified as time-varying functions of covariates. This allows for great flexibility in the dynamics of the modelled process, while retaining the simple interpretation of parametric SDEs.

The derivation of the likelihood of an SDE observed at discrete time intervals requires evaluating its transition density, that is, the function $p(Z_{t+\Delta}|Z_t)$ for each time interval of observation $\Delta > 0$. This transition density is known in closed form for the special cases considered here (Brownian motion and Ornstein–Uhlenbeck process) and in the varying-coefficient setting, we use the value of the parameter at the start of the interval. This is an

approximation based on the assumption that the SDE parameters are constant over the time interval of observation. That is, we use the transition densities:

$$\text{Brownian motion:} \quad Z_{t_1} | Z_{t_0} = z_0 \sim N\big[z_0 + a_{t_0}\Delta, \sigma_{t_0}^2\Delta\big],$$

$$\text{Ornstein–Uhlenbeck:} \quad Z_{t_1} | Z_{t_0} = z_0 \sim N\left[(1 - e^{-b_{t_0}\Delta})a_{t_0} + e^{-b_{t_0}\Delta} z_0, \frac{\sigma_{t_0}^2}{2b_{t_0}}(1 - e^{-2b_{t_0}\Delta})\right],$$

where $\Delta = t_1 - t_0$. More generally, when the transition density is not tractable, a discretisation approach such as the Euler-Maruyama method can be used (Michelot et al. (2021)). This method assumes that the movement parameters ($a_t$, $b_t$, $\sigma_t$) are constant over each time interval; this is a mild assumption for data collected at high resolution, but performance might decrease for coarse data.

We use the formalism of generalised additive models (GAMs) to specify each parameter $\theta_t$ as a function of covariates,

$$h(\theta_t) = \alpha_0 + f_1(x_{1t}) + f_2(x_{2t}) + \cdots, \tag{1}$$

where $h$ is a link function, $\alpha_0$ is an intercept parameter, and the function $f_j$ represents the relationship between the covariate $x_j$ and the parameter (e.g., linear, nonlinear, or random effect). We denote as $\boldsymbol{\alpha}$ the vector of parameters for linear model components (e.g., the intercept). We model nonlinear components with penalised splines, that is, each function $f_j$ is modelled as the linear combination of basis functions $\{\psi_{jl}\}_{l=1}^{L}$, and the basis coefficients $\beta_l^{(j)}$ are penalised to impose smoothness,

$$f_j(x_{jt}) = \sum_{l=1}^{L} \beta_l^{(j)} \psi_{jl}(x_{jt}), \quad \boldsymbol{\beta}^{(j)} \sim N\big(\mathbf{0}, \lambda_j^{-1} \boldsymbol{S}_j^{-}\big), \tag{2}$$

where $\lambda_j$ is a smoothness parameter, and $\boldsymbol{S}_j^{-}$ is the pseudo-inverse of some chosen penalty matrix (Wood (2017), Michelot et al. (2021)). The choice of the basis functions $\psi_{jl}$ and of the penalty matrix $\boldsymbol{S}_j$ determines the type of spline; this flexible formulation includes cubic regression splines and thin plate regression splines as special cases. This approach can also include i.i.d. normal random effects, where the basis functions are indicators and $\boldsymbol{S}_j$ is a diagonal matrix (Wood (2017)). The goal of inference is then to estimate the linear model coefficients $\boldsymbol{\alpha}$, the basis function coefficients $\boldsymbol{\beta}$ for the nonparametric relationships, and smoothness parameters $\boldsymbol{\lambda}$ of nonlinear functions (or precision of random effects).

In a few simple special cases, this model reduces to a GAM or to a GAM for location, scale, and shape (GAMLSS; Rigby and Stasinopoulos (2005)). In particular, the Brownian motion described above can be written as $(Z_{t_1} - Z_{t_0})/\Delta \sim N[a_{t_0}, \sigma_{t_0}^2]$. This is a GAMLSS with response variable $(Z_{t_1} - Z_{t_0})/\Delta$, where the response distribution is normal, and the mean (i.e., location) and standard deviation (i.e., scale) are modelled as nonparametric functions of covariates. As a consequence, GAMLSS software, such as the gamlss package in R, can be used directly in this case (Stasinopoulos and Rigby (2008)), but this does not apply to general SDEs (e.g., the Ornstein–Uhlenbeck process).

SDEs have been used to model various types of animal behaviours, including movement around a central location (Dunn and Gipson (1977)), highly directional movement (Johnson et al. (2008)), and habitat selection (Michelot et al. (2019)). Their continuous-time formulation makes it possible to analyse irregularly-sampled data and to compare or combine studies with different sampling schemes. In this framework, the specification of a baseline model requires the choice of: (1) an appropriate SDE, informed by the type of data and the animal's movement patterns and (2) relevant covariates to be included in the SDE parameters. The SDE should capture a template of the animal's behaviour under normal conditions so that deviations from that template can be quantified. We describe two examples in Section 3.4.

3.2. *Inference and implementation*. Given observations from the process $(Z_t)$ (e.g., locations obtained from a tracking device), our aim is to estimate the relationships between the SDE parameters and covariates (i.e., the functions $f_j$ in equation (1)). In practice, various spline formulations are possible, and we used the R package mgcv to specify the basis functions and penalty matrices for nonlinear and random effects in equation (1) (Wood (2017)). We followed the approach of Michelot et al. (2021) for model fitting, that is, to estimate all linear, nonlinear, and random effects of covariates on the SDE parameters. We view equations (1) and (2) as a mixed effect model, where the basis coefficients $\boldsymbol{\beta}$ of nonlinear terms are treated as random effects. The marginal likelihood of such a model, where the random effects have been integrated out, can be computed using the Laplace approximation, and we implemented it using the Template Model Builder (TMB) R package (Kristensen et al. (2016)). We optimised the marginal likelihood numerically using the function `optim()` in R to obtain estimates of the linear model parameters $\boldsymbol{\alpha}$ and smoothness parameters $\boldsymbol{\lambda}$. Predicted values of the random effects $\boldsymbol{\beta}$ (including basis function coefficients) can also be obtained, similarly to best linear unbiased predictors in mixed effect models, to derive the estimated nonlinear relationships between covariates and SDE parameters. For more detail we refer the reader to Michelot et al. (2021).

3.3. *Modelling the response*. In many animal movement analyses, it is of interest to detect behavioural changes or to compare behaviour over different phases of data. This is particularly relevant to identify effects of internal or external influences on behaviour, such as anthropogenic disturbance or habitat degradation.

In the framework of varying-coefficient SDEs, we propose decomposing the animal's movement parameters into different components for baseline and response behaviours. This can be modelled within the additive structure of equation (1), where time-varying terms can be included to capture behavioural changes after disturbance. The form of these additional terms, and the choice of the parameter on which to include them, will generally depend on the application, as different formulations might be required for different types of deviations from baseline. Perhaps the simplest response model would be to add an intercept term during the sound exposure (or for some set period after start of exposure). This could, for example, capture an unusually high (or low) level of activity directly following disturbance. This simple model requires specifying a time period over which to include the additional intercept and, although this choice could be based on biological expertise, it might be difficult in many applications. Alternatively, in studies where the level of disturbance is measured (e.g., received sound level), this could directly be included as a covariate acting on the SDE parameters. This option is attractive due to its mechanistic interpretation, but direct measurements of disturbance are not always available.

In the following we propose using separate smooth relationships between parameters and covariates for the baseline and response phases of the data. More specifically, we suggest estimating one smooth function for baseline and a "difference smooth" to measure the discrepancy between baseline and postdisturbance periods. Figure 1 illustrates the concept of a difference smooth, and we provide details for the model formulations of interest in the next section. The main assumption of this approach is that the deviation between baseline and response behaviours can be modelled using a smooth function. This would, for example, be violated if the response of interest is a "jerk" reaction corresponding to a large, yet momentary, change in movement dynamics. In cases where the assumption of smoothness holds, however, difference smooths are a convenient formulation, as they make it possible to directly carry out inference (including uncertainty quantification) on the discrepancy between baseline and postdisturbance behaviour.

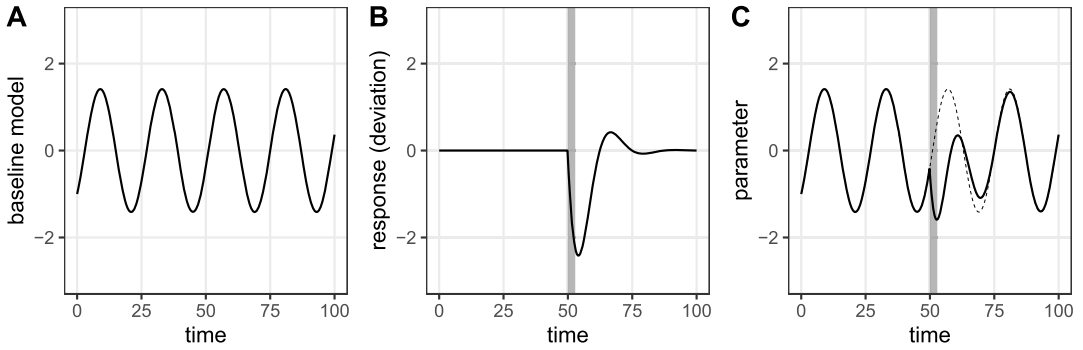


FIG. 1. *Example model terms plotted over time*: (*A*) *baseline covariate effect* $f_1(x_{1t})$, *for example*, *of time of day*, (*B*) *difference smooth measuring deviation from baseline*, $f_d(x_{d,t})$, *which eventually decays to zero*, *and* (*C*) *SDE parameter* $\theta_t$, *obtained as* $\theta_t = f_1(x_{1t}) + f_d(x_{d,t})$. *The shaded band shows the period of exposure*, *and the dashed line in* (*C*) *is the value of the parameter in the baseline model. In a behavioural response study*, *the main focus is the deviation from baseline* (*B*).

### 3.4. *Case studies*.

3.4.1. *Horizontal avoidance*. Avoidance is defined as movement away from a disturbance, and it has been documented in beaked whales (e.g., Tyack et al. (2011)). When this behaviour is observed in the Easting-Northing plane (rather than in the depth dimension), we call it horizontal avoidance. In the framework of varying-coefficient SDEs, we propose modelling horizontal avoidance as follows. We define the location process $(\mathbf{Z}_t)$ of the animal (Easting-Northing) as an isotropic two-dimensional Ornstein–Uhlenbeck process, where each coordinate is defined by

$$dZ_t = b(a_t - Z_t)\,dt + \sigma\,dW_t, \tag{3}$$

where $a_t$ is a time-varying centre of attraction, $b > 0$ is the strength of attraction to $a$, and $\sigma > 0$ is the diffusion parameter. We define the time-varying centre of attraction as

$$a_t = \alpha_0 + f_d(t)\,I_{\{t \geq t_{\text{exp}}\}}, \tag{4}$$

where $I$ is the indicator function, and $t_{\text{exp}}$ is the time of start of exposure. Combining equations (3) and (4), we obtain the following SDEs before and after start of exposure:

$$\text{Before:} \quad dZ_t = b(\alpha_0 - Z_t)\,dt + \sigma\,dW_t.$$

$$\text{After:} \quad dZ_t = b(\alpha_0 - Z_t)\,dt + bf_d(t)\,dt + \sigma\,dW_t.$$

The "before" model is an OU process with centre of attraction $\alpha_0$. The "after" model is a modification of that process with an additional drift (or "advection") term $bf_d(t)$. That is, the animal's movement after start of exposure is driven by two opposing forces: attraction toward a long-term central location $\alpha_0$ and time-varying advection away from that location, corresponding to deviation from baseline. In this example the function $f_d$ can, therefore, be interpreted as the animal's horizontal avoidance.

Gurarie et al. (2017) propose a similar approach, where an animal's home range shift is estimated from tracking data, based on OU or related models. They model a range shift as a linear function; that is, the centre of attraction moves along a straight line from one position to another. Our approach can, therefore, be viewed as a generalisation, where the range shift is estimated as a nonlinear function.

3.4.2. *Disruption of foraging dive behaviour.* Past studies have reported that sound exposure can prompt a beaked whale to stop echolocating during a foraging deep dive, leading to decreased potential energetic gains (e.g., DeRuiter et al. (2013)). Beaked whales have also been observed to extend nonforaging dives to depths beyond that of nonforaging dives observed in baseline, which may increase energetic costs. Here we model the depth $D_t$, using varying-coefficient Brownian motion,

$$dD_t = a_t\,dt + \sigma_t\,dW_t, \tag{5}$$

where $a_t$ and $\sigma_t$ are the time-varying drift and diffusion parameters, respectively. The drift is modelled as a function of proportion of time through dive $x_{1t} \in [0, 1]$ to capture the shape of dives,

$$a_t = \alpha_0^a + f_1^a(x_{1t}). \tag{6}$$

The diffusion $\sigma_t$ is also assumed to depend on $x_{1t}$ with a different relationship during baseline and response phases. We also include a dive-specific random intercept in $\sigma_t$ to capture heterogeneity in the data. Finally, the model is

$$\log(\sigma_t) = \alpha_0^\sigma + \alpha_{d_t}^\sigma + f_B^\sigma(x_{1t}) + \sum_{k=1}^{K} f_{R,k}^\sigma(x_{1t}) \times I_{\{d_t=k\}} \times I_{\{x_{2t}=1\}}, \tag{7}$$

where $d_t \in \{1, \ldots, K\}$ is the dive index at time $t$, $f_B^\sigma$ describes the baseline model, $f_{R,k}^\sigma$ is the difference smooth for exposed dive $k$, and $x_{2t}$ is a binary variable equal to 0 before start of exposure and 1 after. The parameter $\alpha_0^\sigma$ is the population-level mean intercept, and the dive-specific random intercepts are assumed to follow $\alpha_k^\sigma \overset{\text{i.i.d.}}{\sim} N(0, \nu^2)$, where $\nu^2$ measures variance around the population mean. The indicator functions ensure that a separate difference smooth is included for each exposed dive and that it is only added after the start of exposure.

3.5. *Measurement error using state-space models.*

3.5.1. *State-space formulation.* Measurement error is common in animal tracking data, and it is in particular present in the Argos locations analysed in Section 6. State-space models have been proposed to account for observation error in animal movement studies (Anderson-Sprecher and Ledolter (1991), Jonsen, Myers and Mills Flemming (2003)). In this section we describe how the SDEs presented above can be embedded into a state-space formulation. In the case study, measurement error only arises in two-dimensional tracking data used to detect horizontal avoidance. Therefore, we present the methods in the special case of the varying-coefficient Ornstein–Uhlenbeck process (described in Section 3.4.1). However, the approach can be applied directly to other SDEs where the transition density is normal (or approximately normal, e.g., under the Euler–Maruyama discretisation).

Let $\boldsymbol{Z}_t$ be the two-dimensional position of the animal at time $t$, described by an isotropic Ornstein–Uhlenbeck process (i.e., both dimensions are described by the same parameters), and let $\tilde{\boldsymbol{Z}}_i$ be a (noisy) observation obtained at time $t_i$. Assuming that the measurement error can be modelled with a normal distribution, we consider the state-space formulation with the following observation and latent state equations:

$$\text{Observation:} \quad \tilde{\boldsymbol{Z}}_i = \boldsymbol{Z}_{t_i} + \boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \boldsymbol{\Omega}_i), \tag{8}$$

$$\text{Latent state:} \quad \boldsymbol{Z}_{t_{i+1}} \sim N\Big[(1 - e^{-b_{t_i}\Delta_i})\boldsymbol{a}_{t_i} + e^{-b_{t_i}\Delta_i}\boldsymbol{Z}_{t_i}, \frac{\sigma_{t_i}^2}{2b_{t_i}}(1 - e^{-2b_{t_i}\Delta_i})\boldsymbol{I}_2\Big], \tag{9}$$

where $\Delta_i = t_{i+1} - t_i$, $\boldsymbol{\Omega}_i$ is the measurement error covariance matrix at time $t_i$, and $\boldsymbol{I}_2$ is the $2 \times 2$ identity matrix. Here the latent state equation is simply the transition density of the Ornstein–Uhlenbeck process.

3.5.2. *Inference with the Kalman filter.* Inference for this model can be carried out using the Kalman filter, which provides a computationally efficient method to evaluate the likelihood and to obtain one-step-ahead estimates of the latent state variables (Durbin and Koopman (2012)). To apply this method, we rewrite equations (8) and (9) as linear equations in matrix notation. The observation model (equation (8)) can be written as

$$\tilde{\boldsymbol{Z}}_i = \boldsymbol{A}\boldsymbol{Z}_{t_i} + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \boldsymbol{\Omega}_i), \tag{10}$$

where $\boldsymbol{A} = \boldsymbol{I}_2$. Similarly, we can write the latent state model (equation (9)) as

$$\boldsymbol{Z}_{t_{i+1}} = \boldsymbol{T}_i \boldsymbol{Z}_{t_i} + \boldsymbol{B}_i \boldsymbol{u}_i + \boldsymbol{\eta}_i, \quad \boldsymbol{\eta}_i \sim N(\boldsymbol{0}, \boldsymbol{Q}_i),$$

where $\boldsymbol{T}_i = e^{-b_{t_i}\Delta_i}\boldsymbol{I}_2$, $\boldsymbol{B}_i = (1 - e^{-b_{t_i}\Delta_i})\boldsymbol{I}_2$, $\boldsymbol{u}_i = \boldsymbol{a}_{t_i}$, and $\boldsymbol{Q}_i = \sigma^2_{t_i}(1 - e^{-2b_{t_i}\Delta_i})/(2b_{t_i})\boldsymbol{I}_2$. Durbin and Koopman (2012) describe the algorithm for the Kalman filter in terms of those matrices (Section 4.3.2) and the derivation of the model likelihood as a by-product (Section 7.2.1).

Other state-space model methods can be applied directly using the formulation highlighted above, such as Kalman smoothing. Kalman smoothing is an algorithm that returns predictions of the latent state variables, given the full observed time series, as well as uncertainty estimates (Durbin and Koopman (2012)). This might be particularly useful in studies where reconstructing the true trajectory of an animal is of primary interest.

## 4. Uncertainty quantification and model checking.

4.1. *Confidence intervals for nonparametric terms.* A key challenge is to estimate the difference smooths and determine whether they clearly deviate from zero. One approach is to compute confidence intervals on the difference smooths, where overlap with zero may be interpreted as lack of clear deviation. Two types of confidence intervals can be derived for a smooth function with different interpretations. *Pointwise* confidence intervals can only be used to make statements about uncertainty at a given covariate value, whereas *simultaneous* confidence intervals represent joint uncertainty across the domain of definition of the function. Beyond difference smooths, confidence intervals are crucial to interpret relationships between SDE parameters and covariates (e.g., in baseline model).

The two types of intervals are contrasted in Figure 2, and we describe methods to derive them based on posterior simulations. Note that, even though we do not carry out full Bayesian inference, we use terminology from the empirical Bayes view of hierarchical models (Miller

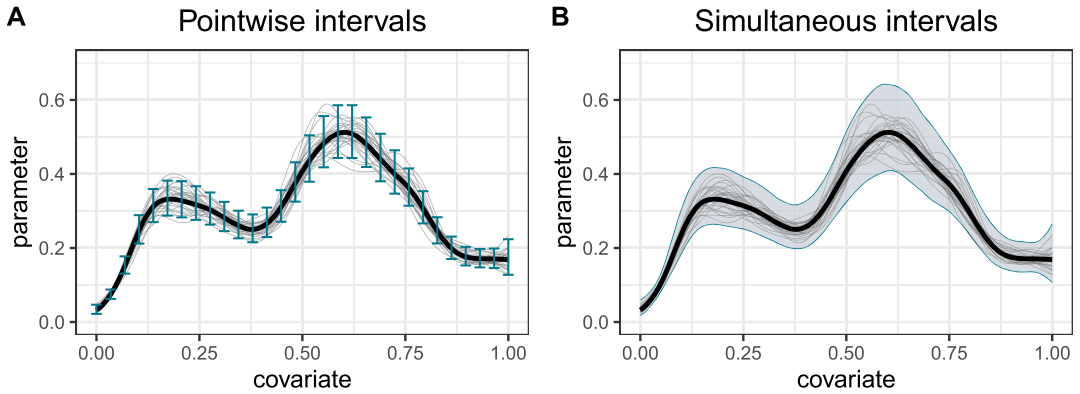


FIG. 2. *Illustration of confidence intervals. The thick black line is the mean estimate, and thin black lines are posterior samples for the spline*: (*A*) *The vertical segments show pointwise* 95% *confidence intervals on a grid of values of the covariate.* (*B*) *The shaded area is a* 95% *simultaneous confidence band.*

(2019)). We, therefore, call the joint distribution of fixed and random effect parameters the "posterior." This is approximated by a multivariate normal distribution centred on the maximum likelihood estimates $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})$, with covariance matrix $\hat{\boldsymbol{\Sigma}}$ derived from the inverse of the Hessian of the log-likelihood (e.g., using the function `sdreport` in the TMB package; Kristensen et al. (2016)).

4.1.1. *Pointwise confidence intervals.* Consider a grid over the range of the covariate of interest, $(x_1, x_2, \ldots, x_M)$. We can obtain pointwise $100(1-\alpha)\%$ confidence intervals, as follows:

1. Generate $K$ posterior draws of all fixed and random effect parameters from $N(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\Sigma}})$. (Note that, because $\hat{\boldsymbol{\gamma}}$ includes $\hat{\boldsymbol{\lambda}}$, uncertainty on the smoothness parameter is propagated to posterior draws of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ used in the following steps.)
2. From these posterior draws, derive $K$ realisations of the smooth function.
3. For each point $x_m$ of the grid, compute quantiles of the $K$ functions with probabilities $\alpha/2$ and $1-\alpha/2$. These correspond to the lower and upper bounds of the confidence interval, respectively.

From a Bayesian viewpoint, the $K$ realisations are draws from the posterior distribution of the smooth, and the interval can, therefore, be interpreted as a credible interval. They can also be viewed as confidence intervals with the expected coverage "across the function" (Marra and Wood (2012)). That is, if we denote as $p_m$ the proportion of such intervals that include the true function at $x_m$, then we expect $(p_1 + p_2 + \cdots p_M)/M \approx 1-\alpha$ (e.g., 0.95 for 95% confidence intervals). Pointwise confidence intervals are illustrated in Figure 2(A).

Although this procedure is for one smooth function, confidence intervals on the SDE parameter $\theta_t$ can be derived similarly. The only modification is that, in step 2, a realisation of the SDE parameter across the covariate grid needs to be computed. This requires adding other model terms (i.e., effects of other covariates, fixed to a given value) and applying the inverse link function.

4.1.2. *Simultaneous confidence intervals.* An alternative approach to quantify uncertainty in a function is to derive simultaneous confidence intervals. A simultaneous $100(1-\alpha)\%$ confidence band has the following interpretation: $100(1-\alpha)\%$ of such confidence intervals will include the true smooth function in its entirety. This requirement is more stringent than for pointwise intervals, and simultaneous intervals, therefore, tend to be wider. Figure 2(B) shows an example of 95% simultaneous confidence band.

Here, we follow the simulation-based method described by Ruppert, Wand and Carroll (2003) to obtain simultaneous confidence intervals over the grid $\boldsymbol{x} = (x_1, \ldots, x_M)$. We outline the main steps but refer to Section 6.5 of Ruppert, Wand and Carroll (2003) for details. In this section we denote as $f$ the true function, $\hat{f}$ the estimated smooth, $\boldsymbol{y} = (f(x_1), \ldots, f(x_M))$, and $\hat{\boldsymbol{y}} = (\hat{f}(x_1), \ldots, \hat{f}(x_M))$:

1. Generate $K$ posterior draws of $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$ from $N(\boldsymbol{0}, \hat{\boldsymbol{\Sigma}})$.
2. From each posterior draw, derive a realisation of the difference between the true function and the estimated smooth, as $\hat{\boldsymbol{y}} - \boldsymbol{y} = \boldsymbol{C}_x(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$, where $\boldsymbol{C}_x$ is the design matrix of basis function evaluations over $\boldsymbol{x}$.
3. From each realisation, approximate the standardised difference between $\hat{f}$ and $f$ by

$$H = \max_{m=1,\ldots,M} \left| \frac{(\boldsymbol{C}_x[\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}])_m}{\widehat{\mathrm{SD}}(\hat{y}_m - y_m)} \right|,$$

where the standard deviation in the denominator is measured from the $K$ replications of $\hat{\boldsymbol{y}} - \boldsymbol{y}$.

4. The simultaneous confidence interval is

$$\hat{\boldsymbol{y}} \pm q_{1-\alpha} \widehat{\mathrm{SD}}(\hat{\boldsymbol{y}} - \boldsymbol{y}),$$

where $q_{1-\alpha}$ is the $(1-\alpha)$ quantile of $H$.

The choice between pointwise and simultaneous intervals depends on the application and on whether joint statements across the range of the smooth are required. We suggest that simultaneous confidence intervals are a natural choice to quantify uncertainty on the difference smooths that measure deviations from baseline behaviour (e.g., $f_d$ in equation (4)). Indeed, to determine whether there is clear evidence of deviation, the question of interest is whether the identically zero function is included in the confidence region (rather than whether the confidence region overlaps zero for some covariate value, which is a weaker statement). As with pointwise intervals, confidence bands for the SDE parameter can also be computed using this method, where additional model terms need to be added to $\boldsymbol{C}_x$.

4.2. *Posterior predictive checks.* We propose a simulation-based approach to model checking for a fitted varying-coefficient SDE (Meng (1994)). The general idea is to simulate from the fitted model and compare patterns in the simulated data and in the observed data, where discrepancies suggest lack of fit. The suggested procedure is as follows:

1. Generate $K$ draws from the posterior distribution of fixed and random effects, $\{\boldsymbol{\gamma}^{(1)}, \ldots, \boldsymbol{\gamma}^{(K)}\}$.
2. Using each posterior draw $\boldsymbol{\gamma}^{(k)}$, simulate a time series $\boldsymbol{z}^{(k)}$ over an appropriate time period for comparison with the observed time series.
3. Compute a relevant summary statistic for each simulated time series, $g(\boldsymbol{z}^{(1)}), \ldots, g(\boldsymbol{z}^{(K)})$, which measures an important feature of the data-generating process.
4. Compute the summary statistic for the observed time series, $g(\boldsymbol{z})$.
5. Compare $g(\boldsymbol{z})$ to the distribution of the $g(\boldsymbol{z}^{(k)})$ to assess how compatible the observed data are with the estimated model. This could, for example, involve the computation of a $p$-value, as the proportion of $g(\boldsymbol{z}^{(k)})$ which are more extreme than $g(\boldsymbol{z})$ (Meng (1994)).

This method is used to check goodness-of-fit of a model of baseline diving behaviour in Section 6.2, where the test statistics are characteristics of a dive (e.g., proportion of time descending, proportion of time spent under 500 m).

**5. Simulation study.** We assessed the performance of the workflow outlined in Section 3.1 to estimate deviations from baseline using simulations. We simulated data from a Brownian motion, with drift and diffusion parameters specified as known functions of a time-varying covariate $x_1$. The relationship between the diffusion parameter and $x_1$ took two different forms, depending on a binary covariate representing disturbance or response behaviour. We then fitted a varying-coefficient SDE with a difference smooth to capture the discrepancy in diffusion between baseline and response behaviour.

We simulated data from a Brownian motion with no drift and with time-varying diffusion parameter defined as

$$\text{Diffusion (baseline):} \quad \sigma_t^B = 0.5 - 1.5(x_{1t} - 0.5)^2,$$

$$\text{Diffusion (response):} \quad \sigma_t^R = 0.05 + 5(x_{1t} - 0.5)^2,$$

where $x_{1t} \in [0, 1]$ was analogous to "proportion of time through dive" in the beaked whale diving study (Section 3.4.2). For each iteration of the simulation, we generated nine independent time series: eight from the baseline model and one that started in the baseline model and switched to the response model when $x_{1t} \geq 0.25$. Each time series contained $n = 200$ points,

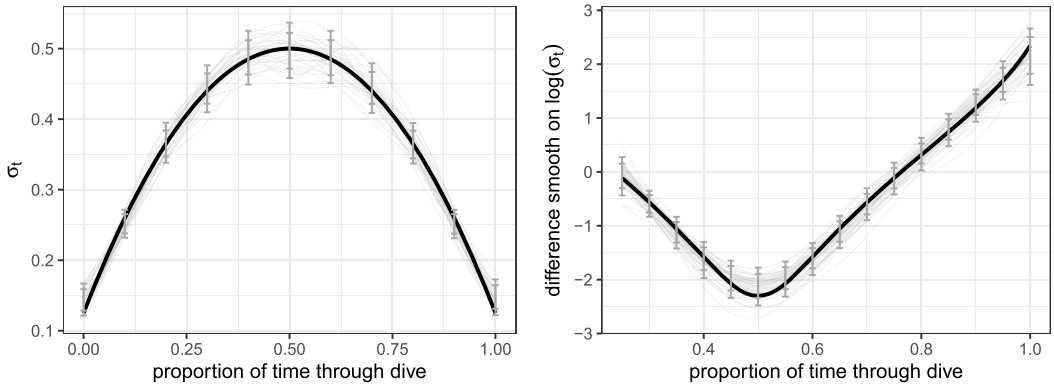


FIG. 3. *Estimated baseline model* (*left*) *and difference smooths* (*right*) *in simulation study. The thin grey lines are* 50 *randomly-selected estimated smooths* (*out of* 2000), *the vertical grey segments show the* 2.5%, 10%, 90%, *and* 97.5% *pointwise quantiles of the* 2000 *estimated smooths*, *and the thick black line is the true function used in the simulation*.

at random times uniformly distributed between $t_1 = 0$ and $t_n = 10$, to check that the method works with irregular time intervals.

For each simulated data set, we fitted a varying-coefficient Brownian motion with a difference smooth on the diffusion parameter to capture the discrepancy between the dynamics of the process during baseline and response phases, similar to equation (7). We repeated this procedure 2000 times, and the results are shown in Figure 3. The results suggest that both the baseline model and the deviation from baseline (difference smooth) were well estimated. In particular, both the smoothness and the shape of the true functions used to simulate were captured well by the fitted splines. We also used these simulations to check the coverage of the simultaneous confidence intervals for the difference smooth. We found that the 95% confidence band included the entire true function in 95.6% of the simulation runs, indicating good coverage.

**6. Beaked whale case study.** For all analyses we used the R package smoothSDE, available on Github at https://github.com/TheoMichelot/smoothSDE (Michelot et al. (2021)). The online Supplementary Material include code and data for the analysis (Michelot et al. (2023a)) and additional details about implementation (Michelot et al. (2023b), Appendix B).

6.1. *Horizontal avoidance.* We analysed the Argos trajectory, described in Section 2.1, using the model for horizontal avoidance, described in Section 3.4.1, embedded in a state-space formulation (Section 3.5) to account for measurement error. The error ellipses from the satellite tag were used to create a covariance matrix for the animal's location at each time of observation, which was then used to account for measurement uncertainty in the likelihood ($\boldsymbol{H}_i$ in equation (10)). The data set was complemented with goniometer observations, which were more precise but only covered a short time period. The data were highly irregular, with intervals ranging from a few seconds to over a day, but the continuous-time approach could still be applied directly.

Figure 4 shows the time-varying centre of attraction parameters in each dimension (Easting and Northing) between the start of exposure and the end of the study. The negative drift in both dimensions suggests that the centre of attraction deviated toward the south-west for about a week following the start of exposure in the afternoon of June 3rd (June 4 to June 11). The maximum deviation in the first coordinate was estimated to be about 50 km to the west, and the maximum deviation in the second coordinate was about 125 km to the south. After

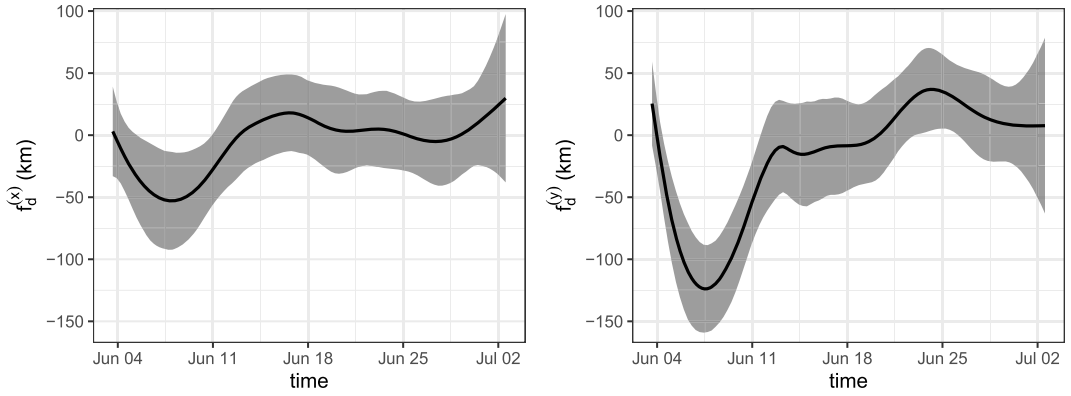


FIG. 4. *Estimated difference smooths for beaked whale centre of attraction, in the x coordinate (left) and y coordinate (right), as functions of time (after start of exposure). Grey areas are 95% simultaneous confidence intervals. Deviations from zero suggest drift away from the whale's baseline centre of attraction (i.e., horizontal avoidance).*

June 12th, the whale seemed to revert to its baseline centre of attraction for the remainder of the study period.

This analytical method has provided evidence of unusual horizontal movement following exposure, and the consequences of such a large-scale movement (in both time and space) for the individual whale are not fully understood. These results should be interpreted in the context of all other available information (e.g., dive data, visual observations, biological knowledge, expert judgement) to inform a conclusion about whether the behaviour change was a response to the sonar exposure, the severity of the response, and the possible consequences (which is beyond the scope of this analysis).

6.2. *Unusual diving behaviour.* We used the model for depth, described in Section 3.4.2, to analyse time series of depth collected from five beaked whales. The data set comprised 13 deep dives and included two controlled exposure experiments. We downsampled depth to a 15-second time resolution to reduce the computational cost of model fitting while retaining information about fine-scale behaviour.

The estimated parameters for the baseline model are presented in Figure 5 as functions of the proportion of time through a dive. The drift parameter, which measures the mean

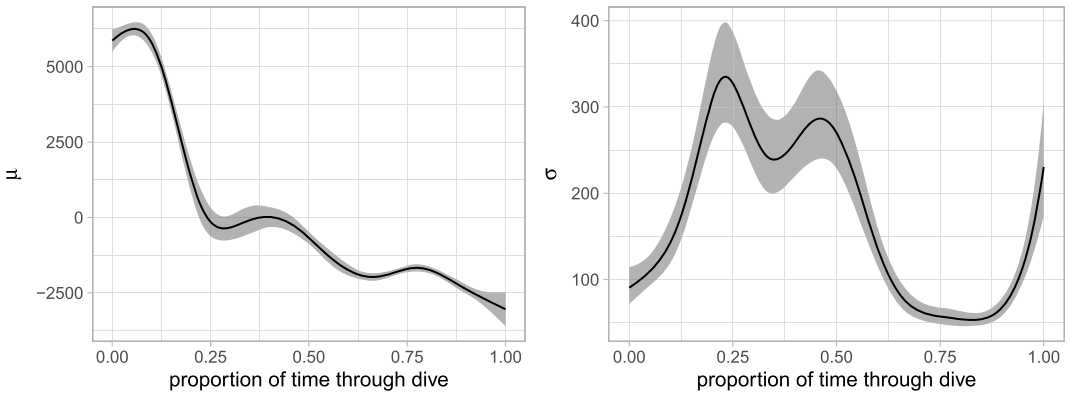


FIG. 5. *Estimated baseline parameters in the study of beaked whale diving behaviour: The drift μ is the expected change (left), and the diffusion σ is the variability (right). The shaded areas show 95% pointwise confidence bands.*

direction of change, was positive during the descent phase of the dive (because the depth increases), then close to zero during the bottom phase, and negative during the ascent phase (when the depth decreases and the animal returns to the surface). The diffusion parameter was highest during the bottom phase of the dive, suggesting high variability due to active foraging behaviour.

We used posterior predictive checks to evaluate whether the chosen Brownian motion was an appropriate model of baseline diving behaviour. We applied the procedure, described in Section 4.2, to compare the true data to simulations from the model. We simulated 1000 baseline dives and compared them to the observed baseline dives based on the following metrics:

- Proportion of time ascending, measured by proportion of time steps, where $D_{i+1} > D_i + 10$ (i.e., depth increases by more than 10 metres over a 15-second interval).
- Proportion of time descending, that is, proportion of time steps where $D_{i+1} < D_i - 10$.
- Maximum depth.
- Proportion of time spent deeper than 500 m.
- Proportion of time spent deeper than 1000 m.
- Persistence in vertical direction of movement, that is, proportion of consecutive pairs of time steps where direction of movement remains the same (either ascending or descending).

These metrics were chosen to assess how well the fitted model captured the shape of baseline dives. The results are shown in Figure 6. For the first five metrics, the mean observed value lay well within the distribution of simulation values, suggesting that features related to the overall shape of dives were well captured by the model. The observed data, however, displayed stronger directional persistence than the simulated dives. This illustrates the limited ability of Brownian motion to capture persistent movement, as there is no built-in mechanism to create directional autocorrelation. It is worth noting that the directional persistence of the simulated dives was between 0.68 and 0.86, which is much higher than the value of 0.5 expected under a simple random walk with no time-varying parameters. This is because some correlation in direction is induced by the model used for the drift parameter of the process. A possible improvement would be to consider the continuous-time correlated random

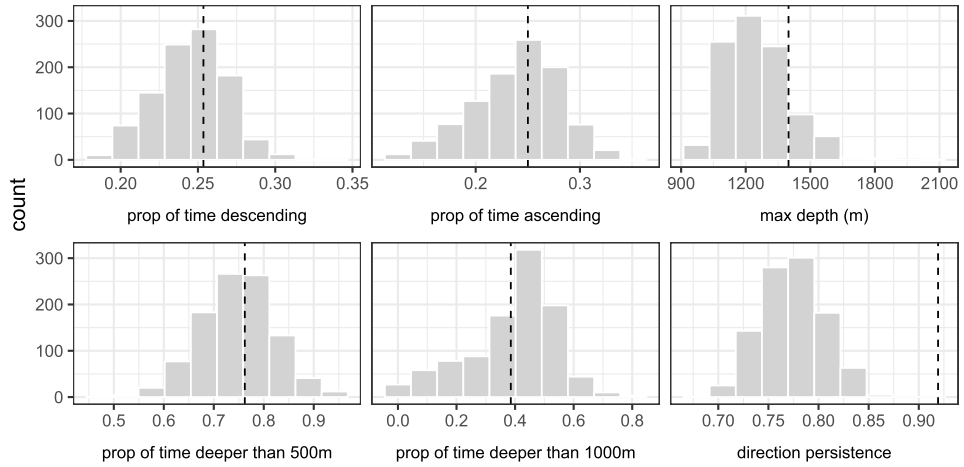


FIG. 6. *Goodness-of-fit for baseline model of beaked whale diving behaviour. Each plot contrasts the observed mean of a metric for baseline dives (vertical dotted line) and a histogram of values obtained from* 1000 *simulated dives. An observed value in the tails of the distribution suggests lack of fit. The metrics are described in the text.*

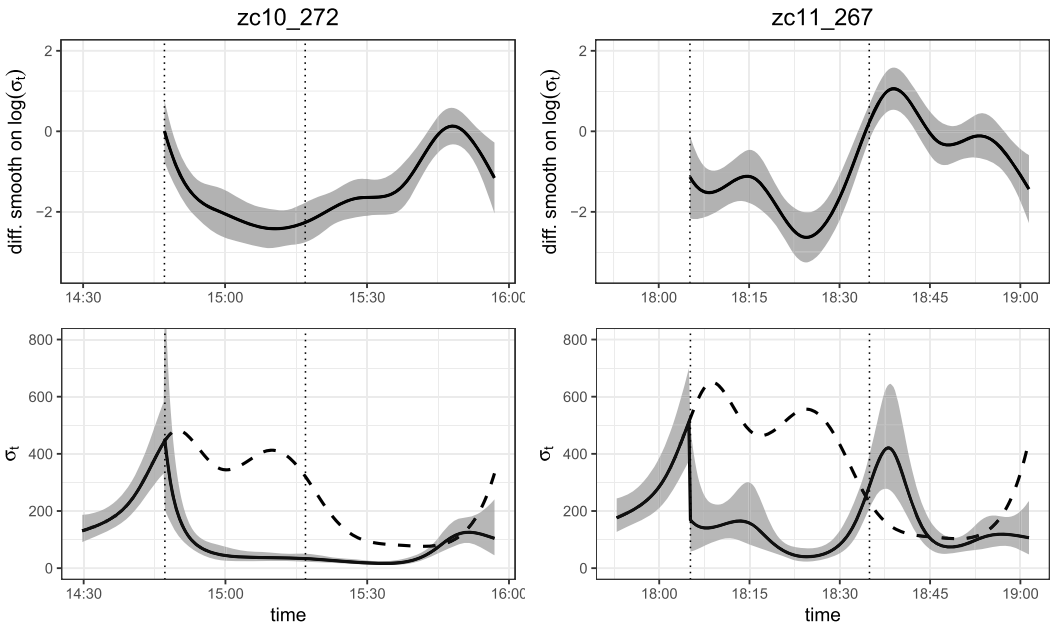


FIG. 7. *Results of analysis of beaked whale diving behaviour. Top row*: *Difference smooths for the diffusion parameter* $\sigma_t$ (*on log scale*) *during the two exposed dives with* 95% *simultaneous confidence bands. Deviation from zero suggests behaviour inconsistent with baseline model. Bottom row*: *Dashed lines show the baseline estimate for* $\sigma_t$, *and solid lines show the response model* (*including difference smooths*) *with* 95% *simultaneous confidence bands. The baseline model is not identical for the two dives* (*and also different from Figure* 5) *due to the dive-specific random intercept. In all plots the vertical dotted lines mark the start and end of sound exposure.*

walk, which directly captures autocorrelation in velocity (Johnson et al. (2008), Michelot et al. (2021)), and we discuss this in Section 7.

In this model behavioural responses were modelled using difference smooths, that is, functions capturing deviations from the baseline model during exposed dives. One difference smooth was estimated for the diffusion parameter $\sigma$ for each of the two exposed dives to estimate the difference in the depth variability compared to a typical baseline dive. Figure 7 shows the estimated smooths with confidence bands. Both curves display large departures from zero after the start of exposure, indicating deviations from baseline behaviour. In the first exposed dive ("zc10_272"), the most noticeable pattern is that the diffusion parameter was much lower than normal during the bottom phase. This decreased variability in depth reflects a reduced level of activity, which suggests that this animal was not searching or chasing prey; that is, this was not a foraging dive. In the second exposed dive ("zc11_267"), $\sigma_t$ was also low during the middle part of the dive, which was followed by a period of unusually high vertical activity during the third quarter of the dive. This period coincides with the bottom phase of this dive, which was delayed because the animal appeared to be engaged in a bout of shallow diving but conducted a deep dive after the start of exposure (see Figure S2 in the Supplementary Material; Michelot et al. (2023b)).

**7. Discussion.** Varying-coefficient SDEs are versatile models to study the effects of covariates on the dynamics of temporal processes. Here we have demonstrated their utility to estimate behavioural changes in cetaceans, including changes in the pattern of diving behaviour and horizontal avoidance. This method stands in contrast with previous statistical approaches applied in this context, where data have often been summarised at the time scale of dives (e.g., dive duration, maximum depth; DeRuiter et al. (2013), DeRuiter et al. (2017)). Our analysis of DTag data at a fine time scale is one of the first attempts to describe detailed

dynamics of within-dive behaviour and contrast them between baseline and exposed dives. Often, conclusions about whether an individual exhibited a behavioural response to sonar exposure requires multiple lines of evidence, and expert biological knowledge is needed to interpret model outputs. In particular, experts are required to evaluate responses in terms of their severity and possible effect on the vital rates of individuals (Southall et al. (2008), Miller et al. (2012)). We have shown that varying-coefficient SDEs can provide such lines of evidence for different types of behaviour (horizontal movement, diving) and data recorded at different resolutions from different types of telemetry devices. Critically, our method provides fine-scale detail about the nature of the response, in the context of relevant covariates, as well as the duration of the response. This is valuable additional information beyond the identification of a behavioural change point. Some other analytical methods, such as Mahalanobis distance, identify observations that are extreme with respect to the full baseline data set. These might not be able to identify behavioural responses that are within the repertoire of baseline behaviour, even if their occurrence is unusual or unexpected, given what the animal was doing at the point of exposure.

In this paper we focused on two model formulations, which were particularly relevant to the application: the varying-coefficient Brownian motion and Ornstein–Uhlenbeck process. The approach is not limited to these models though, and other SDEs could be implemented in different contexts (see examples in Michelot et al. (2021)). One other notable SDE is the continuous-time correlated random walk (or "integrated Ornstein–Uhlenbeck"), which can capture strong autocorrelation in movement and is, therefore, well suited to high-resolution tracking data (Johnson et al. (2008)). The R package crawl implements this model with (possibly nonlinear) covariate effects on the movement parameters, similar to the approach described in Section 3.1, but it does not perform automatic smoothness selection for nonlinear components (Johnson and London (2018)). The correlated random walk model might be better suited to capture autocorrelation in the depth data of Section 6.2; however, it is considerably more complex than Brownian motion and would come with greater computational cost and numerical instability for parameter estimation. In general, many SDE formulations are available, and the choice depends on the data type and research question (e.g., the OU process has a centre of attraction, whereas Brownian motion does not). Future work could further explore the trade-off between parsimony and goodness-of-fit in varying-coefficient SDEs.

The method of inference used to estimate the time-varying SDE parameters, based on the approach of Michelot et al. (2021), requires two (potential) approximations: (1) we assumed that the SDE parameters are constant over each time interval between consecutive observations, and (2) the Euler–Maruyama discretisation is required for SDEs whose closed-form transition density is not known (Section 3.1). The error introduced by both approximations could be mitigated using data augmentation methods. The idea is to define a new time grid, finer than that of observations, and apply the discretisation over those shorter time intervals. This requires integrating over the value of the process at the additional time points, which can, for example, be done using Markov chain Monte Carlo (Elerian, Chib and Shephard (2001)) or the Kalman filter. Note that this approach also requires a method to interpolate covariate values, as these are needed to evaluate the SDE parameters on the finer time grid. This would be straightforward for some covariates (e.g., time of day or proportion of time through dive) but might be challenging in many applications.

SDEs have recently been proposed to model group movement of animals (Niu, Blackwell and Skarin (2016), Milner, Blackwell and Niu (2021)). That approach could be extended to allow for time-varying dynamics, and using the methods presented in this paper, it could be used to estimate behavioural responses for multiple individual animals. This model would not be limited to individual-specific responses but could also capture changes in the interactions between individuals (e.g., group breaking off after disturbance).

The examples that we presented illustrate a general framework to analyse behavioural responses from telemetry data. The first step is to specify a model for baseline periods, typically in terms of spatial or temporal covariates of interest (e.g., time of day, habitat variable). Then additional terms can be added in the model for the SDE parameters to capture deviations from baseline during exposure phases. Difference smooths are powerful for this purpose, as they explicitly model the difference in a smooth relationship between levels of a categorical variable (which could, e.g., represent pre- and postdisturbance). We showcased how difference smooths can be interpreted in terms of behavioural response for two different applications. In particular, simultaneous confidence bands are useful to compare the deviation from baseline to the zero function. Although we checked the coverage of these confidence intervals in simulations, we note that this method might lead to a large rate of false positives in real data applications if the model assumptions are violated (e.g., if the baseline model does not adequately capture heterogeneity in baseline data). For this reason the shape and amplitude of the difference smooth should be inspected as part of the interpretation rather than merely whether it clearly differs from zero. A wide range of varying-coefficient SDEs can be implemented using the smoothSDE R package, and we anticipate that these methods will be a key tool to investigate the potential impact of disturbance, such as sonar, on individuals and populations.

**Acknowledgments.** We are very grateful to Rob Schick, Will Cioffi, Alan Gelfand, Josh Hewitt, Stacy DeRuiter, and Brandon Southall for discussions about the data and models. The data from four of the five DTags were collected as part of the SOCAL-BRS project, primarily funded by the U.S. Navy's Chief of Naval Operations Environmental Readiness Division and subsequently by the U.S. Navy's Living Marine Resources Program. Additional support for environmental sampling and logistics was also provided by the Office of Naval Research, Marine Mammal Program. All research activities for that study were authorized and conducted under U.S. National Marine Fisheries Service permit 14534, Channel Islands National Marine Sanctuary permit 2010-004, U.S. Department of Defense Bureau of Medicine and Surgery authorization, a federal consistency determination by the California Coastal Commission, and numerous institutional animal care and use committee authorizations. The data from the satellite tag and one of the DTags were collected as part of the Atlantic BRS project under National Marine Fisheries Service scientific research permit numbers 17086 and 20605 to Robin W. Baird. The tagging protocol was approved by the Institutional Animal Care and Use Committee at Cascadia Research Collective. This work was supported by the U.S. Fleet Forces Command through the Naval Facilities Engineering Command Atlantic under Contract No. N62470-15-D-8006, Task Order 50, Issued to HDR, Inc. We thank all members of the field teams involved in both the SOCAL and Atlantic BRS projects.

**Funding.** TM, RG, CH, and LT were funded by the U.S. Office of Naval Research, Grant N000141812807.

## SUPPLEMENTARY MATERIAL

**Code and data** (DOI: 10.1214/23-AOAS1776SUPPA; .zip). R code and preprocessed data sets used in the two case studies (Michelot et al. (2023a)).

**Supplementary details about data analyses** (DOI: 10.1214/23-AOAS1776SUPPB; .pdf). Additional details about the data sets and about implementation (Michelot et al. (2023b)).

## REFERENCES


ANDERSON-SPRECHER, R. and LEDOLTER, J. (1991). State-space analysis of wildlife telemetry data. *J. Amer. Statist. Assoc.* **86** 596–602.

ANDREWS, R. D., PITMAN, R. L. and BALLANCE, L. T. (2008). Satellite tracking reveals distinct movement patterns for Type B and Type C killer whales in the southern Ross Sea, Antarctica. *Polar Biol.* **31** 1461–1468.

CIOFFI, W. R., QUICK, N. J., SWAIM, Z. T., FOLEY, H. J., WAPLES, D. M., WEBSTER, D. L., BAIRD, R. W., SOUTHALL, B. L., NOWACEK, D. P. et al. (2022). Trade-offs in telemetry tag programming for deep-diving cetaceans: Data longevity, resolution, and continuity. *BioRxiv*.

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. *Ann. Appl. Stat.* **11** 362–392. MR3634328 https://doi.org/10.1214/16-AOAS1008

DERUITER, S. L., SOUTHALL, B. L., CALAMBOKIDIS, J., ZIMMER, W. M., SADYKOVA, D., FALCONE, E. A., FRIEDLAENDER, A. S., JOSEPH, J. E., MORETTI, D. et al. (2013). First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. *Biol. Lett.* **9** 20130223.

DUNN, J. E. and GIPSON, P. S. (1977). Analysis of radio telemetry data in studies of home range. *Biometrics* 85–101.

DURBIN, J. and KOOPMAN, S. J. (2012). *Time Series Analysis by State Space Methods*, 2nd ed. *Oxford Statistical Science Series* **38**. Oxford Univ. Press, Oxford. MR3014996 https://doi.org/10.1093/acprof:oso/9780199641178.001.0001

ELERIAN, O., CHIB, S. and SHEPHARD, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69** 959–993. MR1839375 https://doi.org/10.1111/1468-0262.00226

FOLEY, H. J., PACIFICI, K., BAIRD, R. W., WEBSTER, D. L., SWAIM, Z. T. and READ, A. J. (2021). Residency and movement patterns of Cuvier's beaked whales Ziphius cavirostris off Cape Hatteras, North Carolina, USA. *Mar. Ecol. Prog. Ser.* **660** 203–216.

GURARIE, E., CAGNACCI, F., PETERS, W., FLEMING, C. H., CALABRESE, J. M., MUELLER, T. and FAGAN, W. F. (2017). A framework for modelling range shifts and migrations: Asking when, whither, whether and will it return. *J. Anim. Ecol.* **86** 943–959. https://doi.org/10.1111/1365-2656.12674

JOHNSON, D. S. and LONDON, J. M. (2018). crawl: An R package for fitting continuous-time correlated random walk models to animal movement data.

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** 1208–1215. https://doi.org/10.1890/07-1032.1

JOHNSON, M. P. and TYACK, P. L. (2003). A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE J. Oceanic Eng.* **28** 3–12.

JONES-TODD, C. M., PIROTTA, E., DURBAN, J. W., CLARIDGE, D. E., BAIRD, R. W., FALCONE, E. A., SCHORR, G. S., WATWOOD, S. and THOMAS, L. (2022). Discrete-space continuous-time models of marine mammal exposure to Navy sonar. *Ecol. Appl.* **32** e02475.

JONSEN, I. D., MYERS, R. A. and MILLS FLEMMING, J. (2003). Meta-analysis of animal movement using state-space models. *Ecology* **84** 3055–3063.

KRISTENSEN, K., NIELSEN, A., BERG, C. W., SKAUG, H. and BELL, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *J. Stat. Softw.* **70** 1–21.

MARRA, G. and WOOD, S. N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scand. J. Stat.* **39** 53–74. MR2896791 https://doi.org/10.1111/j.1467-9469.2011.00760.x

MCCLINTOCK, B. T., LONDON, J. M., CAMERON, M. F. and BOVENG, P. L. (2015). Modelling animal movement using the argos satellite telemetry location error ellipse. *Methods Ecol. Evol.* **6** 266–277.

MENG, X.-L. (1994). Posterior predictive $p$-values. *Ann. Statist.* **22** 1142–1160. MR1311969 https://doi.org/10.1214/aos/1176325622

MICHELOT, T., GLENNIE, R., HARRIS, C. and THOMAS, L. (2021). Varying-coefficient stochastic differential equations with applications in ecology. *J. Agric. Biol. Environ. Stat.* **26** 446–463. MR4292797 https://doi.org/10.1007/s13253-021-00450-6

MICHELOT, T., GLENNIE, R., THOMAS, L., QUICK, N. and HARRIS, C. M. (2023a). Code and data for "Continuous-time modelling of behavioural responses in animal movement." https://doi.org/10.1214/23-AOAS1776SUPPA

MICHELOT, T., GLENNIE, R., THOMAS, L., QUICK, N. and HARRIS, C. M. (2023b). Supplementary materials for "Continuous-time modelling of behavioural responses in animal movement." https://doi.org/10.1214/23-AOAS1776SUPPB

MICHELOT, T., GLOAGUEN, P., BLACKWELL, P. G. and ÉTIENNE, M.-P. (2019). The Langevin diffusion as a continuous-time model of animal movement and habitat selection. *Methods Ecol. Evol.* **10** 1894–1907.

MILLER, D. L. (2019). Bayesian views of generalized additive modelling. ArXiv preprint. Available at arXiv:1902.01330.

MILLER, P. J., KVADSHEIM, P. H., LAM, F.-P. A., WENSVEEN, P. J., ANTUNES, R., ALVES, A. C., VISSER, F., KLEIVANE, L., TYACK, P. L. et al. (2012). The severity of behavioral changes observed during experimental exposures of killer (Orcinus orca), long-finned pilot (Globicephala melas), and sperm (Physeter macrocephalus) whales to naval sonar. *Aquat. Mamm.* **38** 362.

MILNER, J. E., BLACKWELL, P. G. and NIU, M. (2021). Modelling and inference for the movement of interacting animals. *Methods Ecol. Evol.* **12** 54–69.

NIU, M., BLACKWELL, P. G. and SKARIN, A. (2016). Modeling interdependent animal movement in continuous time. *Biometrics* **72** 315–324. MR3515758 https://doi.org/10.1111/biom.12454

POZDNYAKOV, V., MEYER, T., WANG, Y.-B. and YAN, J. (2014). On modeling animal movements using Brownian motion with measurement error. *Ecology* **95** 247–253.

RIGBY, R. A. and STASINOPOULOS, D. M. (2005). Generalized additive models for location, scale and shape. *J. R. Stat. Soc., Ser. C* **54** 507–554. MR2137253 https://doi.org/10.1111/j.1467-9876.2005.00510.x

RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression*. *Cambridge Series in Statistical and Probabilistic Mathematics* **12**. Cambridge Univ. Press, Cambridge. MR1998720 https://doi.org/10.1017/CBO9780511755453

SHEARER, J. M., QUICK, N. J., CIOFFI, W. R., BAIRD, R. W., WEBSTER, D. L., FOLEY, H. J., SWAIM, Z. T., WAPLES, D. M., BELL, J. T. et al. (2019). Diving behaviour of Cuvier's beaked whales (Ziphius cavirostris) off Cape Hatteras, North Carolina. *R. Soc. Open Sci.* **6** 181728. https://doi.org/10.1098/rsos.181728

SOUTHALL, B. L., BOWERS, M., CIOFFI, W., FOLEY, H., HARRIS, C., JOSEPH, J., QUICK, N., MARGOLINA, T., NOWACEK, D. et al. (2020). Atlantic Behavioral Response Study (BRS): 2019 Annual Progress Report. Project report. Prepared for U.S. Fleet Forces Command. Submitted to Naval Facilities Engineering Command Atlantic, Norfolk, Virginia, under Contract No. N62470-15-D-8006, Task Order 19F4029, issued to HDR Inc., Virginia Beach, Virginia. May 2020.

SOUTHALL, B. L., BOWLES, A. E., ELLISON, W. T., FINNERAN, J. J., GENTRY, R. L., JR. GREENE, C. R., KASTAK, D., KETTEN, D. R., MILLER, J. H. et al. (2008). Marine mammal noise-exposure criteria: Initial scientific recommendations. *Bioacoustics* **17** 273–275.

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. (2019). Marine mammal noise exposure criteria: Updated scientific recommendations for residual hearing effects. *Aquat. Mamm.* **45**.

SOUTHALL, B. L., NOWACEK, D. P., MILLER, P. J. and TYACK, P. L. (2016). Experimental field studies to measure behavioral responses of cetaceans to sonar. *Endanger. Species Res.* **31** 293–315.

STASINOPOULOS, D. M. and RIGBY, R. A. (2008). Generalized additive models for location scale and shape (GAMLSS) in R. *J. Stat. Softw.* **23** 1–46.

STIMPERT, A., DERUITER, S. L., SOUTHALL, B., MORETTI, D., FALCONE, E., GOLDBOGEN, J., FRIEDLAENDER, A., SCHORR, G. and CALAMBOKIDIS, J. (2014). Acoustic and foraging behavior of a Baird's beaked whale, Berardius bairdii, exposed to simulated sonar. *Sci. Rep.* **4** 1–8.

TYACK, P. L., ZIMMER, W. M., MORETTI, D., SOUTHALL, B. L., CLARIDGE, D. E., DURBAN, J. W., CLARK, C. W., D'AMICO, A., DIMARZIO, N. et al. (2011). Beaked whales respond to simulated and actual navy sonar. *PLoS ONE* **6** e17009.

WOOD, S. N. (2017). *Generalized Additive Models*: *An Introduction with R. Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3726911