

# Model linearity breeds contempt: using Bayesian non-linear models to uncover general biogeographical patterns

Bernat Bramon Mora<sup>1,\*</sup> and Jake M. Alexander<sup>1</sup>

<sup>1</sup>Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland; \*bernat.bramon@gmail.com

## 1 Abstract

2 Species' distribution models have emerged as one of the most influential methodological  
3 advances in ecology and biogeography of the last decades. Useful to understand how pop-  
4 ulations of species will change along environmental gradients, they have become ecologists'  
5 compass to predict the effects of global climate change. That said, uncovering the mecha-  
6 nisms shaping species' realized niches has been one of the main driving forces behind the  
7 development of these models. That is, recent efforts have been often focused on understand-  
8 ing which biotic and abiotic factors are good predictors of species' niches—with an increasing  
9 effort placed in improving the predictive power of the statistical models. However, we still  
10 lack a general understanding of the shape of species' distributions, and much less is known  
11 about how these distributions compare to each other across gradients. Here, we use a set of  
12 Bayesian non-linear models to uncover the shape of species' realized niches. These models  
13 account for all prior knowledge we have regarding their shape, including expert knowledge on  
14 species' environmental preferences and physiology. With this approach, we are able to shed  
15 light on the true shape of empirical species' distributions. Moreover, they allow us to tackle  
16 long-standing hypothesis regarding general biogeographical patterns. In particular, we found  
17 conclusive evidence of the relationship between several properties of distributions, including  
18 the link between species' range size and elevation and their skewness along gradients. Fi-  
19 nally, we are able to shed light on the extent to which some aspects of the shape of observed  
20 realized niches—such as kurtosis and skewness of the distributions—could be intrinsic prop-  
21 erties of species. Overall, our approach offers a useful statistical framework to understand

the shape of species' distributions, and our results provide an unprecedented perspective of the way systems of many species are distributed along environmental gradients.

## Introduction

One of the central goals of ecology is to understand the ways species are distributed across space and time (ref). Over the last two decades, ecologists have developed multiple distribution models to try to untangle the factors that play a role in defining such distributions (Guisan & Zimmermann, 2000). These models estimate species' realized niches using several covariates, including environmental variables (Guisan & Thuiller, 2005), species ecological traits' (Pollock *et al.*, 2012) and phylogenetic relations (Ives & Helmus, 2011). More recently, some of the focus have shifted towards approaches that estimate and account for biotic factors, such as competitive or facilitative relationships between species (Ovaskainen *et al.*, 2017). The idea is that by untangling the ways in which such biotic and abiotic factors shape species' distributions, we can gain a mechanistic understanding on how ecological communities are established and change over time. However, while these factors can increase the predictive performance of some of the models (Norberg *et al.*, 2019), the interpretation of the corresponding parameter estimates has been recently questioned (Harris, 2016; Thurman *et al.*, 2019; Poggiato *et al.*, 2021). This was best illustrated by Blanchet *et al.* (2020), who used basic statistical arguments to highlight the artefactual nature of the link between co-occurrence and species' ecological interactions drawn by some distribution models.

The value of gaining a mechanistic understanding of species' distributions is unquestionable (ref), with several studies highlighting the importance of factors such as biotic interactions and dispersal ability in setting species' range limits (Wisz *et al.*, 2013; Pollock *et al.*, 2014; Neuschulz *et al.*, 2018). That said, a lot can be learned from taking a phenomenological approach, focussing instead on the description of basic properties of species' realized niches. For example, the study of species' range sizes along environmental gradients can reveal general biodiversity patterns that are crucial from a conservation and management perspective (Stevens, 1992). Differences in species' responses to the environment could shed light on how climatic processes and historical contingencies have shaped their distributions (Rohde, 1992; ?). Other properties, such as the skewness of species' distributions, can also reveal

51 general underlying processes regarding species' physiological tolerance to different environ-  
52 mental conditions (Kaufman, 1995). More generally, understanding the shape of species'  
53 realized niches and the extend to which these vary across species is a crucial issue in ecology  
54 and biogeography (ref); however, we do not have an effective way to parsimoniously compare  
55 the realized niches of many species. Indeed, there is no general agreement on the shape of  
56 species' distributions (ref).

57 Many ecological textbooks (Krebs, 1972) assume the shape of species distributions to  
58 be unimodal and symmetric, but some have warned that empirical distributions can take  
59 many different forms (Austin, 1987, 2002). In practice, distribution frameworks often use  
60 logistic regressions with a linear relationship between covariates (but see ? and ?). This  
61 is useful because it simplifies the optimization process, but it comes with some statistical  
62 shortcomings. First and foremost, such response curve and the linear relationship between  
63 covariates often comes with a set of implicit mathematical constraints that might not be bio-  
64 logically justified. From a purely statistical perspective, if all that we are willing to assume  
65 is that species occupy finite geographic ranges—i.e. their probability distributions have finite  
66 variance—the most conservative statistical approach is to model these as a Gaussian distri-  
67 butions (Frank, 2009). Other factors might then condition species distributions to showcase  
68 heavy-tails or a skewed shapes, revealing interesting ecological processes shaping biodiversity  
69 patterns (Austin, 1976; Minchin, 1987). The starting point, nevertheless, should be the one  
70 that makes the fewest assumptions (i.e. the maximum entropy distribution; Frank 2009), and  
71 every new shape will imply a hypotheses on how communities are distributed (D'Amen *et al.*,  
72 2017). Second, the aforementioned structural constraints also limit our ability to include any  
73 prior information to our parameter estimates. Observations on species' geographic varia-  
74 tion and optimal climatic conditions have long been documented, with extensive databases  
75 compiled by botanists and field ecologists documenting basic knowledge on species' realized  
76 niches (e.g. Landolt *et al.* 2010). That said, this information is rarely accounted for in most  
77 modelling approaches, potentially because there is not a straightforward way to feed this in-  
78 formation into the parameters of a linear model (Scherrer & Guisan 2019; but see ter Braak  
79 & Looman 1986; Ovaskainen *et al.* 2017). Finally, and perhaps most importantly, a direct  
80 biological interpretation of parameter estimates in linear models becomes increasingly diffi-  
81 cult as one moves from unimodal and symmetric distributions (ter Braak & Looman, 1986;

82 Jamil & ter Braak, 2013) to skewed distributions (Huisman *et al.*, 1993), making the tests of  
83 hypothesis on global biodiversity patterns particularly challenging. For example, Huisman  
84 *et al.* (1993) proposed several non-linear models to characterize several features of species’  
85 response curves; however, species’ environmental indicator values, range size or distribution  
86 skewness are difficult to understand altogether following these model structures.

87 The field of ecology has quickly moved towards mechanistic and process-based approaches  
88 to understand species’ distributions (Warton *et al.*, 2015). This has resulted in a plethora  
89 of models accounting for several biotic and abiotic factors into the predictions of species  
90 co-occurrence. Here, we instead rethink traditional modelling approaches and develop a  
91 conceptually simple—and yet statistical and computationally complex—statistical frame-  
92 work to revisit some classic hypothesis in ecology and biogeography. In particular, we develop  
93 a Bayesian hierarchical model that accounts for all prior information that we have regard-  
94 ing the distribution of alpine plant species along an elevation gradient in the Swiss Alps,  
95 including expert knowledge on species environmental indicator values, range sizes, and plant  
96 physiology. We start by considering species’ response curves as Gaussian distributed, and  
97 then we adapt our model to allow for skewed and long-tailed distributions. Using this sta-  
98 tistical framework, we are able to compare the basic properties of the realized niches of  
99 multiple species, testing for the existence of general biogeographical patterns. First, we test  
100 for the Rapopor’s rule, which predicts a positive relationship between range size and eleva-  
101 tion (Stevens, 1992). While this pattern has been largely studied for multiple systems and  
102 across gradients (McCain & Knight, 2013); contrasting evidence suggests this rule not to be  
103 pervasive across species (Ribas & Schoereder, 2006; Bhattarai & Vetaas, 2006; McCain &  
104 Knight, 2013). Our results not only allow us to properly test the existence of this geographi-  
105 cal pattern, but they also showcase variation in how different types of species, such as native  
106 or neophytes, might respond to an environmental gradient. Second, we study whether or not  
107 species’ distributions show steeper declines towards stressful conditions, testing the so-called  
108 abiotic stress limitation hypothesis (ref). Normand *et al.* (2009) tested this for vegetation  
109 data using Huisman *et al.*’s statistical models for several independent species, finding no clear  
110 support for such a hypothesis. Our results are able to shed light on this geographical pattern  
111 as well as to highlight the degree to which different species will showcase different levels of  
112 decline towards stressful conditions. Specifically, we are able to link plant physiological traits

113 to the skewness of their distributions. Overall, we use models that are solely constrained  
114 by the empirical information that we truly have regarding our system, relaxing as much as  
115 possible the structural constraints of the statistical framework. Using these models, we are  
116 able to uncover the approximate shape of empirical plant distributions and answer fundamental  
117 questions regarding the way systems of many species are distributed along environmental  
118 gradients.

## 119 **Methods**

### 120 **Empirical data**

121 We studied the distribution of alpine plant communities along an elevation gradient. To do  
122 so, we combined two different datasets: i) one describing the co-occurrence of species across  
123 multiple open grasslands in the Swiss Alps, and ii) an extensive floristic database containing  
124 environmental and physiological traits for all vegetation across Switzerland (Landolt *et al.*,  
125 2010).

#### 126 *Distribution data*

127 We used data describing the distribution of 798 species across 912 sites covering most of the  
128 mountain region of the Western Alps in the Canton de Vaud (Switzerland; Scherrer & Guisan  
129 2019). Each of these sites is a  $8 \times 8$  m plot placed somewhere along an elevation range from  
130 375 m to 3210 m. In all sites, presence/absence data as well as Braun-Blanquet abundance-  
131 dominance classes were recorded for all species. Additionally, we used meteorological data  
132 provided by Scherrer & Guisan (2019), containing multiple variables characterizing the cli-  
133 mate in each site at high spatial resolution (25 m). This dataset was compiled based on  
134 30 years (1961–1990) of records from national weather stations. Since most of the data is  
135 highly correlated, we calculated the main axes of variation of the following scaled variables:  
136 daily minimum, maximum and average temperature; sum of growing degree-days above  $5^{\circ}\text{C}$ ;  
137 mean temperature of wettest quarter; annual precipitation, precipitation seasonality, and  
138 precipitation of driest quarter (see Supplementary Methods; Supplementary Fig. 1).

## Floristic data

To complement the aforementioned distribution data, we used a floristic database of most vegetation across Switzerland. This database was build based on expert knowledge and field experience of botanists and ecologists, and contains information regarding species' environmental preferences and physiological traits. Species' environmental preferences in this database can be used to inform distribution models—e.g. as an informative prior in a Bayesian framework. These are characterized following the ecological indicator values developed by Landolt *et al.* (2010), providing both an estimate of the average conditions in which a species can be found as well as a broad description of their range of variation. These values are provided for a range of 10 climatic variables, including temperature, continentality, light conditions, as well as moisture, acidity and nutrient content of the soil (see a full list and description of the ecological indicators in the Supplementary Table 1; Landolt *et al.* 2010). On the other hand, the information regarding species' physiological traits represents general descriptions of species' growth and life strategies—examples include their growth forms, nature of the storage organs, dispersal ability and pollinator agents. In total, we identify more than 120 binary traits that characterize the physiology of species (see a full list and description of the ecological indicators in the Supplementary Table 1; Landolt *et al.* 2010). Finally, and in addition to species' environmental preferences and physiological traits, the floristic data also contains information on species types (e.g. identifying those species that are neophytes)—and change tendency (e.g. indicating species that have shown decline or increase in their populations over the recent decades). We describe this information in more detail in the Supplementary Table 1.

## Baseline model

There is a long list of model structures well suited to characterize species' distributions (see Norberg *et al.* 2019). As a baseline model, however, we were interested in a hierarchical model that does not make any assumptions regarding the shape of the distributions, and yet explicitly incorporates all information that we have regarding plant's environmental preferences. More specifically, we wanted to account for the climatic indicator values and range of variation registered in the floristic database for all plants in our dataset. These two

values provide basic information regarding plant's optimal environmental conditions and width of their distributions.

### *Response curve*

To choose an appropriate response curve, we first need to agree on what we truly know about the system. Given the prior information that we have about the system, we know that species occupy specific geographic ranges; therefore, we know that their distributions have finite variance. While we could also assume that many other factors might influence species' presence in a given site—e.g. the biotic interactions among species in the site—we do not necessarily have an *a priori* expectation of how exactly these factors will influence the shape of species' distributions. Therefore, for this baseline model, if all that we are willing to assume about species' realized niches is that these have finite variance, the most conservative assumption and the safest bet—i.e. the one with the largest entropy—is that they follow a Gaussian distribution. That is, given the presence/absence or abundance  $y_{ij}$  of any species  $i$  in any given site  $j$ , and an environmental variable  $x_j$ , we define species' responses to the environment as

$$y_{ij} \sim F(p_{ij})$$
$$\log(p_{ij}) = -\alpha_i - \gamma_i(x_j - \beta_i)^2, \quad (1)$$

where  $F$  is the likelihood function, and  $\alpha_i$ ,  $\beta_i^k$ , and  $\gamma_i$  describe amplitude of the probability  $p_{ij}$ , species' average climatic suitability and range of variation along the environmental gradient, respectively. Notice that  $F$  characterizes a Binomial distribution when considering binary data, and it characterizes an ordered categorical likelihood function when we consider Braun-Blanquet abundance-dominance classes as response variables (see the full description of both models in the Supplementary Methods). For the sake of simplicity, we use only one environmental variable to characterize species' probability distribution. That said, this model can easily be generalized to account for multiple predictors (see Supplementary Methods).

## 191 *Model priors*

192 The model structure described above allows us to explicitly incorporate all prior knowledge  
 193 that we have regarding species’ distributions contained in the floristic database. To do so,  
 194 we define the prior distributions for the parameters in model (1) as:

$$\begin{aligned}
 \beta_i &\sim \text{MVNormal}(\hat{\beta}, \Sigma^\beta) \\
 \log(\gamma_i) &\sim \text{MVNormal}(\hat{\gamma}, \Sigma^\gamma) \\
 \log(\alpha_i) &\sim \text{Normal}(\hat{\alpha}, \sigma_\alpha) \\
 \hat{\beta}, \hat{\gamma}, \hat{\alpha} &\sim \text{Normal}(0, 1) \\
 \sigma_\alpha &\sim \text{Exponential}(1)
 \end{aligned} \tag{2}$$

195 where parameters  $\gamma_i$  and  $\beta_i$  are expressed as multivariate normal distributions—i.e. Gaus-  
 196 sian processes—such that  $\Sigma^\beta$  and  $\Sigma^\gamma$  are variance-covariance matrices describing species’  
 197 similarity in terms of their average climatic suitability and range of variation along the dif-  
 198 ferent environmental gradients, respectively. We define these variance-covariance matrices  
 199 as follows:

$$\Sigma_{ij} = \eta \exp(-\rho D_{ij}^2) + \delta_{ij} \sigma, \tag{3}$$

200 where  $\Sigma_{ij}$  characterizes the covariance between any pair of species  $i$  and  $j$ , and  $\delta_{ij}$  is the  
 201 Kronecker delta. Notice that such a covariance structure declines exponentially with the  
 202 square of a distance matrix  $D_{ij}$ , which characterize differences between species computed  
 203 using our prior information. In the floristic database, this information is represented by the  
 204 set of ordinal traits specified for the different species. While there are many different ways  
 205 to turn ordinal data into distance matrices, we choose to use a mixed-membership stochastic  
 206 block model because it allows us to deal with cases of missing data (see Supplementary  
 207 Methods for extended details; [Godoy-Lorite et al. 2016](#)). In each covariance matrix, the  
 208 hyperparameter  $\rho$  determines the rate of decline of the covariance between any two species,  
 209 and  $\eta$  defines its maximum value. The hyperparameter  $\sigma$  describes the additional covariance  
 210 between the different observations for any given species. For all these hyperparameters, we  
 211 choose weekly informative priors such that  $\sigma, \eta \sim \text{Exponential}(1)$  and  $\rho \sim \text{Exponential}(0.5)$ .



## Alternative variance-covariance structures

The model structure defined above allows us to test how different sources of information characterize each of the different parameters. Specifically, we can do this by modifying Eq. (3). For example, imagine that we have multiple matrices  $D^k$  characterizing species' differences along different axis of variation—e.g. two matrices characterizing physiological and environmental traits. One can modify Eq. (3) for a particular parameter—e.g.  $\beta_i$ —such that

$$\Sigma_{ij} = \eta \exp \left( - \sum_k \rho_k D_{ij}^k{}^2 \right) + \delta_{ij} \sigma, \quad (4)$$

where now  $\rho_k$  are separate relevance hyperparameters for each distance matrix in the total variance of  $\beta_i$ .

## Sampling the posterior

We generated the posterior samples for the Bayesian models with the Hamiltonian Monte Carlo algorithm implementation provided by the R package 'rstan' to (Stan Development Team, 2021). Sampling models like the ones described above can be computationally very expensive. This is especially true when using ordered categorical likelihood functions (see Stan Development Team 2021). Therefore, we focus on those species for which we have at least 20 occurrences when modelling binary data, which corresponds to the majority of the results of this work. When using ordinal data, we limit our study to a random subset of a 100 species of those with at least 20 occurrences to facilitate the sampling of the models.

To test the performance of the model as well as our choice of prior distributions, we modelled simulated data and compared the sampled posterior distributions to the data-generating parameters (see Supplementary Methods; Supplementary Fig. 2). Notice that using the link function in Eq. (1) could cause problems when sampling the model, and some adjustments need to be made when specifying the model (see Supplementary Methods and the Code Availability section).

## 236 Modifying the baseline model

237 We proposed a baseline model that is naive regarding how the data is distributed, and yet  
238 accounts for all prior information that we have about the system. Now, we want to modify  
239 this model to test the extent to which empirical species' distributions showcase different  
240 properties, while preserving both the interpretation of the parameter estimates and the prior  
241 information. More specifically, to propose new species' response curves, we follow three key  
242 criteria: (i) the probability distribution must have a defined variance and mean, (ii) the  
243 Gaussian shape must be a special case of the probability distribution, and (iii) there must  
244 be a re-parametrization of the model that allows us to keep the same prior information and  
245 interpretable parameters.

### 246 *Heavy-tail response curve*

247 Distributions with heavy-tails are very common across fields, as they can capture processes  
248 such as seasonality (e.g. in communications patterns; [Malmgren et al. 2008](#)) or some stochastic  
249 events (e.g. in the spread of infectious diseases; [Wong & Collins 2020](#)). Indeed, heavy-tail  
250 distributions are pervasive in ecology; for example, species' dispersal patterns have been  
251 shown to have heavy-tails due to natural variability among individuals ([Petrovskii et al.,](#)  
252 [2009](#)). Therefore, one might expect these properties to also emerge in alpine communities,  
253 where seasonality and dispersal patterns are crucial factors determining species' distribu-  
254 tions. To accommodate this feature into our baseline model, we could consider a response  
255 curve that follows a generalized error distribution, since the Gaussian shape is a special case  
256 of it and contains parameter that regulates the level of kurtosis (Supplementary Figure XX).  
257 In particular, we can adapt Eq. (1) to present this non-linear form as follows:

$$\log(p_{ij}) = -\alpha_i - \gamma'_i |x_j - \beta_i|^{\nu_i}, \quad (5)$$

258 where  $\gamma'_i = g(\gamma_i, \nu_i)$ , and  $\nu_i$  is a parameter that describes the kurtosis of the distribution,  
259 which we define as  $\nu_i \in (1, \infty)$ . Following this, we choose an adaptive prior for this set  
260 of new parameter such that  $\log(\nu_i - 1) \sim \text{Normal}(\hat{\nu}, \sigma_\nu)$ , where  $\hat{\nu} \sim \text{Normal}(0, 1)$  and  
261  $\sigma_\nu \sim \text{Exponential}(2)$ . Given the relationship between  $\gamma'_i$  and  $\gamma_i$ , we can re-parametrize the

model and follow Eq. (2) to define the corresponding prior distributions (see Supplementary Table 2; Code Availability section; Nadarajah 2005). Notice that the Gaussian distribution will naturally emerge when  $\nu_i = 2$ .

Alternatively, we could have used other distributions that present heavy-tails and fulfil the selection criteria described above. For example, the non-standardized Student's t-distributions is an interesting distribution because, as opposed to the generalized error distribution, it allows for heavy tails without generating a cusp at the center (see Supplementary Fig. XX). However, we avoided using the non-standardized Student's t-distributions because it does not allow for tails that are lighter than normal (e.g.  $\nu_i > 2$  in Eq. 5; Supplementary Fig. XX), and the sampling of the model is somewhat more challenging (ref).

### *Skewed response curve*

When species experience abiotic or biotic pressures that increase or decrease along an environmental gradient, one might expect their distributions to be skewed in one direction. Likewise, this same skewed shape can also be the result of asymmetric environmental tolerance. One way to accommodate this feature to our models is by considering a skewed normal distribution (Supplementary Figure XX). As for the two cases described above, the Gaussian is a special case of this distribution, and it contains a parameter that controls for the level and direction of 'skewness'. Importantly, this distribution presents normal-like tails; therefore, the added skewness does not make additional assumptions regarding how species are distributed along the gradient. To test for the existence of this feature, we modified the species' response curve in Eq. (1) as

$$p_{ij} = \hat{p}_{ij} \left[ 1 + \operatorname{erf} \left( \lambda_i (x_j - \beta'_i) \sqrt{\frac{\gamma'_i}{2}} \right) \right]$$

$$\log(\hat{p}_{ij}) = -\alpha'_i - \gamma'_i (x_j - \beta'_i)^2, \quad (6)$$

where  $\alpha'_i = q_1(\alpha_i, \gamma'_i, \lambda_i)$ ,  $\beta'_i = q_2(\gamma'_i, \lambda_i)$ ,  $\gamma'_i = q_3(\gamma_i, \lambda_i)$ , and  $\lambda_i$  is a parameter that describes the skewness of the distribution, which we define as  $\lambda_i \in (-10, 10)$ . Notice that these boundary conditions for  $\lambda_i$  are arbitrarily set for computational purposes, as this parameter is theoretically not bounded. The function 'erf' is the error function (see Ashour & Abdelhameed 2010). We choose an adaptive prior such that  $\lambda_i \sim \text{Normal}(\hat{\lambda}, \sigma_\lambda)$ , where  $\hat{\lambda} \sim$

Normal(0, 1) and  $\sigma_\lambda \sim \text{Exponential}(1)$ . This model can be re-parametrized following  $q_1$ ,  $q_2$  and  $q_3$ , and set the rest of the prior distributions as described for the baseline model (see Supplementary Table 2; Code Availability section). In this case, the Gaussian distribution is a special case of Eq. (6) when  $\lambda_i = 0$  (Ashour & Abdel-hameed, 2010).

One way to test the extend to which species' physiological traits inform the skewness of the distribution of any species  $i$  is by characterizing  $\lambda_i$  as a Gaussian process. That is, we can consider the prior distribution for  $\lambda_i$  as a multivariate normal with a variance covariance matrix  $\Sigma^\lambda$ . As described before, this variance-covariance matrix can be estimated by the model following Eq. (3), and its structure can shed light on how informative the physiological traits are. Likewise, we could instead use Eq. (4) and test the effectiveness of other prior information in explaining the skewness of species' distributions.

#### *Heavy-tail and skewed response curve*

Finally, one could also consider a response curve with both kurtosis and skewness. A convenient way to achieve this is by considering a response curve that follows a skewed generalized error distribution (also see XX). This is a combination of the two distributions described above, containing two parameters that control for both the level and direction of kurtosis and skewness. The skewed generalized error distribution can be considered by modifying the species' response curve in Eq. (1) as

$$\log(p_{ij}) = -\alpha_i - \left( \frac{\gamma'_i |x_j - \beta'_i|}{1 + \lambda_i \text{sgn}(x_j - \beta'_i)} \right)^{\nu_i}, \quad (7)$$

where  $\gamma'_i = f_1(\gamma_i, \nu_i, \lambda_i)$ ,  $\beta'_i = f_2(\gamma_i, \beta_i, \nu_i, \lambda_i)$ ,  $\nu_i$  and  $\lambda_i$  are parameters that describe the kurtosis and skewness of the distribution, respectively.  $\text{sgn}(x)$  is the sign function. We define  $\nu_i$  and its prior as in Eq. 5. However,  $\lambda_i$  is now defined as  $\lambda_i \in (-1, 1)$ , with an adaptive prior such that  $\text{logit}(2\lambda_i - 1) \sim \text{Normal}(\hat{\lambda}, \sigma_\lambda)$ , where  $\hat{\lambda} \sim \text{Normal}(0, 1)$  and  $\sigma_\lambda \sim \text{Exponential}(1)$ . Again, we can re-parametrize the model following  $f_1$  and  $f_2$ , and set the rest of the prior distributions as for the baseline model (see Supplementary Table 2; Code Availability section). Notice that the generalized error distribution (Eq. 5) and the skew normal distribution (Eq. 6) are special cases of Eq. (7) when  $\lambda_i = 0$  and  $\nu_i = 2$ , respectively.

## Results

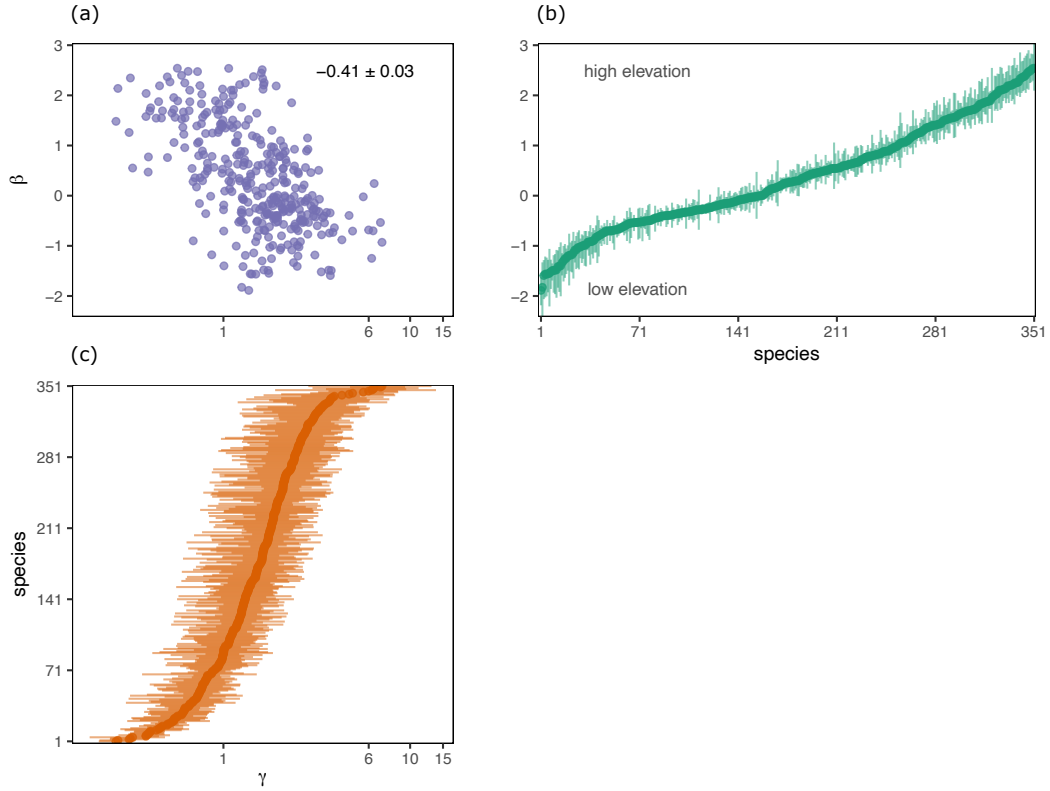
We studied the distribution data to characterize species' realized niches. To do so, we selected the main axis of variation of all environmental variables (Supplementary Fig. XX). Then, using the presence and absence of species across sites as the response variable, we sampled the posterior distributions of the baseline model, accounting for the information in the floristic database regarding species' indicator values and range of variation. Studying the relationship between the mean and variance of the distributions, we found clear evidence of the Rapoport's rule. Specifically, we found that species' range sizes decline with elevation and temperature (i.e.  $\beta_i$  and  $\gamma_i$  in the baseline model are negatively correlated; Fig. 1). The relationship was also found when using instead ordinal data (Supplementary Fig. XX), and it does not seem to depend on species' type or reflect species' abundance change tendency over the years (Fig. 2). This pattern was not present along the second axis of variation of the environmental variables (Supplementary Fig. XX). The comparison between the other parameter estimates revealed additional, somewhat more expected, relationships. In particular, we found the amplitude of distributions to be positively and negatively correlated with the mean and the variance along the elevation gradient, respectively (Supplementary Fig. XX).

Maintaining the symmetry of species' distributions, we then allowed the kurtosis—or shape of the tails—of these to vary in different ways. To do so, we changed the response curve of our Bayesian model to follow either a generalized error distribution (Eq. 5) or a non-standardized Student's t-distribution (Eq. ??). Studying the WAIC values, both non-linear regressions outperformed the baseline model (Supplementary Fig. XX)<sup>†</sup>. In the case of a generalized error distribution, we found the parameter controlling for the kurtosis to be centred around  $\nu_i \sim 2$ , which corresponds to a distribution with a Gaussian shape (Fig. 3). Similarly, we found the shape parameter in the non-standardized Student's t-distribution to present large values, also approximating the normal case<sup>†</sup>. However, in both cases, these parameter estimates displayed a lot of variation between species, which indicates that the shape of the tails of the distribution is species-specific (Supplementary Fig. XX).

Finally, we studied the skewness of species distributions using the skewed response curve in

---

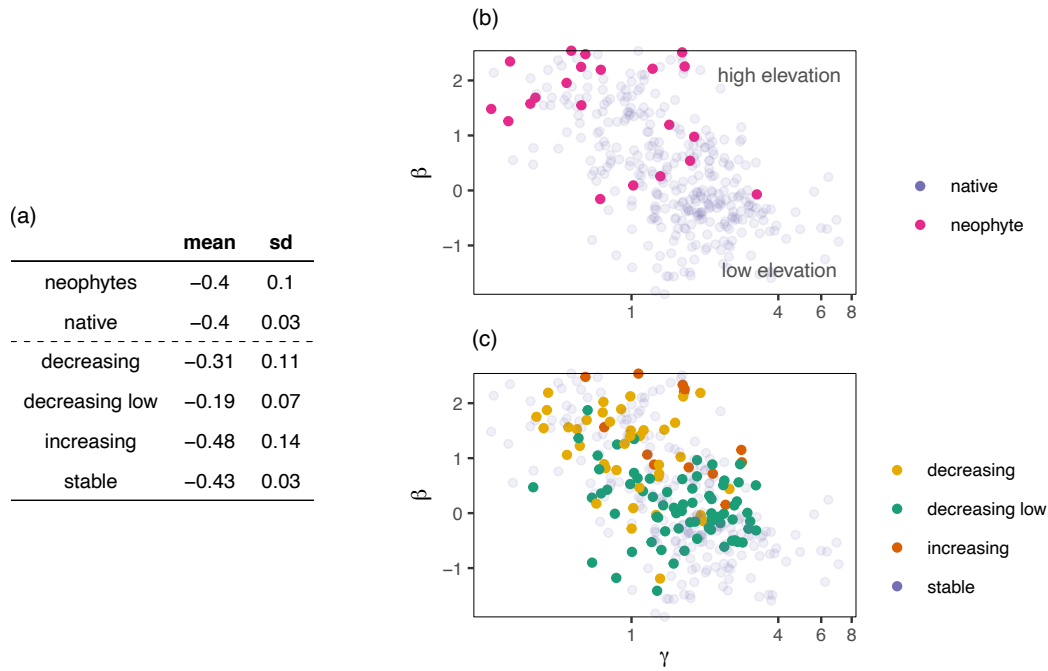
<sup>†</sup> I am still waiting on the results for this (currently running in the cluster), and this is just my prior expectation based on what I've seen in some of the other models I've been working with.



**Figure 1:** Relationship between mean and variance of species' distributions. Posterior distributions for parameters  $\beta_i$  and  $\gamma_i$  from Eq. (1) across species, and the relationship between them. Panel (a) describes the relationship between range size and elevation. Every dot represents the relationship between the mean values for the  $\beta_i$  and  $\gamma_i$  estimates of the different species. The value in the top-right corner of the plot displays the Pearson's correlation between these parameters calculated across samples of the posterior distributions. Panel (b) describes the  $\beta_i$  posterior distribution estimated for all species. Panel (c) describes the  $\gamma_i$  posterior distribution estimated for all species. In (b) and (c), the points represent the mean of the posterior distributions, and the corresponding lines characterize the 89% confidence intervals.

Eq. (6). Based on the estimates for the WAIC values and Akaike weights, this model clearly outperformed the rest (Supplementary Fig. XX), which sheds light on the naturally skewed nature of species' distributions. Studying the mean value of the skewness across species (i.e. parameter  $\hat{\lambda}$ ), we found that species' distributions showed steeper declines towards lower elevations (Fig. 3). Moreover, species' physiology seems to strongly shape this parameter, which suggest that distribution skewness is an intrinsic property of species (New Fig. XX)<sup>†</sup>.

[Note: Reports on the hyperparameters describing the different covariance matrices missing at the moment. I think it's interesting, and I will certainly end up adding things here about this. I will also add a third figure showing how lambda varies with the corresponding



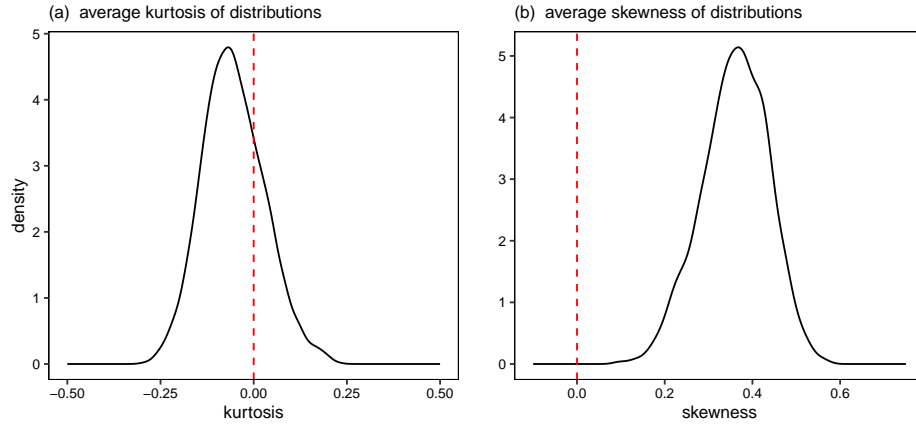
**Figure 2:** Universality of the relationship between mean and variance of species’ distributions. Comparison between how different types of species are mapped in Fig. 1a. Panel (a) describes the correlation coefficient between  $\beta_i$  and  $\gamma_i$  for each type of species. Panel (b) shows the differences between neophytes and native species in the way these are distributed along the environmental gradient. Panel (c) shows the same differences for species that have decreased, decreased in low elevations, increase and remain stable over the last decades (see Supplementary Table 1 for further details). **I might move this to the Supplementary Information.**

variance-covariance matrix of species’ environmental and physiological traits.]

## Discussion

Structure of the discussion section:

1. Summary of results.
2. Proper test of Rapopor’s hypothesis. Different species follow different biogeographical patterns.
3. Proper test of skewed towards high altitude. Is species’ physiology informative to explain the pattern?
4. What is the true shape of species’ distributions? These display heavy-tail and skewed



**Figure 3:** Average kurtosis and skewness of species' distributions. Calculated using the posterior distributions of parameters  $\hat{\nu}$  and  $\hat{\lambda}$  from the models (see Supplementary Table XX), the two panels describe the average (a) kurtosis and (b) skewness of distributions, respectively. Panel (a) displays the results obtained by using a response curve that follows a generalized error distribution. Panel (b) displays the results obtained by using a response curve that follows a skewed normal distribution. In both case, the red dotted line indicates the conditions by which species are normally distributed along the environmental axis.

distributions.

5. Future directions. Missing bimodal curves. Using this information to understand where jSDMs estimate interactions between species. Further test of the ability of traits to predict those parameter estimates.



## References

- Ashour, S. K. & Abdel-hameed, M. A. (2010). Approximate skew normal distribution. *Journal of Advanced Research*, 1, 341–350.
- Austin, M. (1976). On non-linear species response models in ordination. *Vegetatio*, 33, 33–41.
- Austin, M. P. (1987). Models for the analysis of species’ response to environmental gradients. *Vegetatio*, 69, 35–45.
- Austin, M. P. (2002). Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecological Modelling*, 157, 101–118.
- Bhattarai, K. R. & Vetaas, O. R. (2006). Can Rapoport’s rule explain tree species richness along the Himalayan elevation gradient, Nepal? *Diversity and Distributions*, 12, 373–378.
- Blanchet, F. G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, 23, 1050–1063.
- D’Amen, M., Rahbek, C., Zimmermann, N. E. & Guisan, A. (2017). Spatial predictions at the community level: From current approaches to future frameworks. *Biological Reviews*, 92, 169–187.
- Frank, S. A. (2009). The Common Patterns of Nature. *Journal of evolutionary biology*, 22, 1563–1585.
- Godoy-Lorite, A., Guimerà, R., Moore, C. & Sales-Pardo, M. (2016). Accurate and scalable social recommendation using mixed-membership stochastic block models. *Proceedings of the National Academy of Sciences*, 113, 14207–14212.
- Guisan, A. & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8, 993–1009.
- Guisan, A. & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147–186.
- Harris, D. J. (2016). Inferring species interactions from co-occurrence data with Markov networks. *Ecology*, 97, 3308–3314.

391 Huisman, J., Olff, H. & Fresco, L. F. M. (1993). A hierarchical set of models for species  
392 response analysis. *Journal of Vegetation Science*, 4, 37–46.

393 Ives, A. R. & Helmus, M. R. (2011). Generalized linear mixed models for phylogenetic  
394 analyses of community structure. *Ecological Monographs*, 81, 511–525.

395 Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*, vol. 846. John Wiley & Sons.

396 Jamil, T. & ter Braak, C. J. F. (2013). Generalized linear mixed models can detect unimodal  
397 species-environment relationships. *PeerJ*, 1, e95.

398 Kaufman, D. M. (1995). Diversity of New World Mammals: Universality of the Latitudinal  
399 Gradients of Species and Bauplans. *Journal of Mammalogy*, 76, 322–334.

400 Krebs, C. J. (1972). *Ecology: The Experimental Analysis of Distribution and Abundance/by*  
401 *Charles J. Krebs*. 4th edn. Harper & Row, New York.

402 Landolt, E., Bäumler, B., Ehrhardt, A., Hegg, O., Klötzli, F., Lämmler, W., Nobis, M.,  
403 Rudmann-Maurer, K., Schweingruber, F. H., Theurillat, J.-P., Urmi, E., Vust, M. &  
404 Wohlgemuth, T. (2010). *Flora indicativa: Ökologische Zeigerwerte und biologische Kennze-*  
405 *ichen zur Flora der Schweiz und der Alpen*. Haupt, Bern. ISBN 978-3-258-07461-0.

406 Malmgren, R. D., Stouffer, D. B., Motter, A. E. & Amaral, L. A. N. (2008). A Poissonian  
407 explanation for heavy tails in e-mail communication. *Proceedings of the National Academy*  
408 *of Sciences*, 105, 18153–18158.

409 McCain, C. M. & Knight, K. B. (2013). Elevational Rapoport’s rule is not pervasive on  
410 mountains. *Global Ecology and Biogeography*, 22, 750–759.

411 Minchin, P. R. (1987). An evaluation of the relative robustness of techniques for ecological  
412 ordination. *Vegetatio*, 69, 89–107.

413 Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics*, 32,  
414 685–694.

415 Neuschulz, E. L., Merges, D., Bollmann, K., Gugerli, F. & Böhning-Gaese, K. (2018). Bi-  
416 otic interactions and seed deposition rather than abiotic factors determine recruitment at  
417 elevational range limits of an alpine tree. *Journal of Ecology*, 106, 948–959.

418 Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo,  
 419 M. B., Dallas, T., Dunson, D., Elith, J., Foster, S. D., Fox, R., Franklin, J., Godsoe, W.,  
 420 Guisan, A., O'Hara, B., Hill, N. A., Holt, R. D., Hui, F. K. C., Husby, M., Kålås, J. A.,  
 421 Lehtikainen, A., Luoto, M., Mod, H. K., Newell, G., Renner, I., Roslin, T., Soininen, J.,  
 422 Thuiller, W., Vanhatalo, J., Warton, D., White, M., Zimmermann, N. E., Gravel, D. &  
 423 Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species  
 424 distribution models at species and community levels. *Ecological Monographs*, 89, e01370.

425 Normand, S., Treier, U. A., Randin, C., Vittoz, P., Guisan, A. & Svenning, J.-C. (2009).  
 426 Importance of abiotic stress as a range-limit determinant for European plants: Insights  
 427 from species responses to climatic gradients. *Global Ecology and Biogeography*, 18, 437–449.

428 Ovaskainen, O., Tikhonov, G., Norberg, A., Blanchet, F. G., Duan, L., Dunson, D., Roslin,  
 429 T. & Abrego, N. (2017). How to make more out of community data? A conceptual  
 430 framework and its implementation as models and software. *Ecology Letters*, 20, 561–576.

431 Petrovskii, S., Morozov, A., Taylor, A. E. P. D. & DeAngelis, E. D. L. (2009). Dispersal in  
 432 a Statistically Structured Population: Fat Tails Revisited. *The American Naturalist*, 173,  
 433 278–289.

434 Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J. S. & Thuiller, W. (2021).  
 435 On the Interpretations of Joint Modeling in Community Ecology. *Trends in Ecology &*  
 436 *Evolution*.

437 Pollock, L. J., Morris, W. K. & Vesk, P. A. (2012). The role of functional traits in species  
 438 distributions revealed through a hierarchical model. *Ecography*, 35, 716–725.

439 Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk,  
 440 P. A. & McCarthy, M. A. (2014). Understanding co-occurrence by modelling species  
 441 simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and*  
 442 *Evolution*, 5, 397–406.

443 Ribas, C. R. & Schoereder, J. H. (2006). Is the Rapoport effect widespread? Null models  
 444 revisited. *Global Ecology and Biogeography*, 15, 614–624.

445 Rohde, K. (1992). Latitudinal Gradients in Species Diversity: The Search for the Primary  
 446 Cause. *Oikos*, 65, 514–527.

447 Scherrer, D. & Guisan, A. (2019). Ecological indicator values reveal missing predictors of  
448 species distributions. *Scientific Reports*, 9, 1–8.

449 Stan Development Team (2021). RStan: The R interface to Stan.

450 Stan Development Team (2021). Stan Modeling Language Users Guide and Reference Man-  
451 ual.

452 Stevens, G. C. (1992). The Elevational Gradient in Altitudinal Range: An Extension of  
453 Rapoport’s Latitudinal Rule to Altitude. *The American Naturalist*, 140, 893–911.

454 ter Braak, C. J. F. & Looman, C. W. N. (1986). Weighted averaging, logistic regression and  
455 the Gaussian response model. *Vegetatio*, 65, 3–11.

456 Thurman, L. L., Barner, A. K., Garcia, T. S. & Chestnut, T. (2019). Testing the link  
457 between species interactions and species co-occurrence in a trophic network. *Ecography*,  
458 42, 1658–1670.

459 Warton, D. I., Blanchet, F. G., O’Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C. &  
460 Hui, F. K. C. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends*  
461 *in Ecology & Evolution*, 30, 766–779.

462 Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., Dormann,  
463 C. F., Forchhammer, M. C., Grytnes, J.-A., Guisan, A., Heikkinen, R. K., Høye, T. T.,  
464 Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.-C., Normand, S., Öckinger, E., Schmidt,  
465 N. M., Termansen, M., Timmermann, A., Wardle, D. A., Aastrup, P. & Svenning, J.-C.  
466 (2013). The role of biotic interactions in shaping distributions and realised assemblages of  
467 species: Implications for species distribution modelling. *Biological Reviews*, 88, 15–30.

468 Wong, F. & Collins, J. J. (2020). Evidence that coronavirus superspreading is fat-tailed.  
469 *Proceedings of the National Academy of Sciences*, 117, 29416–29418.