Bayesian Distributions of Species Abundance along Environmental Gradients

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science
in
Statistics

at Massey University, Albany,
New Zealand.

Hayden Daniel Rabel

2023

Abstract

Understanding the relationship between species abundance and environmental conditions is crucial for conservation and management efforts. This thesis presents a novel approach for predicting distributions of species abundance along environmental gradients (DAEG) by refining the parameter space of a nonlinear zero-inflated negative binomial with modskurt mean (NZM) model and utilising Bayesian prior probability. Chapter 2 elucidates the NZM model, highlighting the challenges posed by the intricate nature of parameter estimation due to the model's complexity. A refined parameter subspace is proposed to address the issue of multimodality in the likelihood surface, enhancing the reliability of predicting DAEGs. Chapter 3 employs Bayesian inference and proposes a prior distribution for the NZM model that increases the ecological structure used in the parameter estimation process and improves the reliability of making realistic predictions. A step-by-step workflow for using the Bayesian implementation is presented and demonstrated with a case study. The thesis includes as a supplement an R package and interactive resources (Appendix A; `https://hdrab127.github.io/modskurt/`) that enable straightforward fitting of DAEGs using Bayesian NZM models. This work contributes to more accurate predictions of DAEGs, provides a practical tool for ecological research, and promotes effective conservation and management efforts.

Acknowledgements

I am profoundly grateful to Massey University and PRIMER-e for the scholarships that were instrumental in facilitating my research.

My deepest thanks go to my primary supervisor, Dr Adam Smith, and co-supervisor Distinguished Professor Marti Anderson, for your tireless support and guidance throughout my somewhat turbulent research journey. You have taught me so much more than I ever could have hoped in this master's and included me in really cool research projects without hesitation. I had such a wonderful time studying with you, and I am truly grateful; thank you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Species abundance counts refer to the number of individuals per species within an ecological community, a biogeographical area, or other space (Rosenzweig 1995). Understanding the complex relationship between species abundance and the multifaceted environmental conditions they encounter is at the heart of modern ecology (Clark 2007; Chase and Leibold 2009; Vellend 2016) and pivotal in conservation and management decisions (Gaston and Blackburn 2000; Pörtner et al. 2022; Hernández-Blanco et al. 2022). As environmental shifts reshape our world, we must rely on sophisticated models that accurately capture and predict these complex dynamics (Pearman et al. 2008; M. J. Anderson et al. 2022).

Various factors can influence a species' abundance in a particular area. These can include the availability of resources (food, water, and shelter) (Tilman 1982), competition with other species (Connell 1961), predation (Paine 1966), disease (R. Anderson and May 1978), and environmental conditions such as temperature, humidity, and light (Woodward 1987). Changes in environmental conditions can lead to shifts in species abundance (Parmesan 2006). For example, a temperature rise could benefit some species while harming others, causing their respective abundances to increase or decrease (Deutsch et al. 2008; Martins et al. 2023). Understanding and predicting species abundance in rela-

tion to these environmental conditions is of paramount importance (Elith and Leathwick 2009; Waldock et al. 2022).

Traditional approaches for modelling distributions of species abundance along environmental gradients (DAEG) have proven beneficial in conservation and management efforts (Guisan and Zimmermann 2000; Guillera-Arroita et al. 2015). However, these models often fail to fully capture the complexity of ecological data and patterns (Clark 2007; M. J. Anderson et al. 2022). A promising alternative is a new negative binomial regression model with an ecologically backed nonlinear predictor for mean abundance and zero-inflation probability (A. Smith et al. 2012; M. J. Anderson et al. 2022; Martins et al. 2023). This nonlinear zero-inflated negative binomial with modskurt mean (NZM) model has shown potential in effectively learning and predicting nonlinear patterns of realistic abundance count distributions (Martins et al. 2023). However, the NZM model's complexity poses significant challenges, especially regarding the reliability of parameter estimation techniques (Robert et al. 1999; Raue et al. 2009). Incorrect or imprecise parameter estimation can, in turn, lead to less accurate and less reliable predictions of DAEGs (Kéry and Royle 2020). For an ecological model to be useful, its predictions must be accurate and reliable, meaning they can be trusted to hold true across various conditions and data sets (Gelman, Carlin, et al. 2013; Waldock et al. 2022). The challenge, therefore, lies in refining the parameter space and improving parameter estimation for the NZM model to enhance the accuracy and reliability of its predictions.

1.2 Aims and objectives

The primary aim of this research is to enhance the reliability and credibility of the NZM model's predictions of DAEGs by refining the model's parameter space and incorporating Bayesian inference. The research objectives that will guide this work are as follows:

1. To describe the mathematical form of the NZM model, explore the influence of the NZM model's parameter values.

2. To refine the parameter space for the NZM model to reduce unnecessary complexity and focus the model upon a more ecologically credible set of values. This refined parameter space will enhance the reliability of the model's predictions.

3. To develop a Bayesian implementation of the NZM model in which prior probability distributions are used to encode the refined parameter space of objective 2. This Bayesian approach will increase the ecological structure in the parameter estimation process and thus allow more specific model flexibility.

4. To develop a detailed workflow and guide to applying the Bayesian NZM model and online resources for the specification and assessment of prior distributions based on the specific species-environment relationship being studied.

5. To provide a well-documented R package that enables straightforward fitting of DAEGs using Bayesian NZM models.

Through achieving these objectives, this research will provide valuable insights and practical tools for ecological researchers and conservation practitioners, enabling them to understand better and predict species abundance patterns along environmental gradients using the NZM model.

1.3 Thesis structure

Chapter 2 delves into the nonlinear zero-inflated negative binomial with modskurt mean (NZM) model, a powerful tool for predicting distributions of species abundance along environmental gradients (DAEG). The complexity of the NZM model is explored, specifically in relation to parameter estimation, focusing on how different parameter values impact the model's predictions. The chapter culminates in a proposed refined parameter space to enhance the reliability of the model's predictions.

Chapter 3 builds on the foundations laid in Chapter 2; this chapter introduces the Bayesian approach to the NZM model. The concept of specifying prior probability

distributions for parameters is discussed, leading to a refined model with an increased ecological structure. The chapter includes a practical guide to implementing the Bayesian NZM model, featuring a detailed workflow and a real-world case study.

The final Chapter 4 synthesises the insights gained throughout the research process. It provides a summary of key findings, implications for the field of ecology and species conservation, and directions for future research.

The thesis appendix provides supplementary material for Chapter 3. These materials include an open-source R package (R Core Team 2023) to estimate DAEG using the Bayesian NZM model. This supplement is available on GitHub and has accompanying documentation and additional resources on a dedicated website (Appendix A; `https://hdrab127.github.io/modskurt/`). This appendix also links to interactive graphs and tables to assist with prior specification and predictive checks for the distributions of species abundance and the modskurt function that describes how average species abundance changes along an environmental gradient. These resources will be helpful for readers who wish to understand the application of the NZM model further and present much of the original contributions of this thesis to the field of species abundance modelling.

Chapter 2

Abundance-environment distributions

2.1 Overview

Understanding how species abundance and environmental conditions are related is essential for conservation and management (Gaston and Blackburn 2000; Pörtner et al. 2022; Hernández-Blanco et al. 2022). Modelling distributions of abundance along environmental gradients (DAEG) can help to identify patterns in these relationships (M. J. Anderson et al. 2022; Martins et al. 2023). The nonlinear zero-inflated negative binomial with modskurt mean function (NZM) model shows promise for learning and predicting these patterns in distributions of abundance count data. However, the NZM model complexity can create parameter identifiability issues that hinder the reliability of parameter estimation techniques (Robert et al. 1999; Raue et al. 2009).

The aim of this chapter is to show how a refined parameter space could improve the reliability of predicting realistic DAEGs by reducing unnecessary complexity. First, this aim will be achieved by describing the NZM model and the effect of different parameter values. Then second, redundancy in the parameter space is explored, and a refined parameter subspace of credible values is proposed to improve model reliability by increasing the functional separation of parameters. The chapter concludes by discussing implementation options for this proposed parameter subspace.

2.2 The NZM model

Distributions of species abundance count data are well known to exhibit overdispersion (variance in excess of what mean abundance would suggest) (Clapham 1936; Stoklosa et al. 2022) and zero-inflation (zero counts in excess of what overdispersion would suggest) with excess-zero probabilities related to mean abundance (e.g. fewer zeros in areas of higher abundance) (Brown 1984; A. Smith et al. 2012). The zero-inflated negative binomial (NB) distribution with excess-zero probability linked to the mean (ZINBL) of M. J. Anderson et al. (2022) models these properties by supposing that each n^{th} observation of species abundance count, y_n, given gradient value, x_n, is distributed by

$$y_n|x_n \sim \text{ZINBL}\left[\mu_n, \phi, \pi_n\right] \tag{2.1}$$

$$= \pi_n \cdot \mathbb{1}_0(y_n) + (1 - \pi_n) \cdot \text{NB}(\mu_n, \phi) \tag{2.2}$$

$$\text{logit } \pi_n = \gamma_0 - \gamma_1 \cdot \mu_n, \tag{2.3}$$

where the indicator function on line 2, $\mathbb{1}_0(y_n)$ (equal to 1 when $y_n = 0$ and 0 otherwise), inflates the probability of observing a zero count by $\pi_n \in [0, 1]$. This inflated proportion of zeros is predicted by a decreasing logistic function of the mean of the NB distribution, $\mu_n \in \mathbb{R}_{>0}$, with log-odds intercept and slope, $\gamma_0 \in \mathbb{R}$ and $\gamma_1 \in \mathbb{R}_{\geq 0}$. The NB mean abundance. The NB reciprocal dispersion parameter, $\phi \in \mathbb{R}_{>0}$, scales the variation of y_n in relation to μ_n. See Appendix A.2 for an interactive visualisation of how the ZINBL distribution parameters work together to model species abundance.

The relationship between these abundance data and a single environmental gradient is theoretically nonlinear and unimodal (Whittaker 1956; Gauch et al. 1974; Goodall and Johnson 1982; Ter Braak 1987; McGill and Collins 2003; Jamil and Ter Braak 2013). The reproductive success of each species is thought to be optimised for a specific set of environmental conditions, and thus, this set of conditions usually accommodates the highest levels of abundance on average. In contrast, conditions less optimal for the species' reproductive success usually correlate with lower levels of abundance. The pre-

cise shape of this relationship has a long history of debate. It is generally agreed upon that the myriad factors interacting in the species' ecosystem can distort the relationship away from what physiology-based theory would suggest; however, the extent of this distortion is still an active area of research (M. J. Anderson et al. 2022).

Martins et al. (2023) propose a modskurt function to describe the nonlinear relationship between species abundance and an environmental gradient. The modskurt function is appealing in that it is flexible to allow for "some" of the distortions mentioned above yet retains enough ecological structure to avoid issues of overfitting or predicting unrealistic results that commonly plague nonparametric methods like GAMs and Boosted Regression Trees (Hastie et al. 2009). The modskurt function is defined over the environmental variable, x, as

$$\begin{aligned} \text{modskurt}(x_n) = H \Bigg[& r \exp\left(\frac{1}{p}\left[\frac{x_n - m}{sr} - d\right]\right) + \\ & (1-r)\exp\left(\frac{1}{p}\left[d - \frac{x_n - m}{s\,(1-r)}\right]\right) - \\ & \exp\left(-\frac{d}{p}\right) + 1 \Bigg]^{-p}, \end{aligned} \tag{2.4}$$

with shape parameters,

- $H \in \mathbb{R}_{>0}$: the peak mean abundance along the environmental gradient (Height of the curve).
- $m \in \mathbb{R}$: the environmental gradient value at peak mean abundance (modal abundance).
- $s \in \mathbb{R}_{>0}$: that scales the curve's width, describing how abundance decreases away from m.
- $r \in (0, 1)$: that allows for asymmetric relationships, with $r = 0.5$ indicating equal abundance in environmental conditions either side of m, r below 0.5 indicating more abundance below m, and r above 0.5 indicating more abundance above m.

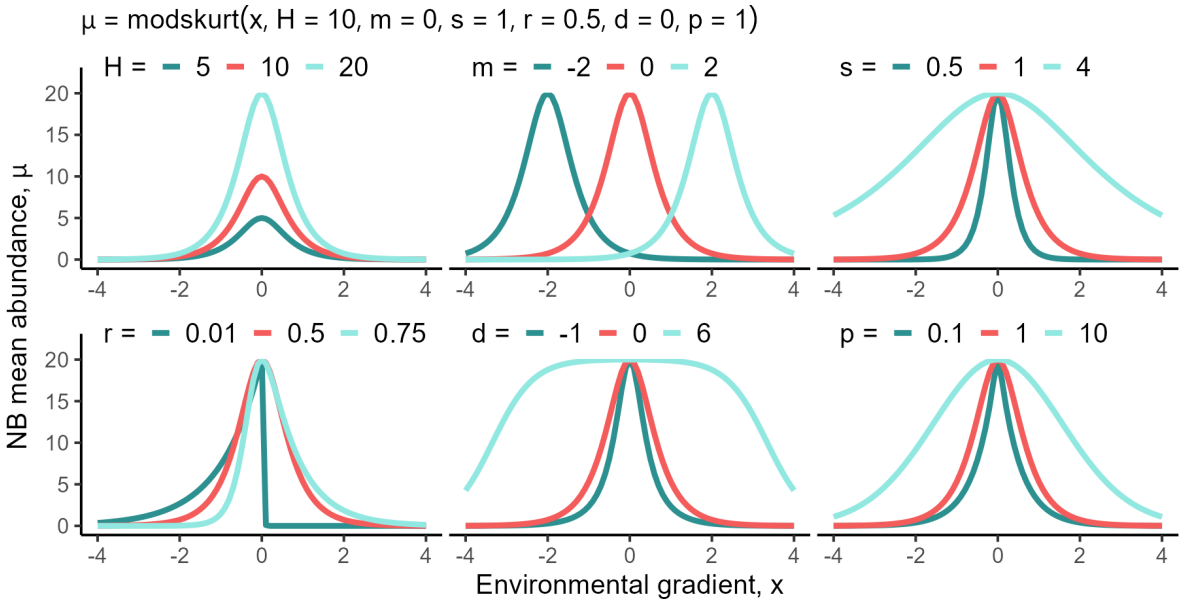

Figure 2.1: Mean abundance shapes produced by the modskurt mean function given a gradient of x and varying parameter values. The top row shows the effects that varying H, m, and s parameter values have on a bell-shaped modskurt curve ($m = 0, s = 1, r = 0.5, d = 0, p = 1$), and the bottom rows does the same for the r, d, and p parameters.

- $d \in \mathbb{R}$: that controls the peakedness ($d < 0$) or broadness ($d > 0$, e.g. plateau-like shapes) of the peak mean abundance around m.
- $p \in \mathbb{R}_{>0}$: that persists higher levels of abundance in relatively extreme conditions; this is similar to increasing the tail weight of a density function.

Figure 2.1 shows how the parameters of this function work together to predict different shapes of mean abundance along an environmental gradient (also see Appendix A.3).

Using the modskurt to predict the NB mean of the ZINBL distribution produces the NZM model for the conditional distribution of y_n given x_n as

$$y_n | x_n \sim \text{ZINBL}\left[\mu_n = \text{modskurt}(x_n), \phi, \pi_n\right]. \tag{2.5}$$

This model presents a powerful tool for predicting species abundance in relation to environmental gradients, enabling us to capture and understand complex ecological phe-

nomena. Its flexibility and robustness make it particularly suited to conservation efforts where understanding and predicting abundance distributions are paramount.

2.3 Refining the parameter space

The NZM (2.5) is appealing for predicting DAEGs over more straightforward methods or those too flexible (M. J. Anderson et al. 2022). However, to achieve such a wide array of shapes and retain good computational properties (e.g. continuous and differentiable), the modskurt mean function has had to sacrifice some parsimony and blur the lines between the functional behaviour that each parameter controls. For example, see how similar the green lines of $s = 0.5$, $d = -1$ and $p = 0.1$ in Figure 2.1 are; these three parameters are designed to capture different relationship patterns but have an interchangeable effect for relatively small values. Is the narrow pointy shape of the relationship caused by small scale (s), lack of broadness (d), or very low persistence to more extreme gradient conditions (p)? Large values of s, d, and p can also create similar shapes of spread and plateaus. Furthermore, another interchangeability issue arises when peak abundance (m) is located near the edge of sampled gradient conditions; this creates asymmetry (i.e. only one side of the modskurt function is visible along the gradient) that could also be captured by r.

Lack of functional separation in the modskurt model can lead to parameter identifiability issues (Raue et al. 2009), where there is no single set of unique parameter values for each possible modskurt shape, and multimodality in parameter estimation (i.e. multiple optima in the likelihood surface that MLE seeks to maximise) causing stability and convergence concerns (Eliason 1993). The model could predict realistic shapes but fail to learn or repeatedly predict the same pattern – as described by the parameter values – or worse, get stuck in the middle and fail to decide on any one pattern at all.

There are also identifiability issues present when considering parameters in isolation. For example, the d parameter that multiplies s to create core width of peak

abundance around m produces an entirely flat line for any value greater than 40. Therefore, if the effect of $d = 40$ and $d = 200$ is the same, parameter estimation could be made less ambiguous by only considering d values less than 50. Similarly, with $s > 6 \cdot \text{range } x$, for conditions sampled, s values that multiply the range of x by more than 6 all produce the same flat lines that large d values do.

Ideally, the parameter space would only contain values that combine to produce unique shapes. There are particular challenges in revising the modskurt functional form or accounting for between-parameter dependencies. However, some quick wins can be achieved by removing redundant regions of the parameter space and, at the least, ensuring each parameter value in the reduced space uniquely affects the modskurt shape in isolation.

Another improvement would be considering a space of values that are "credible" for the DAEG being studied. For example, the m parameter that defines the environmental value associated with peak mean abundance, $H = \max \mu$, is unlikely to be located outside the range of sampled conditions (i.e. there is no abundance data outside of the range of sampled x to suggest the location of H exists out there). Likewise, for H, if the species in question were the mostly solitary polar bear who only congregate to mate (Derocher et al. 2004), it would not be credible to observe a maximum count of individual bears in one spot greater than 10, say, let alone peak mean abundance $H > 10$.

Table 2.1 proposes a refined parameter space for the NZM model (2.5), focusing on values that uniquely affect the shape of the modskurt function in isolation and are credible in the context of the DAEG. The implications of this refined parameter space are illustrated in Figure 2.2 that shows an example of how potentially less credible, but still mathematically optimal (as determined by their log-likelihood value, L), parameter combinations could be avoided. It is a contrived example using fake data but serves as a useful illustration of what can happen when there are fewer data to inform the parameter estimation process and when the linked zero-inflation parameters are considered also.

Table 2.1: Mathematically possible versus ecologically credible values for each parameter in the NZM distribution model (2.5).

Possible	**Credible**	**Brief justification**
$H \in \mathbb{R}_{>0}$	$[0, \max y)$	Peak mean abundance should be less than the max abundance we could observe; in reality, the skewed nature of abundance data would suggest much lower values for H
$m \in \mathbb{R}$	$[\min x, \max x]$	Mode location of x should not be more extreme than what is measurable physically or plausibly (e.g. negative concentrations are impossible, and equator surface sea temperatures below freezing are highly implausible)
$s \in \mathbb{R}_{>0}$	$(0, 6 \cdot \text{range}(x)]$	Spread of abundance values are redundant beyond about $3-6$ times the range of possible x with very similar and mostly uniform flat shapes
$r \in (0, 1)$	$(0, 1)$	Asymmetry of abundance could be anything
$d \in \mathbb{R}$	$[0, 40]$	Broadness of peak multiplies s to create core width, 40 would be uniform for all but $s < 0.05$ and values beyond that redundant
$p \in \mathbb{R}_{>0}$	$[0.05, 2.00]$	Tail persistence values that are tiny cause computational issues, and shapes produced by $p > 2$ can be achieved by s & d and are thus redundant
$\phi \in \mathbb{R}_{>0}$	$(0, \infty)$	Reciprocal dispersion could range from complete, $\phi \to 0$, to Poisson equidispersion, $\phi \to \infty$
$\gamma_0 \in \mathbb{R}$	$[-7, 7]$	Log-odds of an excess-zero intercept values ± 7 would correspond to $[0.001, 0.999]$ in probability space
$\gamma_1 \in \mathbb{R}_{>0}$	$[0, 7]$	Rate of decrease in log-odds of an excess-zero of 7 would be almost no zero-inflation beyond $\mu = 2$

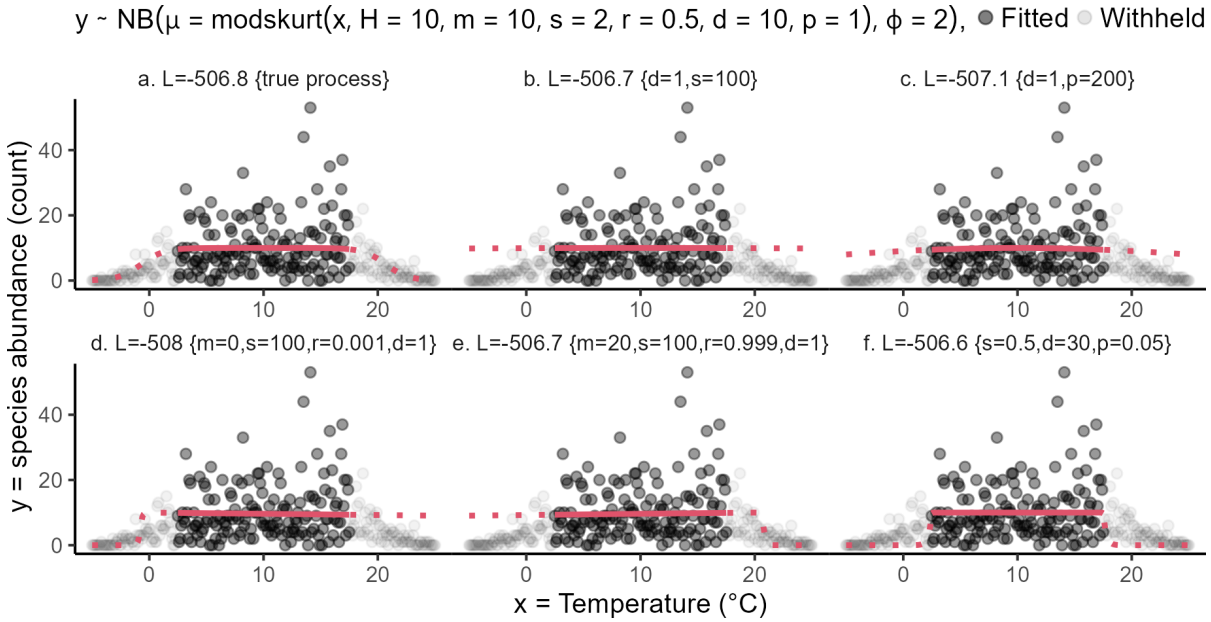

Figure 2.2: The dot plots show count data simulated from an NB+modskurt distribution with parameters $H = 10, m = 10, s = 2, r = 0.5, d = 10, p = 1$, and $\phi = 2$ over a possible temperature gradient. Each subplot shows potential modskurt predictions for the mean, μ, that have trivially different log-likelihood values (L) between 2.5 and 17.5°C (solid red line and dark dots) but wildly different and potentially undesirable (see Table 2.1) modskurt parameter values (changes from the data generating process, subplot a., are denoted in curly braces). The dashed lines show how these different mean shapes would predict average counts poorly calibrated to the true data distribution in more extreme temperatures.

2.4 Discussion

This study offers an in-depth examination of the nonlinear zero-inflated negative binomial with modskurt mean function (NZM) model (2.5) and proposes a refined parameter space that encourages more reliable and realistic predictions of distributions of species abundance along environmental gradients (DAEG). We found that the NZM model, with its advanced modelling of species abundance and environmental conditions, has many applications and potential for improving our understanding of complex ecological dynamics (Martins et al. 2023). However, we recognise the limitations of this model,

particularly regarding parameter identifiability issues and multimodality in parameter estimation, which could result in potential stability and convergence concerns (Raue et al. 2009). By acknowledging these issues and refining the model's parameter space, we aim to improve its performance and applicability in diverse ecological scenarios.

The key to refining the parameter space lies in removing redundant regions and ensuring that each parameter value in the reduced space uniquely affects the modskurt shape in isolation. Moreover, focusing on values that are "credible" in the context of the DAEG being studied further improves the reliability of the model's predictions. Despite these improvements, we must be cognizant that our refined parameter space could still overlook some potentially problematic parameter dependencies, such as the joint effects of s, d, and p on the shape of the mean function. It is also worth noting that the credible regions of the refined parameter space rely heavily on the specific ecological scenario and existing data. Therefore, it requires rigorous verification and adjustment in actual application.

In summary, while the NZM model shows promise for understanding and predicting species abundance patterns in response to environmental gradients, we propose that a refined parameter space can further enhance its reliability and practical use. Future research should explore how to efficiently incorporate these refinements into model implementation and continue to evaluate and refine the parameter space in light of new data and ecological insights.

Chapter 3

Utilising Bayesian prior probability

3.1 Overview

Bayesian data analysis allows prior information about the probability of different parameter values to be directly specified in a model (Gelman, Carlin, et al. 2013; McElreath 2020). These probabilities can be used to increase the level of ecological structure in the parameter estimation process and are invaluable when the model is complex and weakly defined – as is the case for the nonlinear zero-inflated negative binomial with modskurt mean (NZM) model introduced in the previous chapter for analysing distributions of species abundance along environmental gradients. However, correct specification of prior probability can be challenging in practice; encoding too little or too much information can jeopardise the accuracy and integrity of results (Gabry, Daniel Simpson, et al. 2019; Banner et al. 2020). In this chapter, the aim is to guide how to encode prior information for the parameters in a Bayesian implementation of the NZM model. This aim will be achieved by specifying a candidate prior probability distribution that encodes the parameter subspace defined in Table 2.1. Next, a workflow example will be stepped through that details how to check and adjust the default priors for the specific species-environment data in hand. To conclude, the implications and limitations of the research in this chapter will be discussed.

3.2 The Bayesian NZM

Statistical modelling using the Bayesian inference enables the inclusion of prior beliefs in the modelling process (Gelman, Carlin, et al. 2013). The Bayesian data analysis approach to parameter estimation transforms an initial "prior" belief about a probability distribution into an updated "posterior" distribution. For the nine parameters of the NZM model (2.5), $\vec{\theta} = (H, m, s, r, d, p, \phi, \gamma_0, \gamma_1)^\intercal$, the Bayesian posterior distribution is defined as

$$\Pr(\vec{\theta}|y_n, x_n) = \frac{\mathrm{L}(\vec{\theta}|y_n, x_n) \cdot \Pr(\vec{\theta})}{\Pr(y_n, x_n)} \tag{3.1}$$

$$\propto \mathrm{L}(\vec{\theta}|y_n, x_n) \cdot \Pr(\vec{\theta}) \tag{3.2}$$

where $\Pr(\vec{\theta})$ is the joint probability distribution[1] that describes the prior probability of observing each parameter value independent of any observed data, and the joint likelihood function $\mathrm{L}(\vec{\theta}|y_n, x_n)$ describes the probability of observing the data given different parameter values in the NZM model (2.1). The marginal likelihood $\Pr(y_n, x_n)$ is the probability of observing the data averaged over all possible parameter values weighted by their prior probabilities (e.g. the integral of the numerator over all $n \in N$ with respect to $\vec{\theta}$); this term normalises the posterior so that it has a total probability density of 1.

The NZM posterior distribution of (3.1) offers a natural way to encode the refined parameter space developed in the previous chapter (Table 2.1). Instead of specifying binary thresholds on what parameter values are credible or not–as might be required in constrained optimisation (Box 1965)–prior probability can be used to "err" values away from those deemed less credible or redundant and concentrated around values deemed most suitable given ecological knowledge of the DAEG (McElreath 2020). For

[1] For simplicity, we assume that the parameters are independent in the joint prior distribution and compute it as the product of the independent marginal prior distributions. However, as discussed in the previous chapter, the assumption of parameter independence in the NZM model is not necessarily valid.

the modskurt parameters, a possible starting point is

$$H \sim \text{Beta}(2, 3) \cdot \max y \tag{3.3}$$

$$m \sim \text{Beta}(1, 1) \cdot \text{range } x + \min x \tag{3.4}$$

$$\log s \sim \text{Normal}(-2, 0.6) + \log \text{range } x \tag{3.5}$$

$$r \sim \text{Beta}(1.2, 1.2) \tag{3.6}$$

$$\log d \sim \text{Normal}(0.5, 1) \tag{3.7}$$

$$p \sim \text{Beta}(1.2, 1.2) \cdot 1.95 + 0.05, \tag{3.8}$$

where data extremes are used to rescale the priors for H, m, and s for general specification – the effect of parameters r, d, and p on the shape of the modskurt function are already independent of the scale of the data. Figure 3.1 illustrates how these distributions assign a prior probability to different parameter values using an example DAEG, with Appendix sections A.1 and A.3 providing for more information on how to adjust these priors and examine the effect they have on prior predictions for mean abundance along a gradient.

For the overdispersion parameter of the negative binomial distribution, there is no information from Table 2.1 to encode, so we follow the recommendations of Dan Simpson (2018) and A. N. Smith et al. (2020) and define a prior distribution for values of the inverse square-root of the reciprocal dispersion (e.g. square root of dispersion) as

$$1/\sqrt{\phi} = \kappa \sim \text{Exponential}(0.5) \tag{3.9}$$

where as $\phi^{-0.5} \to 0$ the distribution approaches equidispersion, and $\phi^{-0.5} \to \infty$ increases the amount of overdispersion, (bottom-left Figure 3.1). Weakly informative prior distributions for the linked zero-inflation parameters are proposed as

$$\gamma_0 \sim \text{Normal}(3, 1.5) \tag{3.10}$$

$$\gamma_1 \sim \text{Half-Normal}(3), . \tag{3.11}$$

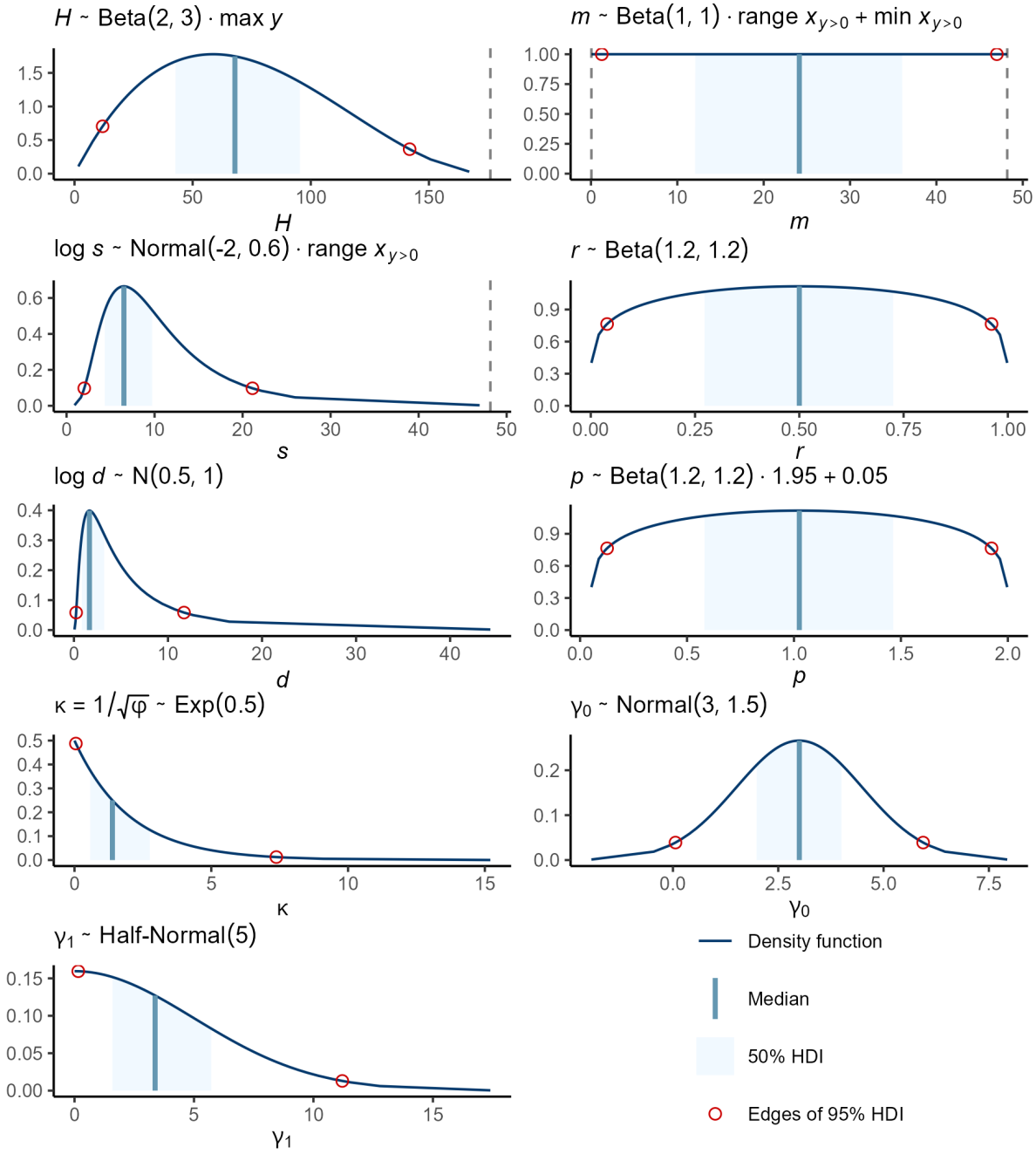

Figure 3.1: Marginal prior probability density plots for each parameter in the NZM model. For this model, $\max y = 176$, range $x = 48.2$, and $\min x = 0.05$. "HDI" stands for highest density interval and is used here to show regions of values assigned different prior probabilities of occurrence.

See Appendix material A.1 and A.2 that detail the choice of priors for the three ZINBL shape parameters. It is important to stress that these suggested prior distributions are just starting points and should be checked in every DAEG context.

3.3 In practice

Building, assessing, and revising a Bayesian statistical model is an iterative process that heavily relies on domain knowledge, statistical reasoning, and computational tools (Gelman, Vehtari, et al. 2020). It involves carefully defining the model's components, selecting appropriate prior distributions, fitting the model to the data, and examining the results to ensure they align with ecological insights and expectations. In this section, we step through a workflow (Figure 3.2) for the Bayesian NZM model (3.1) using the supplementary R package (R Core Team 2023) for this thesis (Appendix A.1; `https://hdrab127.github.io/modskurt/`) on an example DAEG with actual data. This section is not a "how to" on Bayesian data analysis–plenty of resources are much better suited to that task than this thesis–the primary goal is instead to provide some steps specific to the Bayesian data analysis of DAEGs using the NZM model.

A. Specify an initial model

The first step in the workflow is to specify the initial Bayesian NZM model. This involves thinking about the different properties of the DAEG that the NZM model seeks to predict (e.g. overdispersion, zero-inflation, and the shape of the relationship between y and x) and eliciting prior information into the parameters that describe those properties. For a running example, we will analyse the distribution of *Aonides trifida* (a small spionid worm found in intertidal estuaries of New Zealand) along a gradient of sediment mud content. Assume for this example that we have limited prior knowledge of this *Aonides*-mud relationship and specify the initial model as the zero-inflated negative binomial model of (2.1) and modskurt mean function (2.4) with the default priors specified earlier

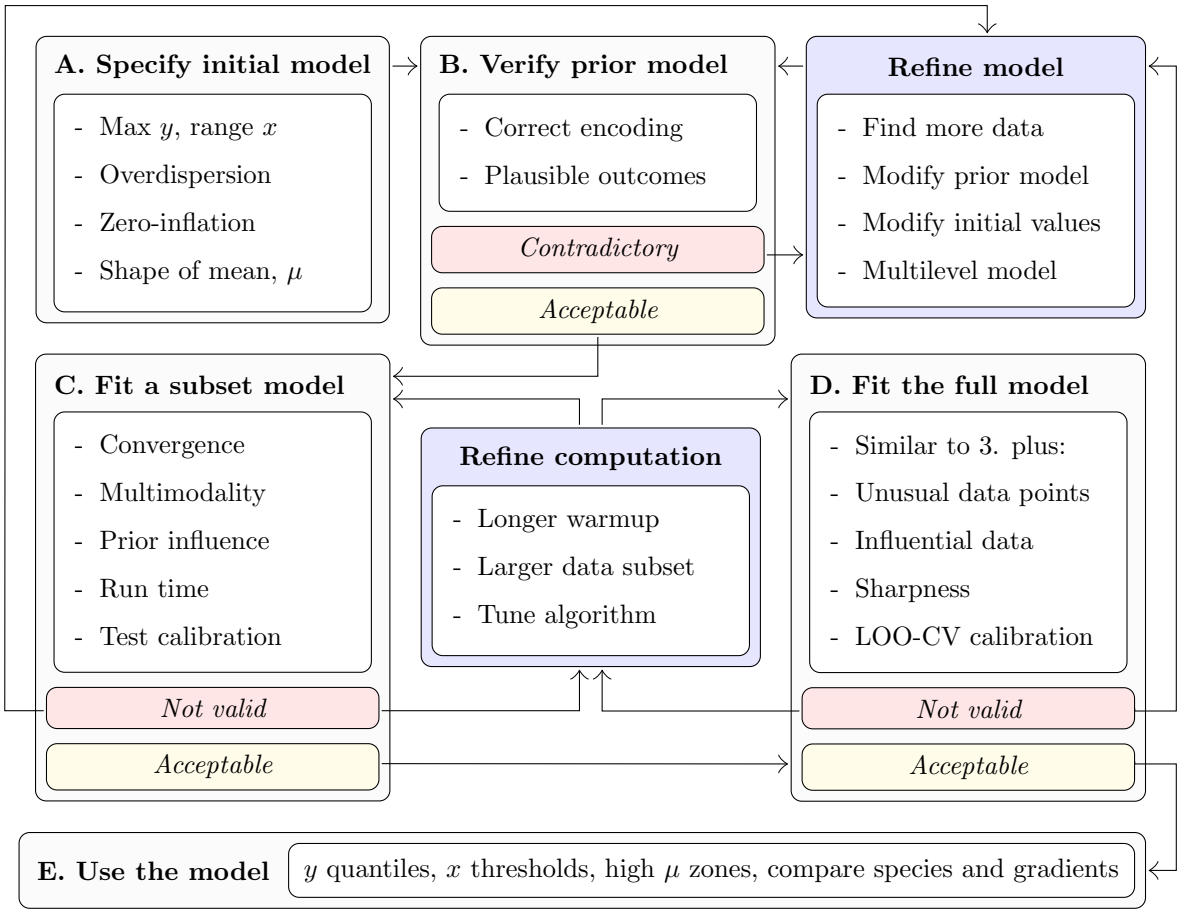

Figure 3.2: High-level view of steps in a Bayesian NZM analysis of DAEG adapted from (Gelman, Vehtari, et al. 2020). The lettered grey boxes indicate core steps to be followed sequentially, while the light blue boxes address specification and computational issues optionally.

(equations 3.3–3.11).

B. Verify the prior model

Once the initial NZM model is specified, we recommend that prior predictive checks (Gabry, Daniel Simpson, et al. 2019) include the range of possible modskurt shapes the prior distribution encourages and the maximum abundances and proportions of zero produced by the prior parameter values for overdispersion and zero-inflation. For our *Aonides* model, we can verify that the marginal distributions in Figure 3.1 reflect the prior knowledge outlined in Table 2.1 and refer to the online prior specification tools of

this thesis (Appendices A.2, A.3) for outputs of the necessary prior predictive checks.

Refine the model

Based on the prior model verification results, it may be necessary to refine the initial model. This step could involve various adjustments, such as acquiring more data to improve parameter estimation, modifying prior distributions if they were not correctly encoded in Step B, tuning the model's initial values, or considering a multilevel model if the data exhibits a hierarchical structure.

C. Fit a subset model

Before fitting the Bayesian NZM model to the entire dataset, it is highly recommended to fit a model to a subset of "training" data and a small sample of the posterior distribution. The idea here is to fail fast and quickly identify any misspecification issues in the model before wasting human and computer time trying to achieve near-perfect Bayesian inference on the full dataset (Gelman, Vehtari, et al. 2020). Common misspecifications here arise from the parameter identifiability concerns raised in Chapter 2; if there is too little data information to estimate the full NZM model, then perhaps the model can be simplified not to include zero-inflation or to fit a simpler approximation of the modskurt shape (M. J. Anderson et al. 2022).

For the *Aonides* example, we use a 30% subset of the available data and draw six chains of samples from the posterior distribution using the Hamiltonian Monte-Carlo methods of Stan (Stan Development Team 2023) implemented in Gabry, Češnovar, et al. (2023). Figure 3.3 shows the marginal posterior distributions for each chain of samples, and given that each chain appears to find similar posterior solutions that do not dramatically contradict the prior information, we can be confident that the model is well specified for now.

Withholding data from the model computation also allows us to test the predictive calibration of the posterior. Calibration refers to the distance between the empirical dis-

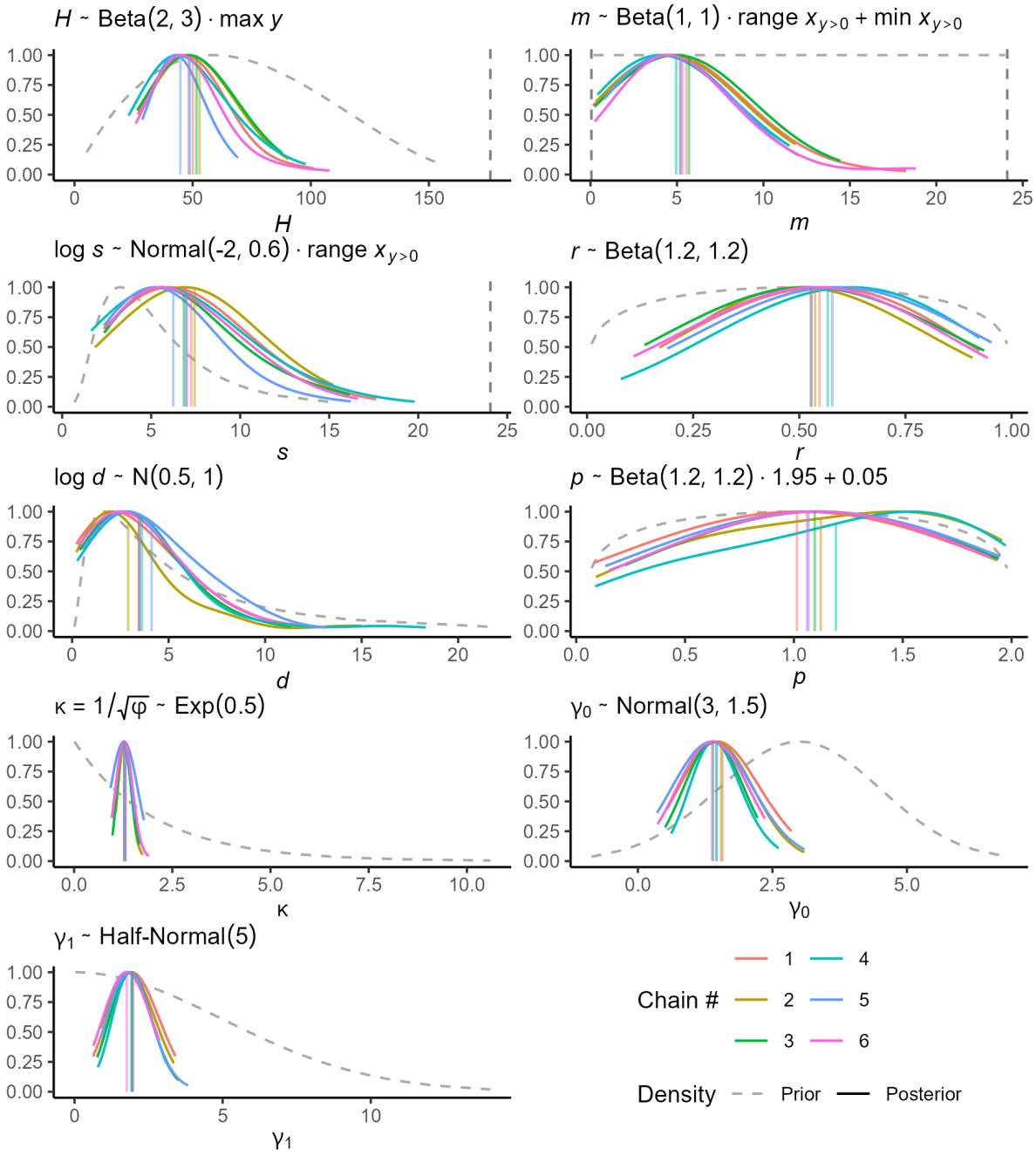

Figure 3.3: Boundary corrected kernel density estimates for the marginal posterior distributions sampled in the subset model.

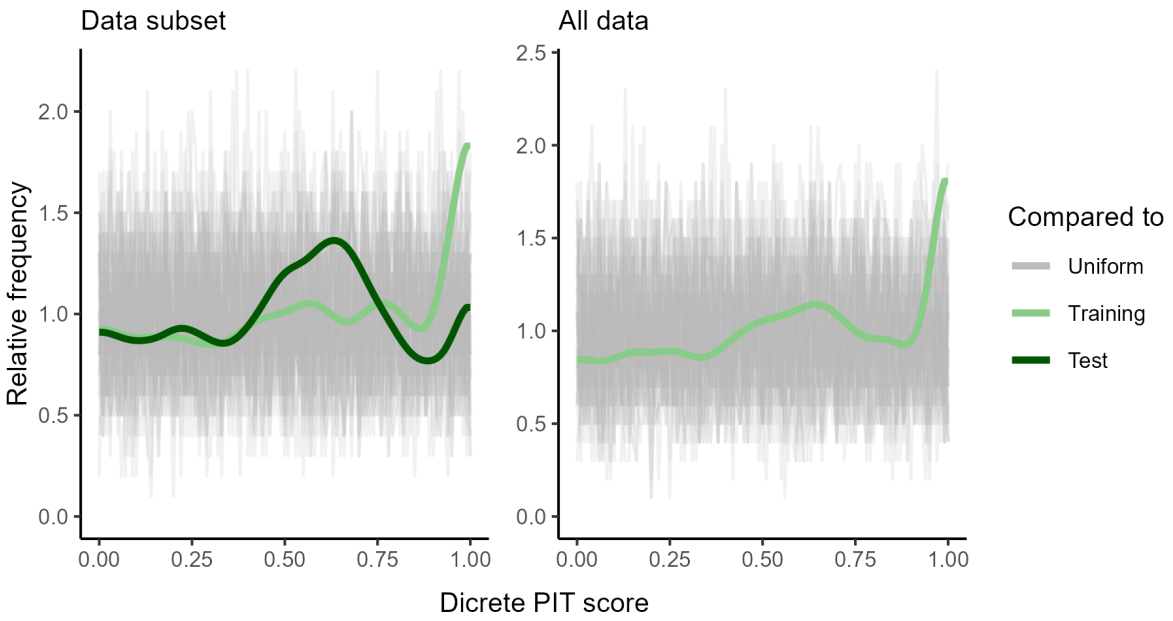

Figure 3.4: Randomised discrete PIT assessment of predictive calibration following the methods described in Czado et al. (2009).

tribution of the withheld DAEG data and the posterior predictive distribution estimated from the training data (Czado et al. 2009). For these discrete abundance distributions, we use the randomised probability integral transformation (PIT) test of Czado et al. (2009). The left-hand subplot of Figure 3.4 compares the frequency of PIT scores for the *Aonides* training and test data. For a model that accurately predicts new data, these PIT scores should be uniformly distributed (Gabry, Daniel Simpson, et al. 2019), and given that the PIT scores for the posterior predictions of test data are well within the range of the random uniform samples, we can be confident for now that the predictive capability of the model is well calibrated.

Optional: Refine the computation

In some cases, the NZM posterior model could be well specified but simply hard to estimate, and refining the computation may be necessary to determine this. Refinements could include adjusting the length of the warmup period, using a larger data subset for computation, or in rare cases, tuning algorithm parameters such as the size of steps used

to explore the posterior curvature.

D. Fit the full model

If the prior model of Step B and subset posterior of Step C pass all checks, then we can proceed to fit the Bayesian NZM model to the entire dataset with a more extensive sample from the posterior. Similar checks for specification and calibration can be performed here; however, leave-one-out cross-validated PIT scores are required in place of the test data set (Vehtari et al. 2017).

E. Use the model

Once the Bayesian NZM model is fitted and validated, it can finally be utilised to analyse the DAEG. The posterior predictive distribution it provides opens up a realm of possibilities for deriving quantities of interest (Gelman, Carlin, et al. 2013). However, two applications for DAEGs may be of particular use; 1, the identification of environmental conditions that predict high levels of mean abundance, and 2, limits of the environmental variable that could "protect" specific densities of predicted total abundance. Figure 3.5 calculates these two summaries for the distribution of *Aonides trifida* along the gradient of sediment mud concentrations, suggesting that the majority of predicted mean abundance ($\mu > 0.5 \cdot H$) is between about 0.2 and 11.6% mud content and 90% of mean abundance is predicted to be found in sediment with less than approximately 14.4% mud content. See the articles in Supplementary link A.1 for ways to present the posterior uncertainty for these DAEGs.

3.4 Discussion

In this chapter, we explored the use of Bayesian prior probabilities to implement the refined parameter space of the nonlinear zero-inflated negative binomial with modskurt mean function (NZM) model in Chapter 2. We showed that by using a well-specified

prior probability distribution and following a practical Bayesian modelling workflow, it is possible to obtain a posterior distribution that reliably reflects our knowledge and uncertainty about the parameters (Gabry, Daniel Simpson, et al. 2019; Gelman, Vehtari, et al. 2020). This Bayesian NZM methodology offers a robust guide for ecologists and modellers that deal with the inherent complexity and variation of distributions of species abundance along environmental gradients (DAEG).

The supplementary materials provide further support to this chapter. Firstly, the open-source R package (R Core Team 2023) (Appendix A.1; `https://hdrab127.github.io/modskurt/`), available on GitHub and further documented on the associated website, provides a ready-to-use tool to estimate DAEG using the NZM model. Secondly, visual prior specifications tools (Appendix A.2; `https://hdrab127.github.io/modskurt/articles/zinbl-priors.html` and A.3; `https://hdrab127.github.io/modskurt/articles/modskurt-priors.html`) aid in understanding and applying the described method, and enable users to grasp the impact of different prior choices better and evaluate their effectiveness. These supplements, in synergy with the guidance given in this chapter, make the Bayesian NZM more accessible to practitioners.

However, this approach has its limitations. The effectiveness of the prior distribution in representing the true state of knowledge can be influenced by many factors, including the quality and quantity of the data and the chosen model structure (Banner et al. 2020). Furthermore, this method requires a certain level of expertise and understanding of Bayesian analysis and NZM models.

In conclusion, this chapter and supporting online materials offer a valuable resource for those working with NZM models in a Bayesian framework. While providing a robust starting point, users must understand the assumptions and constraints of this method. Continued exploration and refinement of methods and interdisciplinary collaboration will be crucial for advancing our understanding of species-environment interactions and making robust ecological predictions.

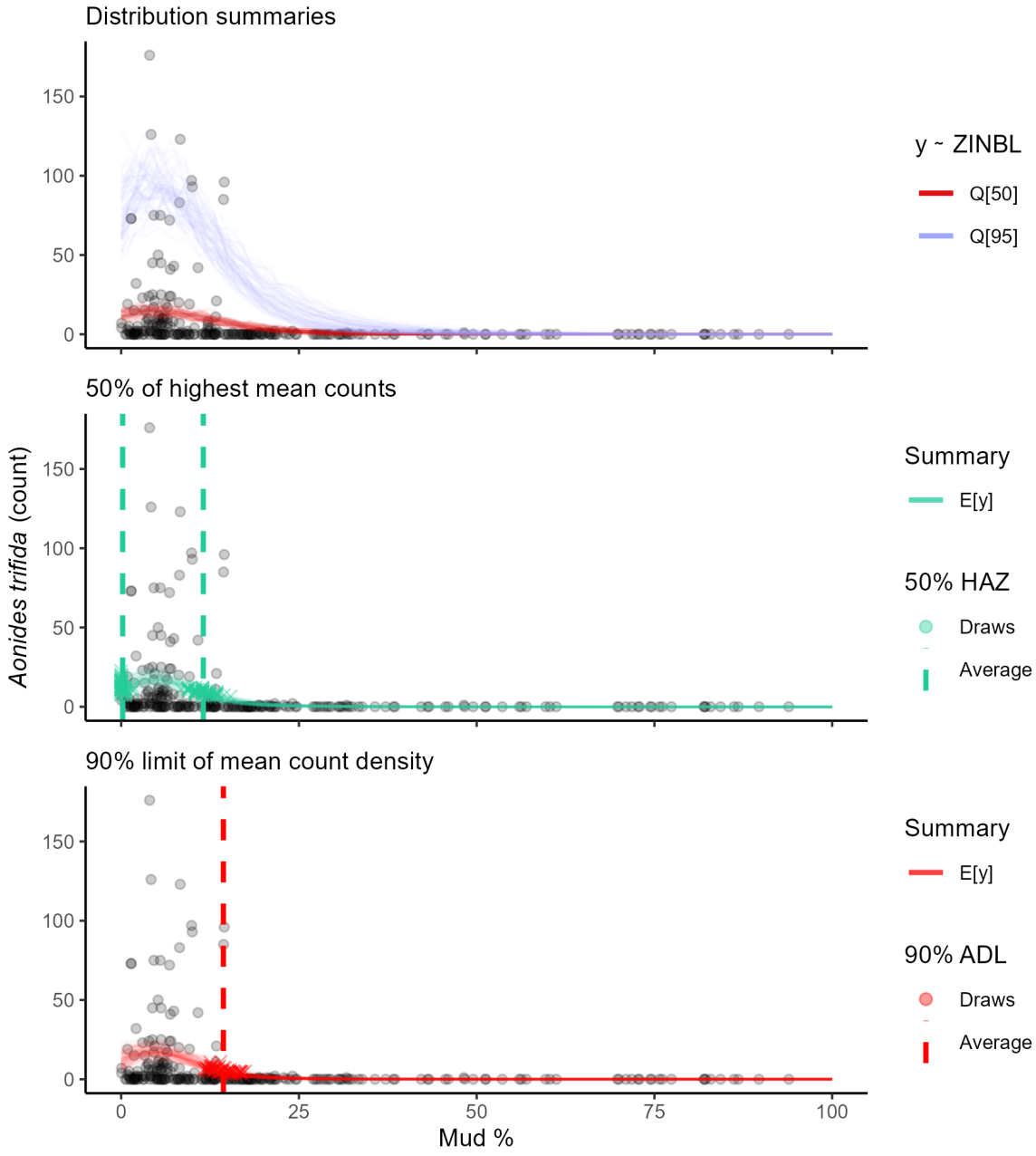

Figure 3.5: The top row shows fifty random draws from the posterior predictive distribution for the median and 95% quantile of the distribution of *Aonides trifida* along the gradient of mud concentrations. The middle and bottom row show posterior predicted 50% high abundance zones (HAZ, middle subplot) and 90% abundance density limits (ADL, bottom) based on fifty random posterior predictions of mean *Aonides trifida* count along the mud gradient. The light grey dots show the actual abundance counts observed.

Chapter 4

Conclusion

4.1 Overview

The aim of this thesis was to enhance the reliability and credibility of the nonlinear zero-inflated negative binomial with modskurt mean function (NZM) model's predictions of distributions of species abundance along environmental gradients (DAEG). It successfully achieved this goal by refining the model's parameter space and incorporating Bayesian inference techniques. As a result, the research represents a significant advancement in understanding and applying the NZM model to predict species abundance patterns along environmental gradients.

4.2 Summary of contributions

The work presented in Chapter 2 extensively explored the NZM model and its sensitivity to different parameter values, revealing complexities and challenges in parameter estimation. Refining the model's parameter space addressed these issues and improved the reliability of predictions. Chapter 3 introduced a Bayesian implementation of the NZM model, incorporating prior probability distributions for the parameters. This innovative approach improved the ecological structure in parameter estimation but necessitated expertise in Bayesian analysis for proper utilisation. The online supplementary materials

(Appendix A) comprise a major part of the original contributions of this thesis. These materials include:

1. An open-source R package (R Core Team 2023) that facilitates the estimation of DAEG using the NZM model. The package is available on GitHub and further documented on the associated website (`https://hdrab127.github.io/modskurt/`).

2. Visual prior specifications tools that aid in understanding and applying the described method, allowing users to evaluate the impact of different prior choices more effectively. Images of these resources in action and their links are provided in Appendix A.2 and A.3.

These supplementary materials, combined with the guidance in Chapter 3, make the Bayesian NZM model more accessible to practitioners.

4.3 Implications for ecological statistics

The findings of this research hold significant implications for ecological researchers and conservation practitioners. The refined NZM model and its Bayesian implementation offer an advanced tool for understanding and predicting species abundance patterns along environmental gradients. The enhanced predictive capability achieved is crucial for accurately assessing species abundance in changing environmental conditions, effectively supporting conservation and management efforts. Furthermore, providing a detailed workflow and case study, along with the open-source R package, makes these sophisticated modelling techniques more accessible to researchers and practitioners.

4.4 Directions for future research

Although this thesis has made substantial progress in refining the NZM model and enhancing its reliability and credibility, further research is needed to advance the field. Specifically:

1. Exploring potential parameter dependencies in the refined parameter space to improve the reliability of parameter estimation. This could involve reparametrisation of the modskurt model or investigating correlated prior distributions in the Bayesian NZM model, extending the work presented in this thesis.

2. Developing more user-friendly tools and resources for researchers and practitioners interested in utilising prior probability for analysing DAEGs, given the complexity of the NZM model and the Bayesian techniques employed.

Additionally, there is excellent potential to extend the Bayesian NZM model for hierarchical data, for example, where data come from different years or geographically separated local populations. Hierarchical modelling would allow us to understand ecological systems' nested and interconnected nature and potentially allow for simpler distributions and mean functions by providing a more comprehensive model of the systems underlying the DAEG. It could also pave the way for analysing the interplay between multiple species and environmental gradients, providing a more holistic view of ecological diversity and its influencing factors. Altogether, this could be particularly beneficial for studies incorporating data from various spatial and temporal scales.

4.5 Conclusion

To conclude, this thesis has advanced our understanding of species abundance distributions along environmental gradients, refining the nonlinear zero-inflated negative binomial model with a modskurt mean function and introducing a Bayesian inference method for parameter estimation. The open-source R package and visual prior specifications tools, provided as supplementary materials, increase the accessibility of these sophisticated techniques. Hopefully, this work will contribute to the ongoing efforts of ecological statistics to understand the complex dynamics of species abundance and support effective conservation and management decisions in our changing environment.

Bibliography

Anderson, Marti J, Daniel CI Walsh, Winston L Sweatman, and Andrew J Punnett (2022). "Non-linear models of species' responses to environmental and spatial gradients". In: *Ecology Letters*. ISSN: 1461-023X.

Anderson, RM and Robert M May (1978). "Regulation and stability of host-parasite population interactions". In: *Journal of Animal Ecology* 47.1, pp. 219–247.

Banner, Katharine M, Kathryn M Irvine, and Thomas J Rodhouse (2020). "The use of Bayesian priors in Ecology: The good, the bad and the not great". In: *Methods in Ecology and Evolution* 11.8, pp. 882–889. ISSN: 2041-210X.

Box, MJ (1965). "A new method of constrained optimization and a comparison with other methods". In: *The Computer Journal* 8.1, pp. 42–52.

Brown, James H. (1984). "On the Relationship between Abundance and Distribution of Species". In: *The American Naturalist* 124.2, pp. 255–279. ISSN: 00030147, 15375323. URL: http://www.jstor.org/stable/2461494.

Chase, Jonathan M and Mathew A Leibold (2009). *Ecological niches: linking classical and contemporary approaches*. University of Chicago Press.

Clapham, A. R. (1936). "Over-Dispersion in Grassland Communities and the Use of Statistical Methods in Plant Ecology". In: *Journal of Ecology* 24.1, pp. 232–251. ISSN: 00220477, 13652745. DOI: 10.2307/2256277. URL: http://www.jstor.org/stable/2256277.

Clark, James S (2007). *Models for ecological data: an introduction*. Princeton University Press.

Connell, Joseph H (1961). "The influence of interspecific competition and other factors on the distribution of the barnacle Chthamalus stellatus". In: *Ecology*, pp. 710–723.

Czado, Claudia, Tilmann Gneiting, and Leonhard Held (2009). "Predictive model assessment for count data". In: *Biometrics* 65.4, pp. 1254–1261. ISSN: 0006-341X.

Derocher, Andrew E, Nicholas J Lunn, and Ian Stirling (2004). "Polar bears in a warming climate". In: *Integrative and Comparative Biology* 44.2, pp. 163–176.

Deutsch, Curtis A, Joshua J Tewksbury, Raymond B Huey, Kimberly S Sheldon, Cameron K Ghalambor, David C Haak, and Paul R Martin (2008). "Impacts of climate warming on terrestrial ectotherms across latitude". In: *Proceedings of the National Academy of Sciences* 105.18, pp. 6668–6672.

Eliason, Scott R (1993). *Maximum likelihood estimation: Logic and practice*. Sage. ISBN: 0803941072.

Elith, Jane and John R Leathwick (2009). "Species distribution models: ecological explanation and prediction across space and time". In: *Annual Review of Ecology, Evolution, and Systematics* 40, pp. 677–697.

Gabry, Jonah, Rok Češnovar, and Andrew Johnson (2023). "cmdstanr: R Interface to 'CmdStan'". In: https://mc-stan.org/cmdstanr/, https://discourse.mc-stan.org.

Gabry, Jonah, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman (2019). "Visualization in Bayesian workflow". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182.2, pp. 389–402. ISSN: 0964-1998.

Gaston, Kevin J and Tim M Blackburn (2000). *Pattern and process in macroecology*. Wiley Online Library.

Gauch, Hugh G., Gene B. Chase, and Robert H. Whittaker (1974). "Ordination of Vegetation Samples by Gaussian Species Distributions". In: *Ecology* 55.6, pp. 1382–1390. ISSN: 00129658, 19399170. DOI: 10.2307/1935466. URL: http://www.jstor.org.ezproxy.massey.ac.nz/stable/1935466.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin (2013). *Bayesian data analysis*. CRC press.

Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák (2020). "Bayesian workflow". In: *arXiv preprint arXiv:2011.01808*.

Goodall, D. W. and R. W. Johnson (1982). "Non-Linear Ordination in Several Dimensions: A Maximum Likelihood Approach". In: *Vegetatio* 48.3, pp. 197–208. ISSN: 00423106. URL: http://www.jstor.org.ezproxy.massey.ac.nz/stable/20037091.

Guillera-Arroita, Gurutzeta, José J Lahoz-Monfort, Jane Elith, Ascelin Gordon, Heini Kujala, Pia E Lentini, Michael A McCarthy, Reid Tingley, and Brendan A Wintle (2015). "Is my species distribution model fit for purpose? Matching data and models to applications". In: *Global Ecology and Biogeography* 24.3, pp. 276–292.

Guisan, Antoine and Niklaus E Zimmermann (2000). "Predictive habitat distribution models in ecology". In: *Ecological Modelling* 135.2-3, pp. 147–186.

Hastie, Trevor, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.

Hernández-Blanco, Marcello, Robert Costanza, Haojie Chen, Dolf DeGroot, Diane Jarvis, Ida Kubiszewski, Javier Montoya, Kamaljit Sangha, Natalie Stoeckl, and Kerry Turner (2022). "Ecosystem health, ecosystem services, and the well-being of humans and the rest of nature". In: *Global Change Biology* 28.17, pp. 5027–5040. ISSN: 1354-1013.

Jamil, Tahira and Cajo JF Ter Braak (2013). "Generalized linear mixed models can detect unimodal species-environment relationships". In: *PeerJ* 1, e95. ISSN: 2167-8359.

Kéry, Marc and J Andrew Royle (2020). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and advanced models*. Academic Press.

Martins, Paulo Mateus, Marti J. Anderson, Winston L. Sweatman, and Andrew J. Punnett (2023). "Significant shifts in latitudinal optima of North American birds". Under peer review.

McElreath, Richard (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC. ISBN: 0429029608.

McGill, Brian and Cathy Collins (2003). "A unified theory for macroecology based on spatial patterns of abundance". In: *Evolutionary Ecology Research* 5.4, pp. 469–492. ISSN: 1522-0613.

Paine, Robert T (1966). "Food web complexity and species diversity". In: *The American Naturalist* 100.910, pp. 65–75.

Parmesan, Camille (2006). "Ecological and evolutionary responses to recent climate change". In: *Annu. Rev. Ecol. Evol. Syst.* 37, pp. 637–669.

Pearman, Peter B, Antoine Guisan, Olivier Broennimann, and Christophe F Randin (2008). "Niche dynamics in space and time". In: *Trends in Ecology & Evolution* 23.3, pp. 149–158.

Pörtner, Hans-O, Debra C Roberts, Helen Adams, Carolina Adler, Paulina Aldunce, Elham Ali, Rawshan Ara Begum, Richard Betts, Rachel Bezner Kerr, and Robbert Biesbroek (2022). *Climate change 2022: Impacts, adaptation and vulnerability*. IPCC Geneva, Switzerland:

R Core Team (2023). "R: A Language and Environment for Statistical Computing". In: URL: `https://www.R-project.org/`.

Raue, Andreas, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller, and Jens Timmer (2009). "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood". In: *Bioinformatics* 25.15, pp. 1923–1929.

Robert, Christian P, George Casella, and George Casella (1999). *Monte Carlo statistical methods*. Vol. 2. Springer.

Rosenzweig, Michael L (1995). *Species diversity in space and time*. Cambridge University Press.

Simpson, Dan (2018). *Justify My Love*. https://statmodeling.stat.columbia.edu/2018/04/03/justify-my-love/. [Online; accessed 3-April-2023].

Smith, Adam NH, David Acuña-Marrero, Pelayo Salinas-de-León, Euan S Harvey, Matthew DM Pawley, and Marti J Anderson (2020). "Instantaneous vs. non-instantaneous diver-operated stereo-video (DOV) surveys of highly mobile sharks in the Galápagos Marine Reserve". In: *Marine Ecology Progress Series* 649, pp. 111–123. ISSN: 0171-8630.

Smith, ANH, MJ Anderson, and RB Millar (2012). "Incorporating the intraspecific occupancy-abundance relationship into zero-inflated models". In: *Ecology* 93.12, pp. 2526–2532. ISSN: 1939-9170.

Stan Development Team (2023). "Stan Modeling Language Users Guide and Reference Manual, 2.28." In: URL: https://mc-stan.org.

Stoklosa, Jakub, Rachel V Blakey, and Francis KC Hui (2022). "An overview of modern applications of negative binomial modelling in ecology and biodiversity". In: *Diversity* 14.5, p. 320. ISSN: 1424-2818.

Ter Braak, Cajo JF (1987). *Unimodal models to relate species to environment*. Wageningen University and Research.

Tilman, David (1982). *Resource competition and community structure*. Princeton University Press.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry (2017). "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC". In: *Statistics and Computing* 27 (5), pp. 1413–1432. DOI: 10.1007/s11222-016-9696-4.

Vellend, Mark (2016). "The theory of ecological communities (MPB-57)". In: *The Theory of Ecological Communities (MPB-57)*. Princeton University Press. ISBN: 1400883792.

Waldock, Conor, Rick D Stuart-Smith, Camille Albouy, William WL Cheung, Graham J Edgar, David Mouillot, Jerry Tjiputra, and Loïc Pellissier (2022). "A quantitative

review of abundance-based species distribution models". In: *Ecography* 2022.1. ISSN: 0906-7590.

Whittaker, R. H. (1956). "Vegetation of the Great Smoky Mountains". In: *Ecological Monographs* 26.1, pp. 2–80. ISSN: 00129615. DOI: 10.2307/1943577. URL: http://www.jstor.org/stable/1943577.

Woodward, Frank Ian (1987). *Climate and plant distribution*. Cambridge University Press.

Appendices

Appendix A

Supplementary material for Chapter 3

A.1 R package and online documentation

The contents of Appendix A.1 are available at:

- `https://github.com/hdrab127/modskurt`, and

- `https://hdrab127.github.io/modskurt/`.

Description: This supplement is an R package (R Core Team 2023) that enables the Bayesian estimation of distributions of species abundance along environmental gradients using the nonlinear modskurt mean function with the zero-inflated negative binomial distribution of counts with the probability of excess-zero linked to the mean (ZINBL). The second link is a website with additional resources to accompany the R package.

A.2 ZINBL prior specification

The contents of Appendix A.2 are available at:

- `https://hdrab127.github.io/modskurt/articles/zinbl-priors.html`.

Description: This supplementary tool provides a set of interactive graphs and tables to assist prior specification and prior predictive checks for the negative binomial and ZINBL distributions of species abundance (Figure A.1).

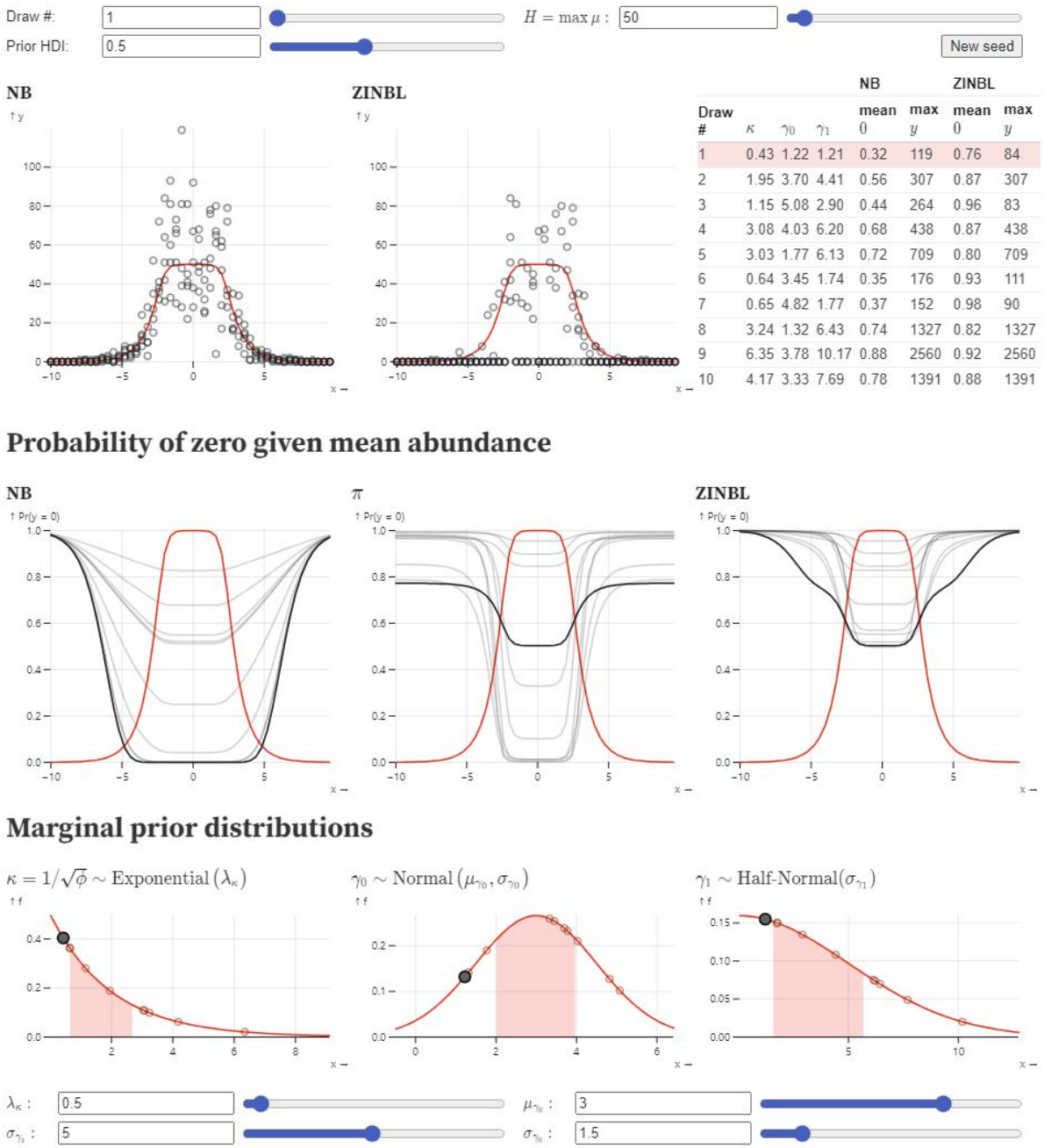

Figure A.1: Screenshot of the interactive tool for prior specification and verification for NB and ZINBL distribution parameters. The sliders allow adjustments of the prior specification, while the charts and table update in real-time to illustrate the assumptions specified in the model.

A.3 Modskurt prior specification

The contents of Appendix A.1 are available at:

- https://hdrab127.github.io/modskurt/articles/modskurt-priors.html.

Description: Similarly to A.2, this supplementary tool provides a set of interactive graphs to assist prior specification and prior predictive checks for the modskurt function that describes how average species abundance changes along an environmental gradient (Figure A.2).

A.4 References

R Core Team (2023). "R: A Language and Environment for Statistical Computing". In: URL: https://www.R-project.org/.

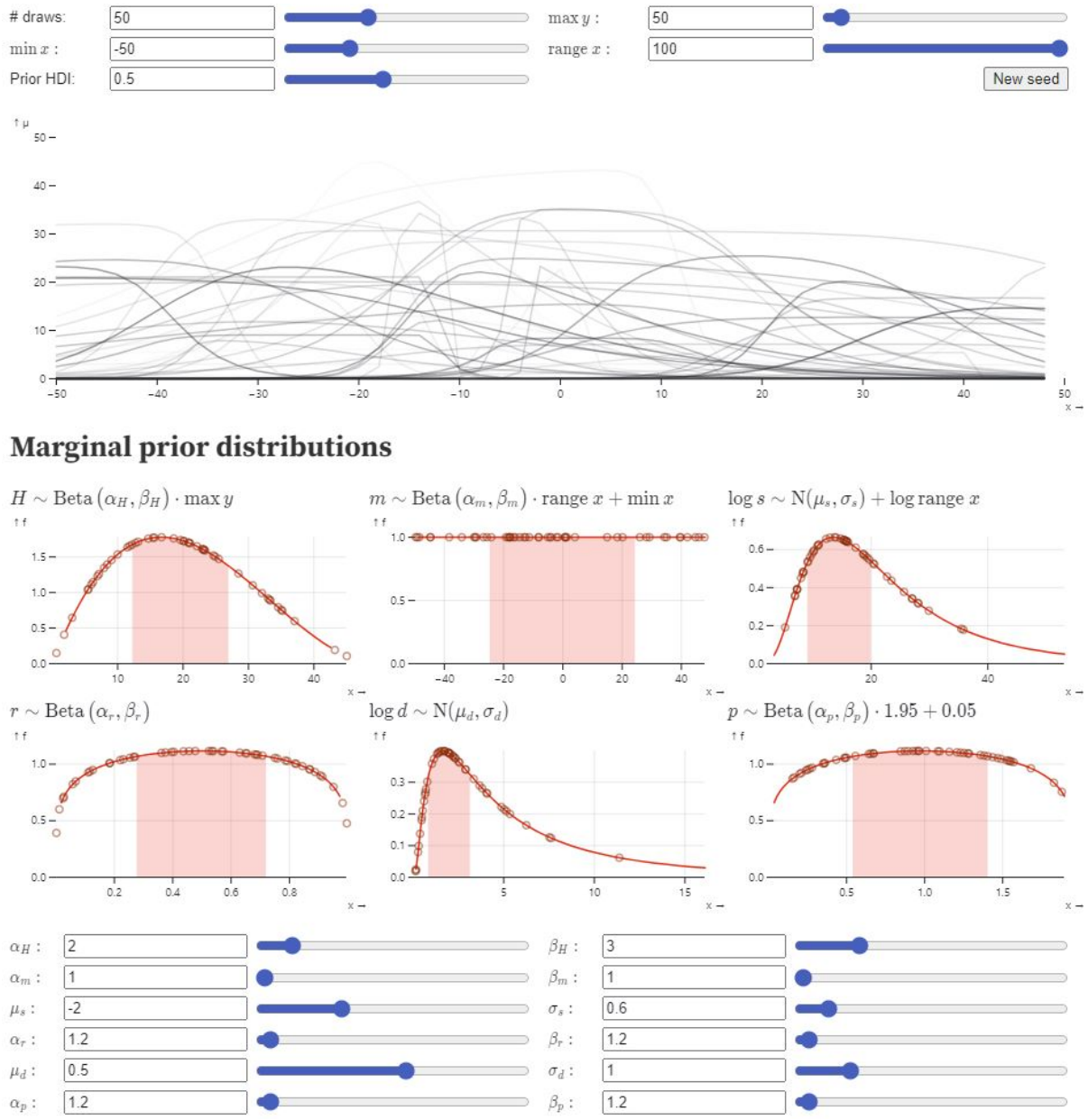

Figure A.2: Screenshot of the interactive tool for prior specification and verification for parameters of the modskurt mean function. The sliders allow adjustments of the prior specification, while the charts and table update in real-time to illustrate the assumptions specified in the model.