

## METHOD

# Non-linear models of species' responses to environmental and spatial gradients

Marti J. Anderson<sup>1,2</sup>  | Daniel C. I. Walsh<sup>3</sup> | Winston L. Sweatman<sup>4</sup> | Andrew J. Punnett<sup>2</sup>

<sup>1</sup>New Zealand Institute for Advanced Study (NZIAS), Massey University, Auckland, New Zealand

<sup>2</sup>PRIMER-e (Quest Research Limited), Auckland, New Zealand

<sup>3</sup>Statistical Consulting Centre, University of Auckland, Auckland, New Zealand

<sup>4</sup>School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand

## Correspondence

Marti J. Anderson, New Zealand Institute for Advanced Study (NZIAS), Massey University, Albany Campus, Auckland, New Zealand.  
Email: [m.j.anderson@massey.ac.nz](mailto:m.j.anderson@massey.ac.nz)

## Funding information

Royal Society Te Apārangi, Grant/Award Number: Marsden Grant 19-MAU-145; Royal Society of New Zealand; Institute for Advanced Study; Massey University

Editor: Forest Isbell

## Abstract

Species' responses to broad-scale environmental or spatial gradients are typically unimodal. Current models of species' responses along gradients tend to be overly simplistic (e.g., linear, quadratic or Gaussian GLMs), or are suitably flexible (e.g., splines, GAMs) but lack direct ecologically interpretable parameters. We describe a parametric framework for species–environment non-linear modelling ('*senlm*'). The framework has two components: (i) a non-linear parametric mathematical function to model the mean species response along a gradient that allows asymmetry, flattening/peakedness or bimodality; and (ii) a statistical error distribution tailored for ecological data types, allowing intrinsic mean–variance relationships and zero-inflation. We demonstrate the utility of this model framework, highlighting the flexibility of a range of possible mean functions and a broad range of potential error distributions, in analyses of fish species' abundances along a depth gradient, and how they change over time and at different latitudes.

## KEY WORDS

abundance, biomass, counts, cover, ecological statistics, environmental variables, gradient analysis, latitude, species distribution models, zero-inflation

## INTRODUCTION

Gradient analysis has a long history in ecology, reaching back to Whittaker's seminal work (1956, 1967). Responses of species along broad-scale spatial gradients (such as latitude, depth or elevation), or environmental gradients (such as nutrients, light, temperature or moisture) are typically unimodal (Gauch et al., 1974; Jamil & ter Braak, 2013; ter Braak, 1996; Westman, 1980; Whittaker, 1956, 1967). Each species is thought to have an 'optimum' (modal) value along a given gradient (e.g., the temperature which best suits the species' survival—neither too cold nor too hot, but 'just right'). Different species are expected to have different optimal positions and different 'tolerances' (spread) along the gradient,

according to their realised niche (Colwell & Rangel, 2009; Whittaker et al., 1973). Multiple species typically show a pattern of (partially) overlapping unimodal curves along a gradient of interest (ter Braak, 1996; Whittaker, 1956, 1967).

Unimodal responses of species to environmental gradients have historically been modelled using a bell-shaped or Gaussian curve (Gauch et al., 1974; Jamil & ter Braak, 2013; Johnson & Goodall, 1980; ter Braak, 1985, 1986, 1996; Westman, 1980; Yee, 2004). A generalised linear model (GLM) that includes a quadratic term might also be used (Austin et al., 1984; Makarenkov & Legendre, 2002; ter Braak & Looman, 1986; Warton et al., 2015). Although the physiological response of a species might be approximately bell-shaped, other

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Ecology Letters* published by John Wiley & Sons Ltd.

factors, such as biogeographical history, unmeasured environmental variables, dispersal limitation, sampling extent, predation, herbivory or competition can all substantially alter this shape (Austin, 1976, 1980; Bradshaw et al., 2014; Oksanen & Minchin, 2002). Empirically, mean species' responses show a wide variety of shapes; they can be asymmetric, j-shaped, truncated, peaked, flattened, plateaued, bimodal or even multi-modal. Clearly, a broader range of non-linear mathematical forms should be explored (Oksanen & Minchin, 2002).

Two logistic curves can be combined to create a suitable shape for the mean unimodal response that allows for both asymmetry and peakedness/flattening (Huisman et al., 1993; Jansen & Oksanen, 2013), but these have thus far only been applied to presence/absence or cover data with an upper bound. Generalised additive models (GAMs, Hastie & Tibshirani, 1990; Yee, 2015), including B-splines (de Boor, 2001) and P-splines (Eilers & Marx, 2021), are more flexible, and can take virtually any shape (Anderson, 2008; Fraaije et al., 2015; Rigby & Stasinopoulos, 2005; Yee & Mitchell, 1991). However, they can over-fit (so need some form of cross-validation, Gu & Wahba, 1991), may be adversely affected by excess zeros and do not yield directly interpretable parameters. More sophisticated predictive models, such as artificial neural networks (Harrison et al., 2006; Lek & Guégan, 1999), maximum entropy models (Phillips et al., 2006), regression trees/splines (Chipman et al., 2010; Elith et al., 2008; Leathwick et al., 2005; Stoklosa & Warton, 2018), or ensemble models (Araújo & New, 2007) also lack directly interpretable parameters, and many deal only with presence/absence data (Elith & Leathwick, 2009).

Useful parameters for characterising a species' mean response include the modal position,  $m$ , of the species along the gradient (directly interpretable as its optimum), and the mean abundance of the species at that modal position, that is, the maximum height,  $H$ , of the non-linear response curve. Armed with estimates of such parameters for a given species, one may then track changes in their values over time or across space. For example, a warm-water species' modal position along a latitudinal gradient may shift towards the poles due to climate change, as temperatures increase at higher latitudes over time.

The mean response is, however, just part of the story. Real field data are quite messy (Legendre, 1993). Counts of individuals (or density or biomass) have no upper bound, and typically display large residual variance (Chapman & Underwood, 2008), zero-inflation (Martin et al., 2005; Smith et al., 2012), over-dispersion and intrinsic variance–mean relationships (McArdle et al., 1990; McArdle & Anderson, 2004; Taylor, 1961).

Here, we describe a modular approach for modelling the non-linear response of a given species along an environmental or spatial gradient of interest, termed '*senlm*' (for species-environment non-linear models). This model framework couples together two key

elements: (i) a flexible, parametric, non-linear mathematical function for the mean response; and (ii) a statistical error distribution to model residual variation around the mean that will accommodate noisy and generally zero-inflated ecological field data. We focus here on a suite of statistical error distributions appropriate for the types of data used to quantify the abundance of organisms in the field (i.e., counts, biomass, densities, frequencies, cover or presence/absence). We demonstrate the utility of our proposed model framework for uncovering novel insights on species' distributions along gradients *via* analyses of North-east Pacific groundfish species *versus* depth.

## DESCRIPTION OF THE METHOD

### General modelling framework

Let  $Y$  be a non-negative random variable with observed values  $y_i$  (counts, densities, biomass, percentage cover, or presence/absence, etc.) quantifying the abundance (or relative abundance, biomass or occurrence) of a single species obtained from  $i = 1, \dots, N$  standardised sampling units situated at positions  $x_i$  along gradient  $X$  of interest. For presence/absence or percentage cover, values are bounded inclusively between 0 and 1 (or 0 and 100%). Our general model considers  $y_i$  to be drawn independently from probability distribution  $\mathcal{P}(\mu_i, \theta_E)$ , where:

- $\mu_i$  is the parameter specifying the mean of  $\mathcal{P}$ , defined by a non-linear function  $f(x; \theta_M)$  of  $x$ , requiring a set of parameters  $\theta_M$ ; so  $\mu_i = f(x_i; \theta_M)$  at position  $x_i$ ; and
- $\theta_E$  is the set of all other required parameters of  $\mathcal{P}$  excluding the mean.

For example, if  $\mathcal{P}$  is a zero-inflated negative binomial (ZINB),  $\theta_E = \{\phi, \pi\}$ , with  $\phi$  being the dispersion parameter and  $\pi$  the probability of an excess zero, then the model is  $y_i \sim \text{ZINB}(\mu_i, \phi, \pi)$ , with  $\mu_i = f(x_i; \theta_M)$ .

### The mean response

The mean response function first must be sufficient to characterise a classic unimodal, bell-shaped form, as predicted by theory, with a special focus on estimating the maximum height ( $H$ ) of the curve (i.e., its mode), particularly for data types (abundance, density, biomass) that have no known upper bound, and also the modal position ( $m$ ) along the gradient at which that maximum height occurs (i.e., the species' 'optimum'). The mean function must also allow for: (i) asymmetry, either to the right or the left; for example, a species may display a more rapid decrease in mean abundance if its tolerance/adaptability to non-optimal conditions is exceeded more quickly in

one direction (e.g., too hot) *versus* the other (e.g., too cold); (ii) flattened or pointed/peaked shapes (e.g., a species may require highly specific environmental conditions, or it may be very broadly distributed); (iii) j-shaped, to either the left or right (e.g., due to the sampling window covering only part of a species' range); and (iv) potential bi- or multi-modal mean responses, with each mode being potentially of different heights (e.g., a species may encounter heavy competition or may be limited by some other environmental parameter at its physiological optimum position along a measured gradient, hence show a 'dip' in its response, with peaks either side of this).

Mean response functions in our *senlm* framework (Table 1) include two new mathematical functions that allow a suitably flexible variety of shapes typical of species' responses to gradients: a modified *beta* function (following Austin, 1976), and a modified *sech* function. We highlight here, in particular, the modified *sech* function, designed specifically to provide ecologically relevant estimates of parameters  $m$  and  $H$  for a species' mean response directly.

A basic *sech* function has the following form:  $\text{sech}(x) = 2/(e^x + e^{-x})$ . This is a symmetric, unimodal function centred on zero. Our new proposed function has parameters  $\theta_M = \{H, m, s, r, p\}$  and is given by the following equation:

$$f_{\text{sech}}(x; \theta_M) = \frac{H}{H_m} \exp\left(\frac{rp}{s}(x - (m - x_m))\right) \left( \text{sech}\left(\frac{x - (m - x_m)}{s}\right) \right)^p$$

where:

$H$  = Maximum value of the mean function ( $H > 0$ .)

$m$  = Location of the maximum ( $-\infty < m < \infty$ )

$s$  = Spread parameter ( $s > 0$ )

$r$  = Symmetry parameter ( $-1 < r < 1$ )

$p$  = Peakedness parameter ( $p > 0$ )

**TABLE 1** Mean functions and error distributions currently included in the *senlm* framework for modelling non-linear species' responses to environmental or spatial gradients. Any mean function (column 1) can be paired with any error distribution (column 2) for a given species. Acronyms identify zero-inflated error distributions that have a probability of an excess zero that is constant ('ZI...'), linked to the log of the mean ('ZI...L') or linked directly to the mean itself ('ZI...L.mu') along the gradient

Mean functions	Error distributions
Constant	<i>Discrete (counts; abundances)</i>
Uniform	Poisson
Gaussian	Zero-inflated Poisson (ZIP; ZIPL; ZIPL.mu)
HOF	Negative binomial (NB)
Beta	Zero-inflated negative binomial (ZINB; ZINBL; ZINBL.mu)
Sech	<i>Continuous (biomass; densities; traits)</i>
Mixed Gaussian	Gamma
	Zero-inflated gamma (ZIG; ZIGL; ZIGL.mu)
	Zero-inflated inverse Gaussian (ZIIG; ZIIGL; ZIIGL.mu)
	Tweedie
	<i>Cover (percentages or proportions)</i>
	Binomial
	Tail-adjusted beta (TAB)
	Zero-inflated tail-adjusted beta (ZITAB)
	<i>Binary (presence/absence)</i>
	Bernoulli

The values  $x_m = \frac{s}{2} \log\left(\frac{1+r}{1-r}\right)$  and  $H_m = \exp\left(\frac{rp}{s} x_m\right) (\sqrt{1-r^2})^p$  are used to fix the curve so that  $m$  is the location of the mode and  $H$  is the maximum height of the curve (at  $m$ ). If  $r$  is positive (or negative) the curve is asymmetric with a broader right-hand (or left-hand) tail. Note that a simpler model (with fewer parameters) may be obtained by setting fixed values *a priori* for peakedness ( $p = 1$ ) and/or symmetry ( $r = 0$ ).

We also extended and re-parameterised mixed logistic functions (called *HOF* models after Huisman et al., 1993; see also Oksanen & Minchin, 2002; Jansen & Oksanen, 2013) so they may accommodate unbounded and zero-inflated data types. Classic Gaussian mean response curves (Gauch et al., 1974; Jamil & ter Braak, 2013; ter Braak, 1985, 1986, 1996; Westman, 1980; Yee, 2004) are also included within our framework. In addition, we propose the use of mixture models to accommodate bi-modal or multi-modal distributions; the Gaussian mixture is an obvious choice here, but mixtures of *sech*, *HOF* and/or *beta* functions (allowing asymmetry, flattened/peaked shapes for any given mode, etc.) might also be considered.

An intuitive understanding of these mean functions may be obtained by referring to the 'Articles' page of the *senlm* package on GitHub (<https://primer-e.github.io/senlm/articles/>), where visual interactive tools demonstrating each function are provided; specifically, one can move a slider for any parameter in any given function and witness its effect on the shape of the resulting species' non-linear mean response curve along the gradient.

## The error distribution

Species' responses typically are measured from standardised sampling units as non-negative values. We can

classify and characterise typical data types used to quantify species abundances or occurrences and, for each of these, we propose appropriate probabilistic error distributions,  $\mathcal{P}$ , as follows (Table 1):

- *count data* (abundances)—discrete, non-negative, no upper bound, typically over-dispersed, and with variance–mean relationships;  $\mathcal{P}$  is Poisson (P), negative binomial (NB) or zero-inflated versions of these (ZIP or ZINB).
- *biomass data or densities*—like counts, but continuous rather than discrete;  $\mathcal{P}$  is Gamma (G), zero-inflated Gamma (ZIG), inverse Gaussian (IG), zero-inflated inverse Gaussian (ZIIG) or Tweedie (T).
- *percentage cover* (or *proportional*) data—continuous with a lower bound at zero and an upper bound of 100% (or 1.0);  $\mathcal{P}$  is binomial (Bin), or a tail-adjusted Beta (TAB) with rounding parameter  $\delta$ . The TAB distribution is identical to a standard Beta distribution between  $\delta$  and  $1 - \delta$ , but the probability density below  $\delta$  is distributed uniformly over the region  $[0, \delta)$  and the probability density above  $1 - \delta$  is distributed uniformly over the region  $(1 - \delta, 1]$ . This tail-adjustment allows the distribution to model data that range from 0 to 1 (inclusive). A zero-inflated version of this is ZITAB.
- *presence/absence* data—binary (0,1);  $\mathcal{P}$  is Bernoulli.

## Zero-inflation

We wish to cater explicitly for potential zero-inflation, regardless of the data type. Considering responses along a single gradient ( $X$ ), we expect that a species might well be absent from many samples, even at its optimal position ( $m$ ) along  $X$ , due to variation in the species' response across a host of other unmeasured environmental or biological parameters (e.g., Anderson, 2008). For zero-inflated (ZI) models, the probability of an excess zero ( $\pi$ ) is an additional parameter in  $\theta_E$ . For example, if the error distribution were (say) ZINB, then we would have:

$$P(Y_i = y_i \mid \theta_M, \theta_E, x_i) \sim \text{ZINB}(\mu_i, \pi, \phi) \\ \sim \begin{cases} \text{NB}(\mu_i, \phi) & \text{with probability } (1 - \pi) \\ 0 & \text{with probability } \pi \end{cases}$$

Note that for all zero-inflated models in this *senlm* framework, we shall use  $\mu_i$  to denote the mean of that portion of the specified error distribution that does *not* include excess zeros. In the above example,  $\mu_i$  is the mean of the NB distribution within the ZINB mixture. For error distributions that do not have excess zeros (implicitly,  $\pi = 0$ ), the expected value of the response variable is  $E(Y_i) = \mu_i$ . However, for zero-inflated models,  $E(Y_i) = (1 - \pi)\mu_i$ . Although here we shall maintain our focus on  $\mu_i$ , one may

clearly choose to focus instead on  $E(Y_i)$  in different inferential contexts.

## Linking the probability of an excess zero to the mean

We acknowledge the well-known occupancy–abundance relationship in ecology (e.g., Borregaard & Rahbek, 2010; Brown, 1984). Consequently, we can expect more zeros to occur where mean abundances are low. Thus, the degree of zero-inflation generally is expected to increase with decreases in mean abundance (Nielsen et al., 2005; Sileshi et al., 2009; Smith et al., 2012). To accommodate this parsimoniously, one may allow  $\pi$  to vary along the gradient by linking it directly to the mean (Lambert, 1992; Smith et al., 2012): that is,  $\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \gamma_0 + \gamma_1 \log(\mu_i)$ , resulting in *linked zero-inflated models*. Note that we expect the parameter,  $\gamma_1$ , to be negative.

We shall denote these linked models by adding an ‘L’ (for ‘linked’) to each of the potential zero-inflated distributions considered here: *viz.* ZIPL, ZINBL, ZIGL or ZIIGL. In some cases, a better-fitting model may be achieved by linking the zero-inflation parameter to the mean directly, rather than to the log of the mean; that is,  $\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \gamma_0 + \gamma_1 \mu_i$ . Such models are denoted here by adding ‘.mu’ to the acronym (e.g., ZINBL.mu). One might also consider allowing dispersion parameters to vary along the gradient (e.g., McArdle & Anderson, 2004), but this refinement is not pursued here. Further details and mathematical descriptions of these error distributions are given in *senlm* R package documentation (see ‘Implementation’ below).

## IMPLEMENTATION

We provide an R package, *senlm* (‘species–environment non-linear models’), available at the following GitHub repository: <https://primer-e.github.io/senlm/>, to implement the model framework. The end-user can: (i) fit any one or more chosen mean functions with any one or more chosen error distributions to a given set of data ( $Y, X$ ), in any (or all possible) combinations; (ii) estimate parameters using maximum likelihood (ML) and calculate information criteria (e.g., AIC, AICc or BIC) to aid in choosing among competing potential models; and (iii) produce graphics to visualise species' responses along gradients.

For example, once the *senlm* package has been installed and loaded (`library(senlm)`), along with data (e.g., suppose one has a data frame called ‘my.data’ containing a gradient variable called ‘env.gradient’ and a response variable called ‘my.species’), then fitting a given model using the `senlm()` function is straightforward; the user simply provides the data and identifies the

mean function and the error distribution desired for the model fit, as follows:

```
fit <- senlm(data = my.data, xvar =
"env.gradient", yvar = "my.species",
mean_fun = "sech", err_dist = "zinbl")
```

Standard errors on ML parameter estimates may be obtained in the usual way *via* Fisher's information matrix (McCullagh & Nelder, 1983; McCulloch & Searle, 2001) or *via* a jack-knife or bootstrap (Chernick, 2008; Davison & Hinkley, 1997). Although we stayed within a classical ML framework here, a Bayesian implementation (Gelman et al., 2013), with specification of suitable priors on parameters, is also possible (e.g., McElreath, 2020). We provide here a vignette with example code (Rcode\_S1.txt) and data (Data\_S1.csv) to demonstrate use of the *senlm* R package (see 'Vignette\_S1.pdf' in Supporting Information).

We chose to create a tailored *senlm* R package to focus on specific non-linear functions of interest for modelling species–environment relationships, and to couple these with error distributions that are most suitable for ecological field data. Note that the identity link is our preference for these models and, in our experience, works well in practice. The *senlm* R package is an open-source work in progress. Further contributions/improvements are welcome. We recognise that other packages (or combinations of packages) may be used to fit non-linear statistical models (e.g., Bolker et al., 2013), although many require sophisticated coding skills. Our *senlm* package is intended to simplify and enhance ecologists' toolkit for obtaining direct parametric models of species–environment non-linear relationships for data types commonly encountered in broad-scale macro-ecology.

## CAVEATS AND PITFALLS

As with many optimisation problems, the choice of optimiser and initial values can affect resulting parameter estimates. The *senlm* package takes a practical empirical approach—initial values for parameters in  $\theta_M$  are estimated using the method of moments from splines. Initial values for other parameters depend on the model being fitted; further details can be found in the section entitled 'Initial Parameter Estimates' (and the 'Init.R' file of the *senlm* code) available on GitHub. For optimisation, the *senlm* package uses simulated annealing *via* the function `mle2()` with arguments `optimiser = "optim"` and `method = "SANN"` in the R package 'bbmle' (Bolker et al., 2021). A quasi-Newton method is subsequently used to further refine these estimates, or in the event of failure (`optimiser = "nllminb"`).

One motivation for developing this framework was to enable estimation of useful parameters for comparative purposes. However, different mean functions (e.g., sech, HOF, beta) generally have different parameters, so to track changes in a parameter (such as  $m$ ) through time or space or across species, it is advisable to stick with a single mean+error combination (e.g., sech + ZINBL) for the full suite of data being analysed.

Finally, a note of caution regarding rare species: the utility of any given *senlm* model will depend on there being some reasonable number of non-zero values in the response data vector. Although the *senlm* R package will provide estimates of parameters even if there is only one non-zero value(!), this clearly represents a case where scepticism is warranted. Simulations suggest that sample sizes of  $n \leq 20$  yield parameter estimates that are highly variable, particularly for asymmetric responses (see 'Comparison with splines' below). It is a topic for future research to identify the percentage (or number) of non-zero values required to construct adequate formal models of a species' response. A practical rule of thumb may lie somewhere between 1% and 5%, depending on the overall sample size. We leave this decision to the experimenter.

## EXAMPLE

### North-east Pacific groundfish species *versus* depth

We demonstrate the utility of the method *via* analyses of North-east Pacific groundfish species along a depth gradient. Data were obtained from the West Coast Groundfish Bottom Trawl (Slope and Shelf Combination) Survey, conducted annually by the National Oceanic and Atmospheric Association (NOAA)'s Northwest Fisheries Science Center (NOAA Fisheries, NWFS/FRAM, 2725 Montlake Blvd. East, Seattle, WA 98112) and available online (<https://www.nwfsc.noaa.gov/data/map>).

We focused on count data obtained from 1999 to 2004, inclusive. We removed trawls that were not identified as 'satisfactory' in the database, and also removed trawls whose swept area was more extreme than the inter-quartile range (1.682–2.025 ha) to ensure commensurability. This yielded a total of 1280 trawls, from which 239 fish taxa were identified to species. Many species were too rare to permit formal modelling (69 species occurred in only one or two trawls). We fit individual species-specific *senlm* models to a sub-set of 137 species that occurred in at least 7 trawls (i.e., present in  $\geq 0.5\%$  of the 1280 trawls). For each of these 137 species, we fit all combinations of six potential mean functions: {uniform, Gaussian, mixed Gaussian, beta, sech, HOF} and eight potential error distributions for count data: {Poisson,

NB, ZIP, ZINB, ZIPL, ZINBL, ZIPL.mu, ZINBL.mu}. The best model for each species (out of 48 possible models) was then identified using AICc.

More than 91% of the fish species (125 of 137) required some type of non-Gaussian shape to characterise their mean response along the depth gradient, and either a *sech* or mixed Gaussian model were found to be most suitable (lowest AICc) for the majority (60%) of species examined (i.e., 82 of 137; Table 2). Distributions of  $\Delta$ AICc values from models fitted using different *senlm* mean functions (with the error distribution from the best AICc model) indicated that either the *sech* or HOF mean functions would likely be suitable for modelling most of these species (Figure S1). More than 85% of the species (117 of 137) had lower AICc values for models with NB (rather than Poisson) error distributions. In addition, the AICc best model for 73% of the species (101 of 137) included zero-inflation, and 90% of these (91 of 101) indicated that linkage to the mean was most useful to model excess zeros (Table 2).

Patterns in abundances (counts of individuals) per trawl for individual species *versus* depth (in m) showed all of the salient features we aimed specifically to accommodate in our model framework (e.g., Figure 1). Namely, within the sampled gradient's range (i.e., the sampling frame): (i) there is a modal depth, where larger abundances of a given species occur (e.g., *Sebastolobus altivelis*); (ii) in some cases, there is more than one mode (e.g., *Antimora microlepis*); (iii) asymmetry can occur either to the right (*Sebastolobus alascanus*) or left (*Coryphaenoides acrolepis*); (iv) the modal range may be broad (platykurtic, *Alepocephalus tenebrosus*) or narrow (*Lyopsetta exilis*) and (v) there is, in almost all cases, an excess of zeros (Figure 1). These features were all handled capably by *senlm* models (Figure 1; Table S2).

We may focus on a single species (*Sebastolobus alascanus*, Shortspine thornyhead) as an exemplar for

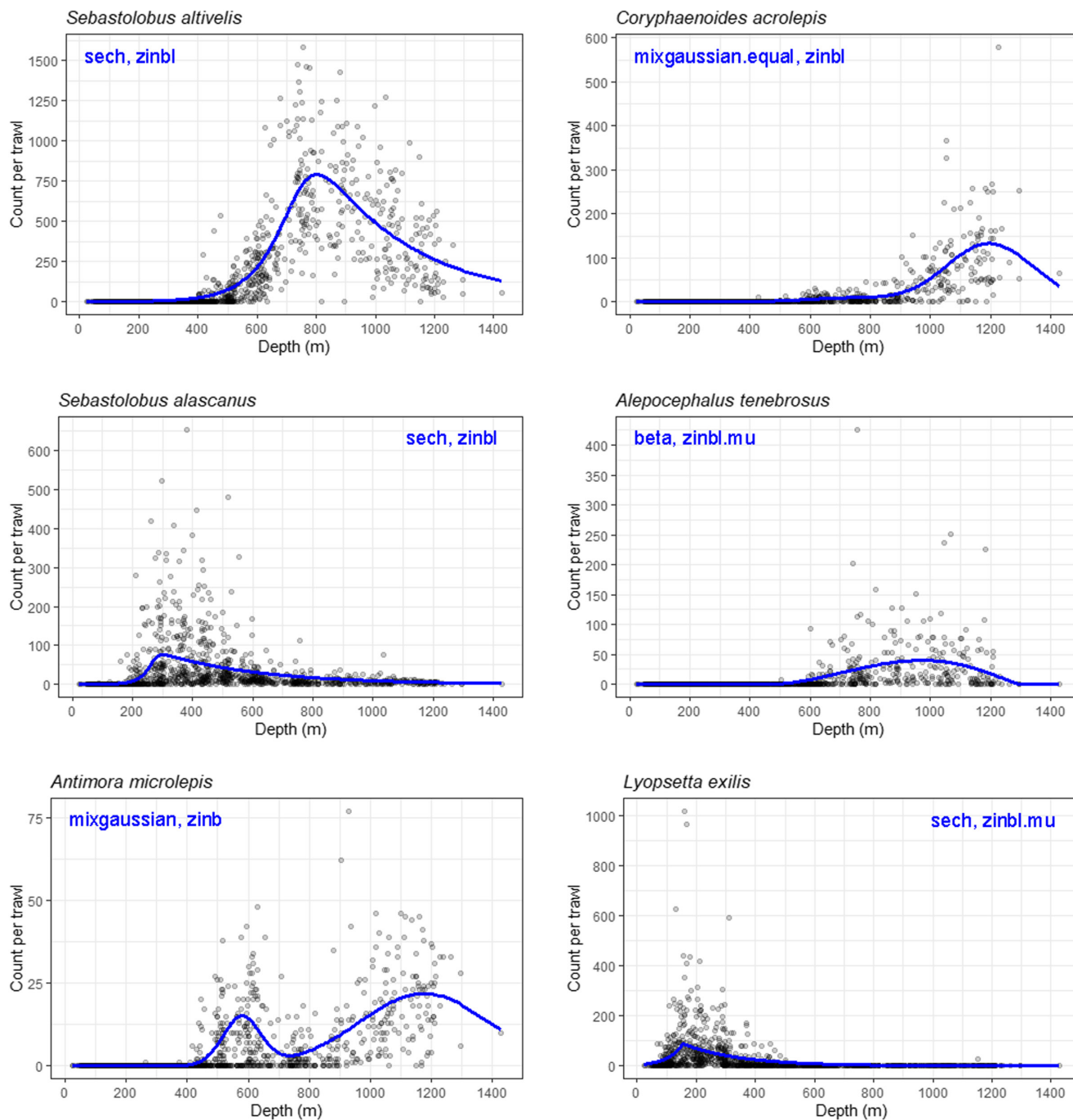
comparative purposes (Figure 2). The *senlm* model (*sech* [ $p = 1$ ] mean function with ZINBL errors, AIC = 8512) provided a better model (lower AIC) than: (i) a quadratic GLM (log link) with either NB errors (AIC = 9233) or ZINB errors (AIC = 9235); (ii) a Gaussian mean function with ZINB errors (AIC = 9235) or (iii) a GAM (P-spline) model with NB errors (AIC = 8538). Distributions of AIC (and  $\Delta$ AIC) values obtained from fitting these four models to each of 50 random jack-knife sub-samples from the original data (where each jack-knife sample contained one-quarter of the data, drawn in a proportional depth-stratified manner to cover the gradient), showed that, for this species, the *sech* and P-spline models consistently out-performed Gaussian or quadratic GLM models (Figure 2d, e), and that the *sech* model (marginally) out-performed the flexible GAM model as well (Figure 2e).

The *senlm* models also outperformed these three other potential models (Gaussian, quadratic GLM and P-spline) for all other fish species shown in Figure 1, based on AIC (Table S3). Essentially, Gaussian models or quadratic GLMs failed to identify correct modal positions along the gradient, while P-splines typically identified modal positions rather well but did not accommodate excess zeros as well as *senlm* models.

Mean functions from *senlm* models fitted to a set of individual species may be drawn in multi-species plots to characterise changes in fish communities along a gradient (Figure 3; Table S4), using either absolute (Figure 3a) or relative mean abundances (Figure 3b). In addition, *senlm* models for a single species can be fitted in a variety of different contexts to explore and quantify potential changes in species' responses to a gradient (and hence, in *senlm* parameters) through space, time, or along some other environmental parameter or factor of interest. For example, constructing separate *senlm* models of *Sebastolobus altivelis*

**TABLE 2** Tallies of the best non-linear model (i.e., having the lowest AICc value), constructed using a given mean function (columns) coupled with a given error distribution (rows), for the responses of each of 137 fish species (counts from 1280 trawls) *versus* a depth gradient in the Northeast Pacific

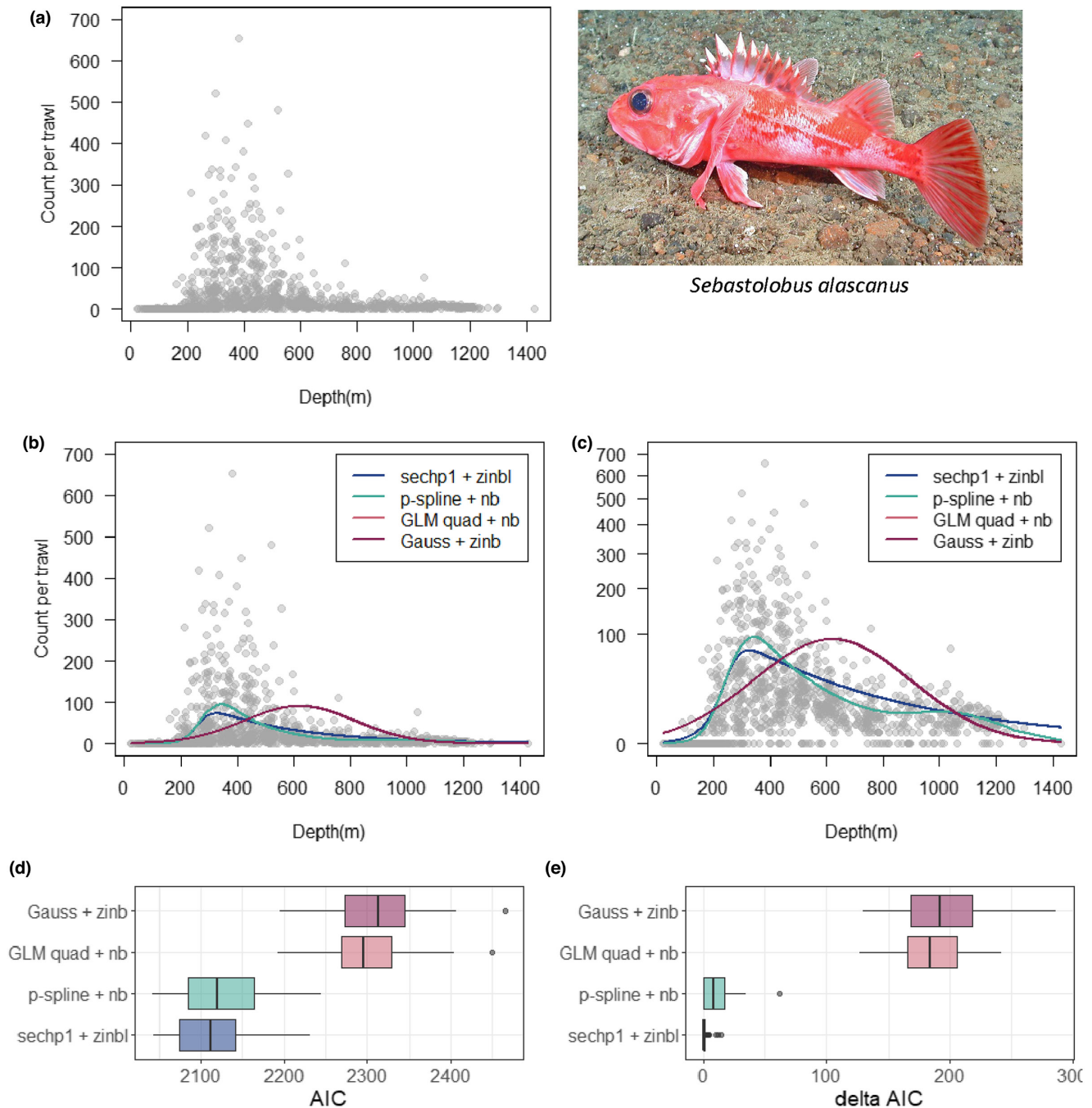
Error Distributions	Mean functions							Total	Percentage
	Sech	Mixed Gaussian	HOF	Uniform	Gaussian	Beta	Total		
ZINBL.mu	25	9	11		5	3	<b>53</b>	38.69%	
NB	2	12	2	12	2		<b>30</b>	21.90%	
ZINBL	10	8	5		3	3	<b>29</b>	21.17%	
Poisson	1	1		3	1		<b>6</b>	4.38%	
ZIPL	2	3	1				<b>6</b>	4.38%	
ZINB	1	2		1	1		<b>5</b>	3.65%	
ZIP		3		2			<b>5</b>	3.65%	
ZIPL.mu		3					<b>3</b>	2.19%	
<b>Total</b>	<b>41</b>	<b>41</b>	<b>19</b>	<b>18</b>	<b>12</b>	<b>6</b>	<b>137</b>		
Percentage	29.93%	29.93%	13.87%	13.14%	8.76%	4.38%			



**FIGURE 1** Scatter plots showing patterns of counts per trawl for each of six species of fish *versus* depth (in metres) from NOAA's North-east Pacific dataset, with the AICc best non-linear model, in each case, shown in blue. The modal nature of observed abundance values along the depth gradient is apparent in all cases, as is a preponderance of zero counts, particularly when mean abundances are low. Note that some of these model curves are asymmetric, bimodal, plateaued or peaked, all fairly typical of species' mean responses to a gradient and for which the suite of general mathematical functions of the *senlm* model framework have been designed to accommodate.

(Longspine thornyhead) *versus* depth at different latitudes (2-degree bins) along the western US coastline (Figure 4a) showed that the greatest estimated values of peak mean abundance per trawl ( $\hat{H}_m \sim 1000$  individuals) occurred between latitudes  $\sim 37\text{--}43^\circ\text{N}$  (Figure 4b), spanning a region of pronounced upwelling (Cape Mendocino; Jacox et al., 2018).

Species' responses to natural gradients may also vary through time, due to climate change. For example, we fit separate *senlm* models of *Sebastes crameri* (Darkblotched rockfish) obtained from the NOAA database (filtered in the same way as previously described), but for the years 1999–2018, inclusive. There has been a significant increase in the estimated position along the

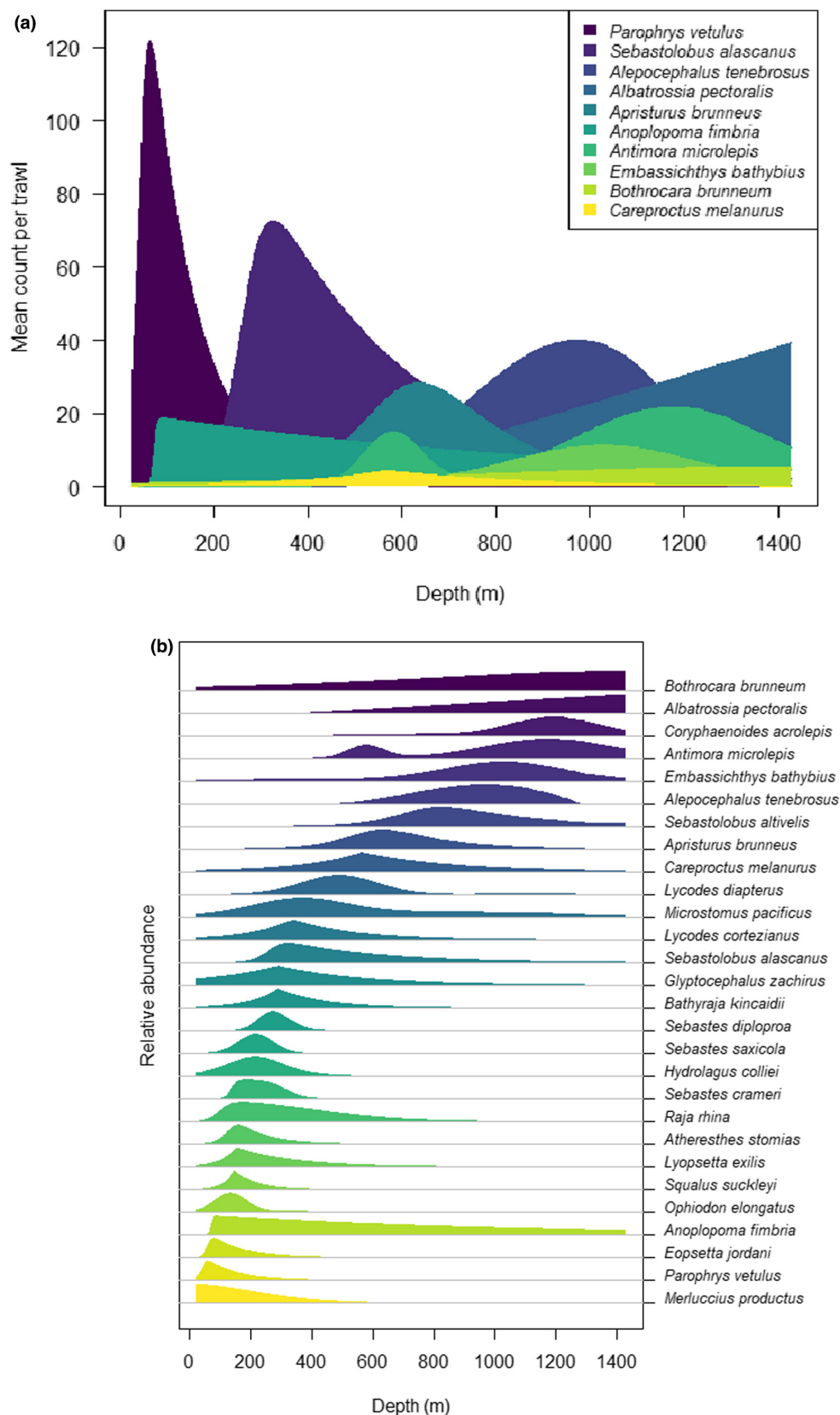


**FIGURE 2** Scatter plot of counts per trawl for *Sebastolobus alascanus* (pictured) versus depth (in metres) shown (a) on its own and also shown with fitted lines corresponding to each of four non-linear models (colours), with the y-axis being either (b) on a raw abundance scale or (c) on a square-root-transformed scale to better see patterns for smaller values. The four models are indicated on each plot as: the *senlm* sech ( $p = 1$ ) mean function with ZINBL errors (blue); a GAM (P-spline) with NB errors (green); a GLM with quadratic mean function (log link) with NB errors (peach) and a Gaussian mean function with ZINB errors (burgundy), as indicated. Note that the quadratic GLM and the Gaussian model are effectively identical here (peach line is not discernible from the burgundy line). Distributions of: (d) AIC values and (e)  $\Delta$ AIC values from fitting each model to 50 jackknife samples of the data. Each jackknife sample was a random proportional depth-stratified sample of one-quarter of the data. (Photo image provided courtesy of Milton Love, University of California, Santa Barbara).

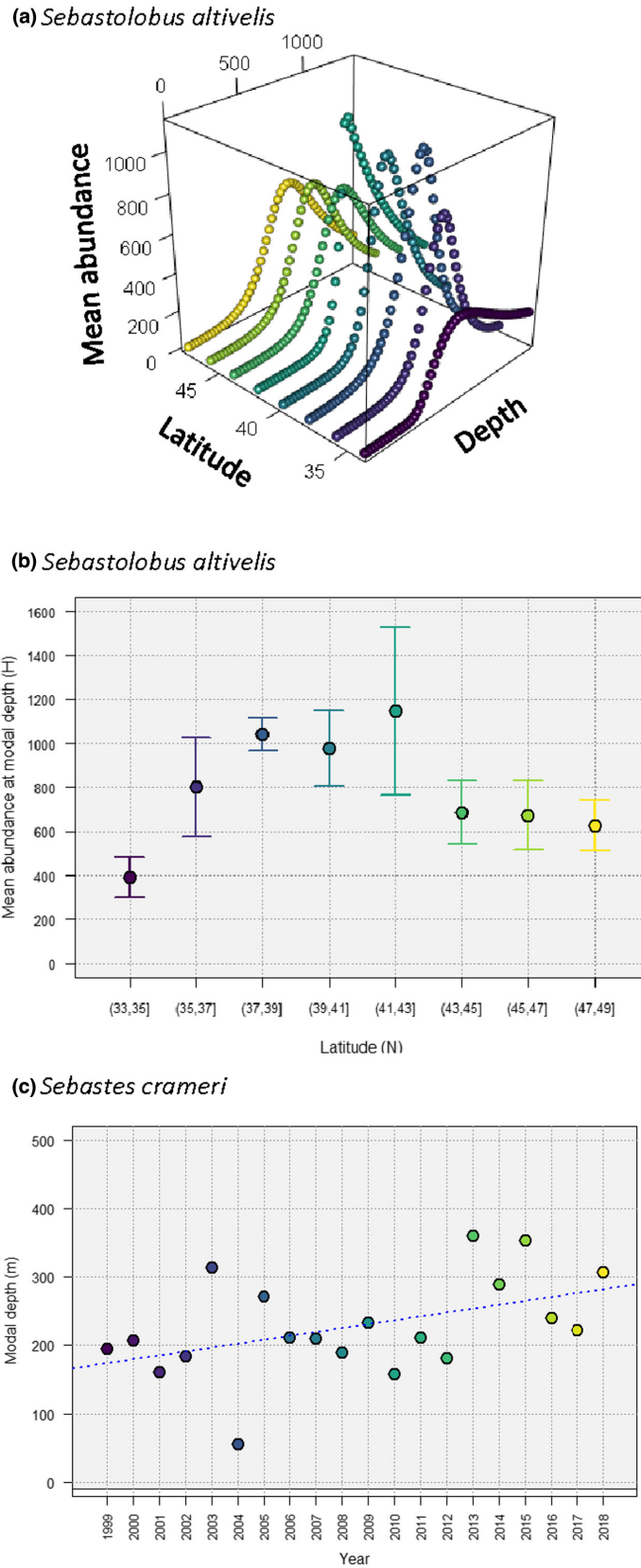
depth gradient at which this species reaches its peak mean abundance ( $m$ , the modal depth) over the past 20 years (Figure 4c,  $F_{1,18} = 6.70$ ,  $p = 0.019$ ). A retreat to greater (deeper) depths, as documented here, may signal a physiological response to warming surface temperatures (e.g., Kingsbury et al., 2020; Rijnsdorp et al., 2009).

## COMPARISON WITH SPLINES

One may fit non-linear shapes to any ( $X$ ,  $Y$ ) data using a variety of flexible empirical methods, such as non-parametric GAMs or P-splines. From the resulting fitted curve one may derive estimates of certain properties of interest, such



**FIGURE 3** (a) Estimated mean abundance per trawl for each of 10 prominent fish species versus depth (in metres) obtained using individual AICc best *senlm* models. The layering of the species response curves (and their associated 'viridis' colour scale) are ordered by decreasing maximum height ( $\hat{H}$ ). (b) Estimated relative mean abundance per trawl (as a fraction of the maximum) for the top 28 most frequently occurring fish species (in at least 20% of the samples) from the NOAA dataset extract (1999–2004) versus depth (in metres) obtained using individual AICc best *senlm* models. Colour and vertical ordering of the species (from top to bottom) is by decreasing modal position of their mean response along the depth gradient.



**FIGURE 4** (a) Mean *senlm* response curve for *Sebastolobus altivelis* versus depth, estimated separately in each of eight 2-degree latitudinal bins along the U. S. west coast using the sech mean function with ZINBL errors; (b) plot of the estimated peak abundance ( $\hat{H} \pm 1$  jack-knife standard error) versus latitude ( $^{\circ}$ N) for the models shown in (a); and (c) relationship between the estimated modal position ( $\hat{m}$ ) along the depth gradient for the mean response of *Sebastes crameri* (modelled using the sech mean function with ZINB errors) versus time (years). The blue dotted line shows the fitted linear regression.

as the maximum fitted value, which may correspond to the mode. However, standardising degrees of freedom (spacing of knots, penalties, etc.) to make comparison across multiple datasets or species may not be straightforward. An extensive comparison of the statistical properties of non-linear curve-fitting methods is beyond the scope of this contribution. However, we performed a modest simulation study to compare the ability of spline-based tools versus our proposed parametric *senlm* models to recover modal characteristics of species under several pertinent scenarios (see ‘Simulation\_study\_S4.pdf’ in Supporting Information). Simulations also permitted us to empirically estimate the coverage of confidence intervals (CIs) built using Fisher's Information matrix for *senlm* models.

First, for symmetric mean response, we found that there was greater variance in the estimated mode obtained using B-splines or P-splines compared to the parametric models (Gaussian or *sech*), even for large sample sizes ( $n \geq 100$ ). Second, when the mean response was asymmetric, both the spline approaches and Gaussian models yielded biased estimates for modal position ( $m$ ) and height ( $H$ )—specifically,  $m$  was dragged towards the larger tail and  $H$  was underestimated, while estimates from *sech* models were unbiased and also had lower variance. Furthermore, both the bias and the variance of  $m$  and  $H$  increased for spline approaches with increases in zero-inflation, while estimates from *sech* models with ZI errors remained unbiased (see ‘Simulation\_study\_S4.pdf’ in Supporting Information for details). Finally, we found empirical coverage of CIs for either  $m$  or  $H$ , built using Fisher's Information matrix in *senlm* models, readily converged to the nominal 95% level for moderate sample sizes (Figure S4.4 in Simulation\_study\_S4.pdf, Supporting Information). Overall, these results suggest that spline-based methods, although clearly more flexible than parametric models, are also more strongly affected by idiosyncrasies of individual datasets.

## DISCUSSION

Despite ample evidence for the prevalence of non-linear (modal) species' responses to environmental (and spatial) gradients (e.g., Austin, 1976, 1980; Oksanen & Minchin, 2002; ter Braak, 1985, 1996; Whittaker, 1956, 1967), there have been surprisingly few attempts to develop a suitably flexible parametric modelling framework to characterise these for abundance data. Although mathematically elegant, the utility of symmetric modal distributions (such as Gaussian curves) for modelling species' responses tends to be the exception, rather than the rule, for most real ecological datasets (e.g., Table 2). Although non-parametric GAMs are useful to help visualise the general non-linear pattern of responses along a gradient (e.g., Yee, 2015; Yee & Mitchell, 1991), they do not yield ecologically interpretable parameters. Furthermore, estimates of modal parameters derived from spline-based models were found to be biased in

simulations where the mean species' response was asymmetric (Supporting Information S4).

We have provided here a suite of non-linear mathematical functions for the mean response of species to environmental gradients that yield interpretable parameters, yet flexibly accommodate asymmetry, peakedness/flatness or bi-modal shapes—features that are empirically readily apparent in natural systems. A non-linear function for the mean response must be coupled with an appropriate error distribution. Importantly, the mean function and the error distribution work together in *senlm* models to capably track the overall shape of the species' response appropriately. Most species display excess zeros (e.g., Table 2). Furthermore, in our example, almost all models requiring zero-inflated errors worked best when the probability of an excess zero was linked to the mean response. This clearly supports the genuine utility of the linked models offered in the *senlm* framework (e.g., ZIPL, ZINBL, ZINBL.mu, ZIGL, etc.).

The suite of error distributions offered by the *senlm* package caters to a large array of data types commonly used in ecology under a single umbrella. We have also tailored certain known statistical distributions (e.g., beta) to accommodate ecological applications better (e.g., ZITAB). In practical terms, using the *senlm* R package, the end-user may simply identify the data type (e.g., discrete, continuous, cover, etc., see Table 1), then information criteria can be used to choose among competing potential models for that data type.

We found for the example dataset that three of the available mean functions (*sech*, mixed Gaussian and HOF) modelled the majority of species rather well. Some species were better modelled, however, using a beta mean function, usually when the mean response was fairly flat, or when there were hard upper/lower bounds on a species' occurrence along the gradient. In contrast, the *sech* function did well when the mean response was more peaked. Development of a single mean function that can be adjusted parsimoniously to model the majority of mean response shapes is clearly desirable.

We consider the *senlm* model framework presented here to be merely the beginning—it provides a core parametric approach for modelling one species along one environmental gradient—upon which more complex models may be built. Future developments might provide for: (i) inclusion of additional factors, random effects or interactions; (ii) non-linear models of species' responses along two or more gradients ( $X$  variables) simultaneously (a species' realised environmental niche); (iii) multivariate non-linear models of two or more species ( $Y$  variables) simultaneously; (iv) methods to integrate errors in parameters (including autocorrelation) through space and time, perhaps using an hierarchical Bayesian approach. Associations among species might also be modelled formally (e.g., Anderson et al., 2019; Ovaskainen & Abrego, 2020; Warton et al., 2015); positive (or negative) relationships could arise from species having similar (or dissimilar) fitted responses to the gradient,

while other types of inter-species associations may be evident across their residuals.

We invite contributions to the *senlm* package to enhance its utility across a broader range of applications. For example, with increasing interest in climate change, non-linear curves explicitly identifying species' tolerances to temperature (including 'tipping points', e.g., see Kenek et al., 2011) would be welcome. Other error distributions could also be added, such as generalised Poisson distributions (Aitchison & Ho, 1989; Clarke et al., 2006; Coly et al., 2016), truncated distributions (e.g., Nadarajah & Kotz, 2006), distributions catering to under-dispersion (Rogers, 1974) or hurdle models (Mullahy, 1986; Zeileis et al., 2008). Extensions to model functional traits along gradients are also desirable.

We anticipate that the *senlm* framework will empower ecologists to construct bespoke non-linear models for species of every conceivable type (e.g., plankton, invertebrates, microbes, trees, fish, birds, mammals, etc.). Our aim is for large datasets covering a variety of places, times and taxonomic groups to be tackled with confidence, regardless of how abundance has been quantified (e.g., biomass, counts, cover, presence/absence, etc.), allowing rigorous quantification of individual species–environment relationships.

#### AUTHOR CONTRIBUTIONS

MJA developed all of the conceptual ideas for the model framework, led the research program, wrote and ran code for all simulations and wrote the manuscript and supplements; DCIW developed the new parameterisation of the beta distribution and wrote R code for the *senlm* package; WLS developed the new parameterisation of the sech function; AP wrangled data, developed the R package with DCIW, developed code and ran analyses to create Table 2 and helped create some of the Figures. All authors contributed refinements to the final draft of the manuscript.

#### ACKNOWLEDGEMENTS

This research was supported by a Royal Society of New Zealand Marsden Grant (19-MAU-145) and a Strategic Science Investment Fund, administered by the Ministry of Business Innovation and Employment (MBIE), Aotearoa/New Zealand. This work was also supported by PRIMER-e (Quest Research Limited). The authors also thank the New Zealand Institute for Advanced Study (NZIAS), particularly Distinguished Professor P. Schwerdtfeger and the Centre for Theoretical Chemistry and Physics (CTCP) at Massey University, for letting us use the CTCP super-computer for our simulation work. Open access publishing facilitated by Massey University, as part of the Wiley - Massey University agreement via the Council of Australian University Librarians.

#### PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/ele.14121>.

#### DATA AVAILABILITY STATEMENT

Data used in examples are available on Dryad at DOI <https://doi.org/10.5061/dryad.c59zw3rbp>. The R package is provided on GitHub at: <https://primer-e.github.io/senlm/>.

#### ORCID

Marti J. Anderson  <https://orcid.org/0000-0002-4018-4049>

#### REFERENCES

- Aitchison, J. & Ho, C.H. (1989) The multivariate Poisson-log normal distribution. *Biometrika*, 76, 643–653. <https://doi.org/10.1093/biomet/76.4.643>
- Anderson, M.J. (2008) Animal-sediment relationships re-visited: characterising species' distributions along an environmental gradient using canonical analysis and quantile regression splines. *Journal of Experimental Marine Biology and Ecology*, 366, 16–27.
- Anderson, M.J., de Valpine, P., Punnett, A. & Miller, A.E. (2019) A pathway for multivariate analysis of ecological communities using copulas. *Ecology and Evolution*, 9(6), 3276–3294. <https://doi.org/10.1002/ece3.4948>
- Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22, 42–47.
- Austin, M.P. (1976) On non-linear species response models in ordination. *Vegetatio*, 33, 33–41.
- Austin, M.P. (1980) Searching for a model for use in vegetation analysis. *Vegetatio*, 42, 11–21.
- Austin, M.P., Cunningham, R.B. & Fleming, P.M. (1984) New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio*, 55, 11–27.
- Bolker, B.M., Gardner, B., Maunder, M., Berg, C.W., Brooks, M., Comita, L. et al. (2013) Strategies for fitting nonlinear ecological models in R, AD Model Builder, and BUGS. *Methods in Ecology and Evolution*, 4, 501–512.
- Bolker, B.M., R Development Core Team, Giné-Vázquez, I. (2021). *bbmle: Tools for general maximum likelihood estimation*. R package version 1.0.24.
- Borregaard, M.K. & Rahbek, C. (2010) Causality of the relationship between geographic distribution and species abundance. *The Quarterly Review of Biology*, 85, 3–25.
- Bradshaw, C.J.A., Brook, B.W., Delean, S., Fordham, D.A., Herrando-Pérez, S., Cassey, P. et al. (2014) Predictors of contraction and expansion of area of occupancy for British birds. *Proceedings of the Royal Society*, 281, 20140744.
- Brown, J.H. (1984) On the relationship between abundance and distribution of species. *The American Naturalist*, 124, 255–279.
- Chapman, M.G. & Underwood, A.J. (2008) Scales of spatial variation of gastropod densities over multiple spatial scales: comparison of common and rare species. *Marine Ecology Progress Series*, 354, 147–160.
- Chernick, M.R. (2008) *Bootstrap methods: a guide for practitioners and researchers*, 2nd edition. Hoboken, New Jersey, USA: Wiley.
- Chipman, H.A., George, E.I. & McCulloch, R.E. (2010) BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4, 266–298.
- Clarke, K.R., Chapman, M.G., Somerfield, P.J. & Needham, H.R. (2006) Dispersion-based weighting of species counts in assemblage analyses. *Marine Ecology Progress Series*, 320, 11–27.
- Colwell, R.K. & Rangel, T.F. (2009) Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences of the United States of America*, 106(Suppl2), 19651–19658.
- Coly, S., Yao, A.-F., Abrial, D. & Charras-Carrido, M. (2016) Distributions to model overdispersed count data. *Journal de la Société Française de Statistique*, 157, 39–63.

- Davison, A.C. & Hinkley, D.V. (1997) *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.
- de Boor, C. (2001) *A practical guide to splines*, revised edition. New York: Springer-Verlag.
- Eilers, P.H.C. & Marx, B.D. (2021) *Practical smoothing: the joys of P-splines*. Cambridge, UK: Cambridge University Press.
- Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology and Systematics*, 40, 677–697.
- Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *The Journal of Animal Ecology*, 77, 802–813.
- Fraaije, R.G.A., ter Braak, C.J.F., Verduyn, B., Breeman, L.B.S., Verhoeven, J.T.A. & Soons, M.B. (2015) Early plant recruitment stages set the template for the development of vegetation patterns along a hydrological gradient. *Functional Ecology*, 29, 971–980.
- Gauch, H.G., Chase, G.B. & Whittaker, R.H. (1974) Ordinations of vegetation samples by Gaussian species distributions. *Ecology*, 55, 1382–1390.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2013) *Bayesian data analysis*, 3rd edition. Boca Raton, Florida, USA: Chapman & Hall/CRC Press.
- Gu, C. & Wahba, G. (1991) Minimizing GCV/GML scores with multiple smoothing parameters via the Newton method. *SIAM Journal on Scientific and Statistical Computing*, 12, 383–398.
- Harrison, P.A., Berry, P.M., Butt, N. & New, M. (2006) Modelling climate change impacts on species' distributions at the European scale: implications for conservation policy. *Environmental Science & Policy*, 9, 116–128.
- Hastie, T. & Tibshirani, R. (1990) *Generalized additive models*. London: Chapman and Hall.
- Huisman, J., Olf, H. & Fresco, L.F.M. (1993) A hierarchical set of models for species response analysis. *Journal of Vegetation Science*, 4, 37–46.
- Jacox, M.G., Edwards, C.A., Hazen, E.L. & Bogard, S.J. (2018) Coastal upwelling revisited: Ekman, Bakun, and improved upwelling indices for the U.S. West Coast. *Journal of Geophysical Research: Oceans*, 123, 7332–7350.
- Jamil, T. & ter Braak, C.J.F. (2013) Generalized linear mixed models can detect unimodal species environment relationships. *PeerJ*, 1, e95.
- Jansen, F. & Oksanen, J. (2013) How to model species responses along ecological gradients—Huisman-Olf-Fresco models revisited. *Journal of Vegetation Science*, 24, 1108–1117.
- Johnson, R.W. & Goodall, D.W. (1980) A maximum likelihood approach to non-linear ordination. *Vegetatio*, 41, 133–142.
- Kenek, S., Berendonk, T.U. & Petzoldt, T. (2011) Thermal performance curves of *Paramecium caudatum*: a model selection approach. *European Journal of Protistology*, 47, 124–137.
- Kingsbury, K.M., Gillanders, B.M., Booth, D.J., Coni, E.O.C. & Nagelkerken, I. (2020) Range-extending coral reef fishes trade-off growth for maintenance of body condition in cooler waters. *Science of the Total Environment*, 703, 134598.
- Lambert, D. (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34, 1–14.
- Leathwick, J.R., Rowe, D., Richardson, J., Elith, J. & Hastie, T. (2005) Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology*, 50, 2034–2052.
- Legendre, P. (1993) Real data are messy. *Statistics and Computing*, 3, 197–199.
- Lek, S. & Guégan, J.F. (1999) Artificial neural networks as a tool in ecological modelling: an introduction. *Ecological Modelling*, 120, 64–73.
- Makarek, V. & Legendre, P. (2002) Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. *Ecology*, 83, 1146–1161.
- Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J. et al. (2005) Zero-tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, 8, 235–246.
- McCullagh, B.H. & Anderson, M.J. (2004) Variance heterogeneity, transformations, and models of species abundance: a cautionary tale. *Canadian Journal of Fisheries and Aquatic Sciences*, 61, 1294–1302.
- McCullagh, B.H., Gaston, K.J. & Lawton, J.H. (1990) Variation in the size of animal populations: patterns problems and artifacts. *The Journal of Animal Ecology*, 59, 439–454.
- McCullagh, J.A. & Nelder, J.A. (1983) *Generalized linear models*, 2nd edition. London: Chapman & Hall.
- McCulloch, C.E. & Searle, S.R. (2001) *Generalized, linear and mixed models*. New York: John Wiley & Sons.
- McElreath, R. (2020) *Statistical rethinking: a Bayesian course with examples in R and Stan*, 2nd edition. Boca Raton, Florida, USA: Chapman and Hall/CRC Press.
- Mullahy, J. (1986) Specification and testing of some modified count data models. *Journal of Econometrics*, 33, 341–365.
- Nadarajah, S. & Kotz, S. (2006) R programs for computing truncated distributions. *J Statist Softw*, 16 Code Snippet 2, 1–8.
- Nielsen, S.E., Johnson, C.J., Heard, D.C. & Boyce, M.S. (2005) Can models of presence-absence be used to scale abundance? Two case studies considering extremes in life history. *Ecography*, 28, 197–208.
- Oksanen, J. & Minchin, P.R. (2002) Continuum theory revisited: what shape are species responses along ecological gradients? *Ecological Modelling*, 157, 119–129.
- Ovaskainen, O. & Abrego, N. (2020) *Joint species distribution modelling with applications in R*. Cambridge, UK: Cambridge University Press.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, 190, 231–259.
- Rigby, R.A. & Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, 54, Part 3, 507–554.
- Rijnsdorp, A.D., Peck, M.A., Engelhard, G.H., Möllmann, C. & Pinnegar, J.K. (2009) Resolving the effect of climate change on fish populations. *ICES Journal of Marine Science*, 66, 1570–1583.
- Rogers, A. (1974) *Statistical analysis of spatial dispersion: The quadrat method*. London, UK: Pion Limited.
- Sileshi, G., Hailu, G. & Nyadzi, G.I. (2009) Traditional occupancy-abundance models are inadequate for zero-inflated ecological count data. *Ecological Modelling*, 220, 1764–1775.
- Smith, A.N.H., Anderson, M.J. & Millar, R.B. (2012) Incorporating the intraspecific occupancy-abundance relationship into zero-inflated models. *Ecology*, 93, 2526–2532.
- Stoklosa, J. & Warton, D.I. (2018) A generalized estimating equation approach to multivariate adaptive regression splines. *Journal of Computational and Graphical Statistics*, 27, 245–253.
- Taylor, L.R. (1961) Aggregation, variance and the mean. *Nature*, 189, 732–735.
- ter Braak, C.J.F. (1985) Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics*, 41, 859–873.
- ter Braak, C.J.F. (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67, 1167–1179.
- ter Braak, C.J.F. (1996) *Unimodal models to relate species to environment*. DLO-Agricultural Mathematics Group, Wageningen, the Netherlands.
- ter Braak, C.J.F. & Looman, C.W.N. (1986) Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio*, 65, 3–11.
- Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. et al. (2015) So many variables: joint modeling in community ecology. *Trends in Ecology & Evolution*, 30, 766–779.

- Westman, W.E. (1980) Gaussian analysis: identifying environmental factors influencing bell-shaped species distributions. *Ecology*, 61, 733–739.
- Whittaker, R.H. (1956) Vegetation of the Great Smoky Mountains. *Ecological Monographs*, 26, 1–80.
- Whittaker, R.H. (1967) Gradient analysis of vegetation. *Biological Reviews of the Cambridge Philosophical*, 49, 207–264.
- Whittaker, R.H., Levin, S.A. & Root, R.B. (1973) Niche, habitat, and ecotope. *The American Naturalist*, 107, 321–338.
- Yee, T.W. (2004) A new technique for maximum-likelihood canonical Gaussian ordination. *Ecological Monographs*, 74, 685–701.
- Yee, T.W. (2015) *Vector generalized linear and additive models: with an implementation in R*. New York, USA: Springer Series in Statistics.
- Yee, T.W. & Mitchell, N.D. (1991) Generalized additive models in plant ecology. *Journal of Vegetation Science*, 2, 587–602.
- Zeileis, A., Kleiber, C. & Jackman, S. (2008) Regression models for count data in R. *Journal of Statistical Software*, 27(8), 1–25. <https://doi.org/10.18637/jss.v027.i08>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Anderson, M.J., Walsh, D.C.I., Sweatman, W.L. & Punnett, A.J. (2022) Non-linear models of species' responses to environmental and spatial gradients. *Ecology Letters*, 25, 2739–2752. Available from: <https://doi.org/10.1111/ele.14121>