

Virtual experiment with two species

Contents

1 Virtual experiment with two species	1
1.1 Perfect knowledge model : generating response values.	1
1.2 Partial knowledge model : interpreting the response values with only one explanatory variable.	5
1.3 Spatial autocorrelation of individual response	15
1.4 Similarity between conspecific individuals compared to heterospecific individuals and consequences for species coexistence.	16
2 Code implementation	17
References	18

1 Virtual experiment with two species

Intraspecific variability (IV) is often seen as an intrinsic property of individuals, associated with genetics, which is unstructured in space and time. Here, we investigated whether it can result from the species responses to the spatial variations of the environment. Our hypothesis was that the individuals can be clones, i.e. have no genetic differences between them, and still have different measured attributes because the environment in which they strive varies, as shown in the previous analysis of a clonal dataset. The response of an individual results from the integrated effects of the environmental conditions which the individual has encountered during its life (local environmental variation, also called microsite effect) and its genetic features. Moreover, the differences between individuals due to environmental variation does not imply a broader overlap of the species niches. Therefore, this type of IV has no direct effect on species coexistence.

Here, we designed and conducted a virtual experiment that aims at illustrating that IV, or “individual effects,” can result from variation in unobserved environmental variables (Clark *et al.* 2007), therefore accounting for multidimensional species differences which cannot be observed on a few niche axes.

1.1 Perfect knowledge model : generating response values.

To do so, we first considered a “perfect-knowledge” model that depicts the ecological response (e.g. growth) of individual clones for two different species as a function of a set of environmental variables. This model represents the perfect knowledge of the process as it occurs in the field and thus includes no residuals :

$$Y_{ijt} = \beta_{0j} + \beta_{1j} * \ln(X_{1ijt}) + \beta_{2j} * X_{2ijt} + \dots + \beta_{Nj} * X_{Nijt} \quad \text{“Perfect knowledge model”}$$

where $1 \leq i \leq 500$ are the individuals ; $1 \leq j \leq 2$ are the species ; Y is a response variable, for instance growth ; X_1 to X_N are explanatory variables, for instance environmental variables, that can vary with time t ; and the $\beta_{k,i}$, $1 \leq k \leq N$ depict the species-specific responses to these environmental variables. As we consider

clones, individuals of the same species respond the same way to environmental variables, and variation in Y among individuals within species is due to differences in the environment where each individual thrives. The first explanatory variable was natural log-transformed to represent a log-log relationship as would be the relationship between tree growth and light for example.

We fixed the species parameters of the “Perfect knowledge model” as follows, using 10 environmental variables ($N=10$) :

Parameters of species 1 : $\beta_0 = 0.25$, $\beta_1 = 0.15$, β_2 to β_6 were chosen randomly between -0.05 and 0.05 and β_7 to β_{10} were chosen randomly between 0 and 0.05.

Parameters of species 2 : $\beta_0 = 0.2$, $\beta_1 = 0.1$, β_2 to β_{10} were chosen the same way as for species 1.

The difference between the species is imposed by those parameters. Here, species 1 was more competitive on average thanks to its higher intercept (β_0) and response to the first environmental variable (β_1). The first environmental variable (X_1) has a higher weight in the computation of the response variable, as would be the most limiting factor for the response variable.

To account for genetic variability in our generated data, we added an IV in species parameters by sampling each individual parameter in a normal distribution centered on the species mean parameter and with a standard deviation of a quarter of the species mean parameter.

To represent the spatialised environment, we build a 2D matrix of dimension 500×500 for each environmental variable at a time t_0 , by randomly generating them with spatial autocorrelation. We here considered that all environmental variables were independent. Each variable was simulated by using the gstat R package (Pebesma 2004 ; Gräler *et al.* 2016), enabling to create autocorrelated random fields. A spherical semivariogram model was used for each of the ten environmental variables, with a mean of 0 for each explanatory variable (beta = 0), a sill of 1 (psill = 1) for each and a range of 50 (range = 50).

We then considered that Y had been measured at two times, t_0 and t_1 , for each individual as it would be under periodic forest plot censuses for instance. A pair of coordinates within this spatialised environment was randomly assigned to each of the 500 individuals. We considered that some of the environmental variables (light, temperature, humidity, nutrient availability for instance) had changed between t_0 and t_1 and others had not (slope, altitude for instance). For the first environmental variable which had the stronger impact on Y values and another randomly drawn environmental variable, we computed values at t_1 as $t_0 + \epsilon$, $\epsilon \sim \mathcal{N}(0, 0.1)$ (they randomly increase or decrease). For two other environmental variables randomly drawn, $X_{t_1} = X_{t_0} + |\epsilon|$, $\epsilon \sim \mathcal{N}(0, 0.1)$ (they increase) and for two other environmental variables, $X_{t_1} = X_{t_0} - |\epsilon|$, $\epsilon \sim \mathcal{N}(0, 0.1)$ (they decrease).

This led to two repeated measurements of Y for each individual of each species, i.e. 2000 values of $Y_{i,j,t}$, with $i = [1; 2]$; $j = [1, 50]$; $t = [t_0; t_1]$.

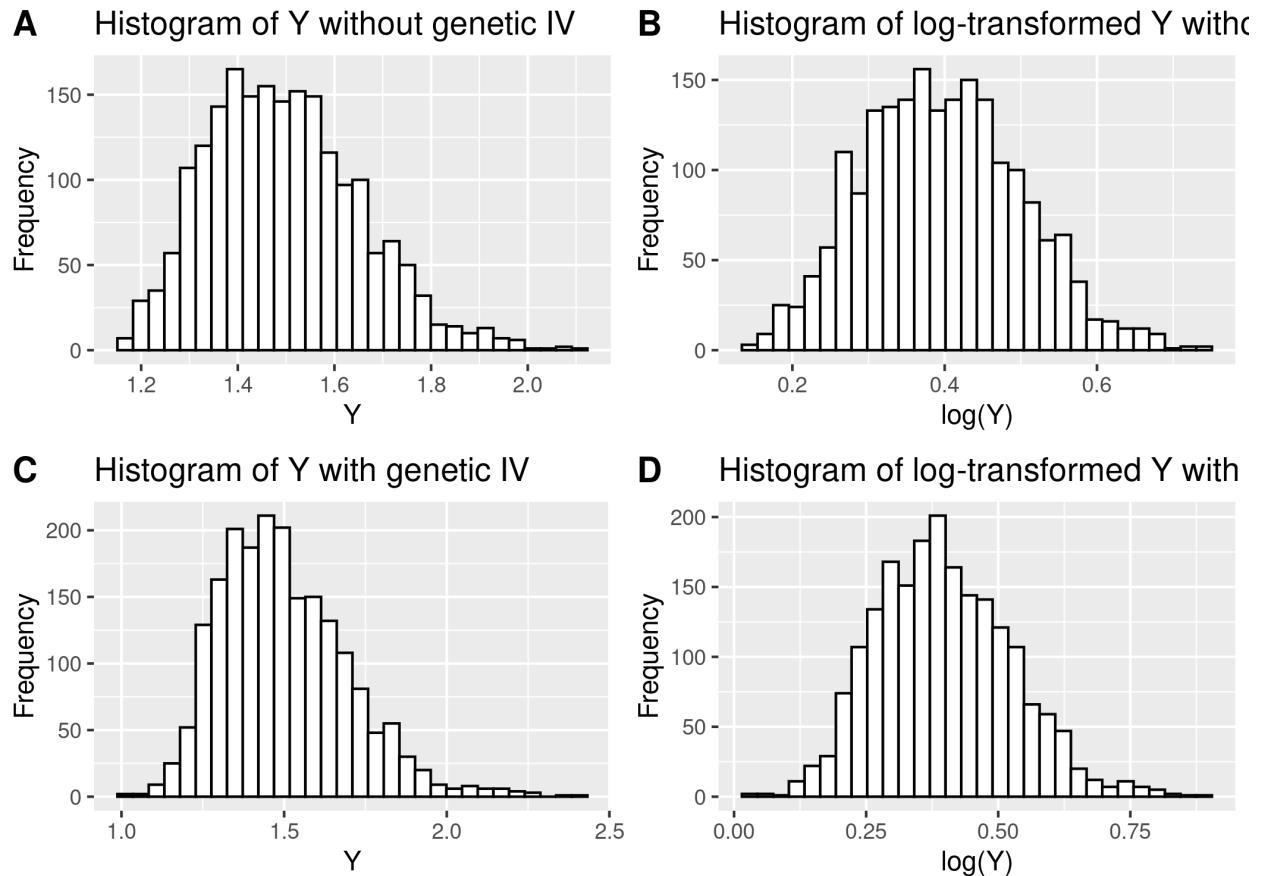


Figure 1.1: Histogram of raw and log-transformed Y with and without genetic IV

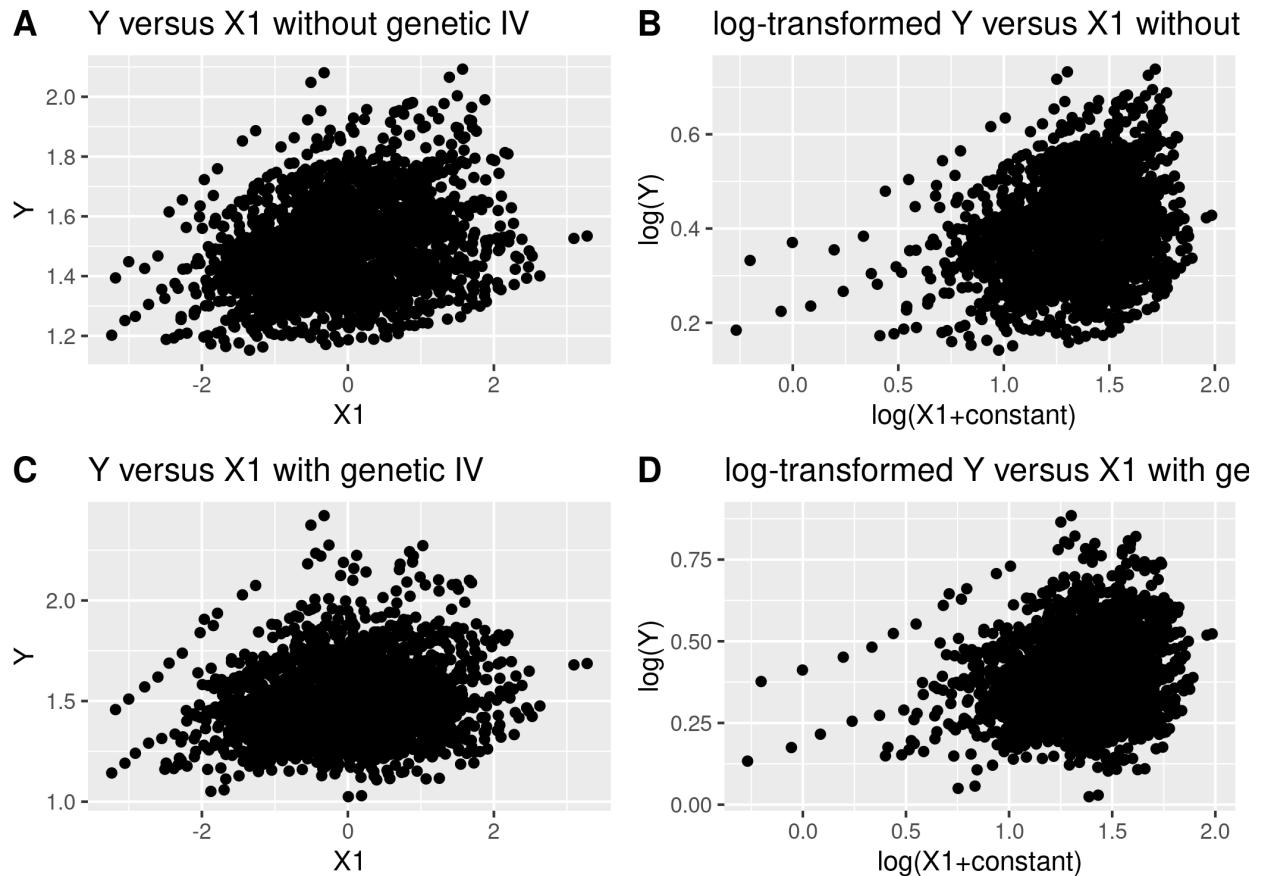


Figure 1.2: Raw and log-transformed Y versus X_1 plots with and without genetic IV

Figure 1.1 shows the distribution of the data and Figure 1.2 shows the relationship between the response Y and the environmental variable X_1 .

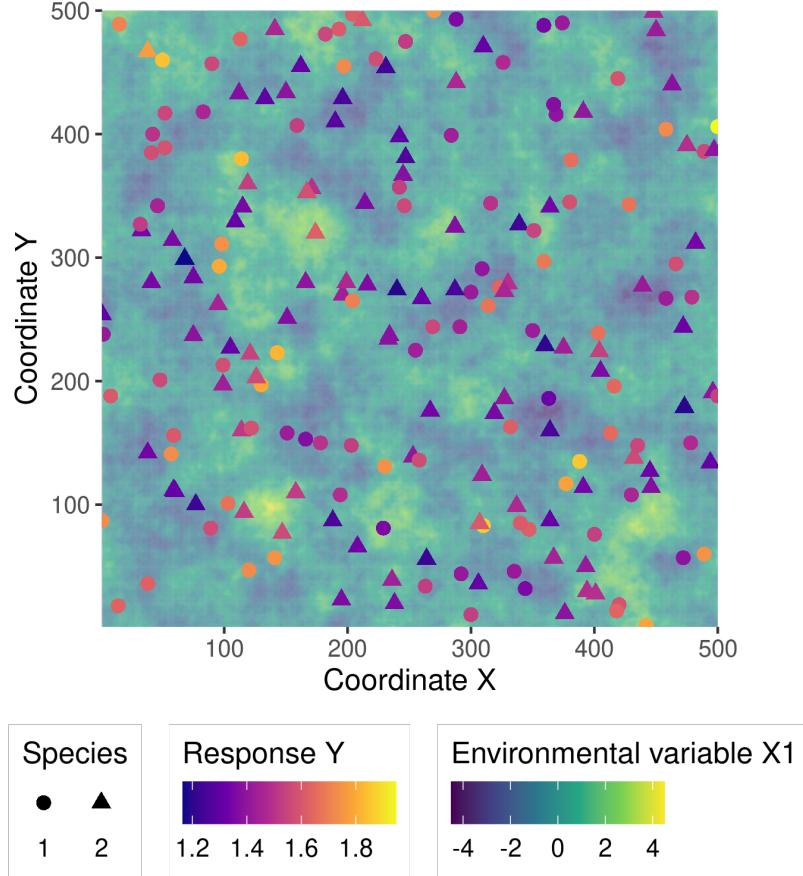


Figure 1.3: Virtual landscape representing X_1 and the individual with their respective Y value.

Figure 1.3 shows that microsite effects, which are the effects of the local multidimensional environment on the response variable, can result in local reversals of the competitive hierarchy between species. Indeed, the local outcome of competition can be opposite to the mean hierarchy : at one point of the space-time, an individual of Species 1 can overcome an individual of Species 2, whilst Species 1 is, on average, the fittest of the two species. Consequently, niche multidimensionality and variation of the environment in space allow the coexistence of Species 1 and Species 2.

1.2 Partial knowledge model : interpreting the response values with only one explanatory variable.

These two virtual datasets (with and without genetic variability) were then analysed with a “Partial knowledge model,” which represents the model derived by ecologists when using datasets derived from an incomplete characterisation of the environment : only a few variables (here only one, X_1) are actually measured and taken into account.

$$\ln(Y_{ijt}) = [\beta'_{0j} + b_{0i}] + \beta'_{1j} \ln(X_{1ijt}) + \varepsilon_{ijt} \quad \text{“Partial knowledge model”}$$

Priors :

$$\varepsilon_{ijt} \sim \mathcal{N}(0, V_j), j = [1, 2], \text{iid}$$

$$\beta'_{kj} \sim \mathcal{N}_2(0, 1), k = [0, 1], j = [1, 2], \text{iid}$$

$$b_{0i} \sim \mathcal{N}(0, V_{bj}), i = [1, 500], j = [1, 2], \text{iid}$$

Second level priors :

$$V_j \sim \mathcal{IG}(10^{-3}, 10^{-3}), j = [1, 2], \text{iid}$$

$$V_{bj} \sim \mathcal{IG}_2(10^{-3}, 10^{-3}), j = [1, 2], \text{iid}$$

This model includes a random individual effect on the intercept (b_{0i}) to account for the variability among individuals within species. We ran this model twice, with the datasets generated with and without genetic IV. These models were fitted in a hierarchical Bayesian framework using the brms package (Bürkner 2017, 2018) with 100000 iterations, a warming period of 50000 iterations and a thinning of 50. We obtained 1000 estimates per parameter.

We visualised the convergence and the results of the models thanks to trace and density plots.

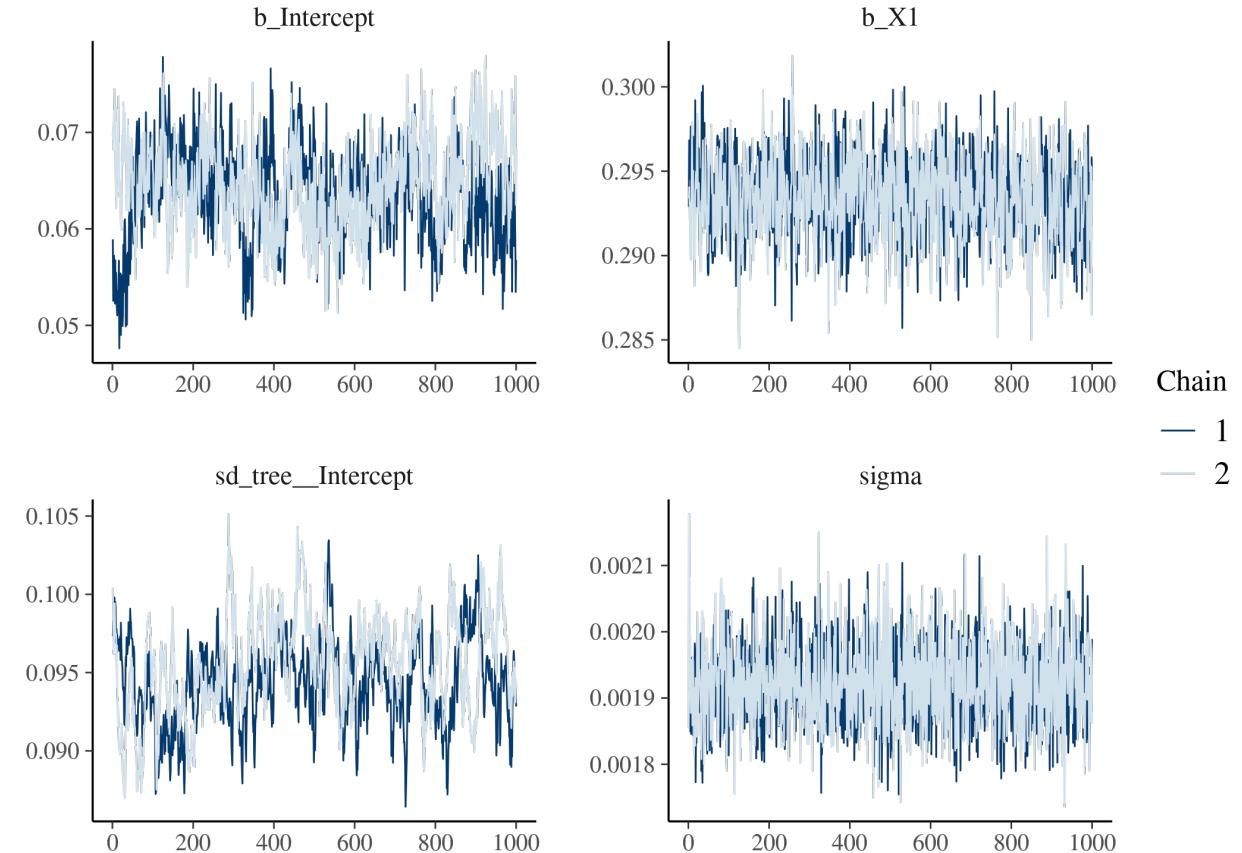


Figure 1.4: Trace of the posterior of the model for Species 1 without genetic IV

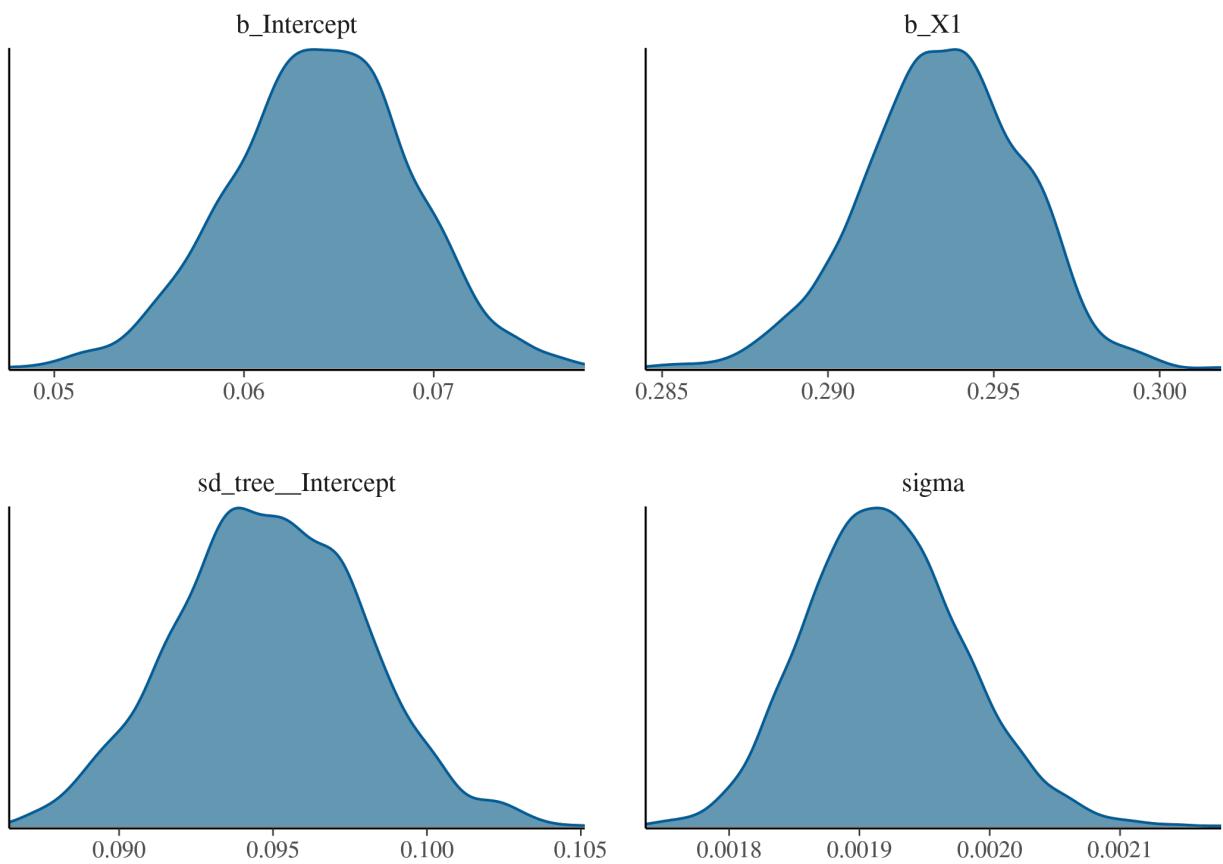


Figure 1.5: Density of the posterior of the model for Species 1 without genetic IV

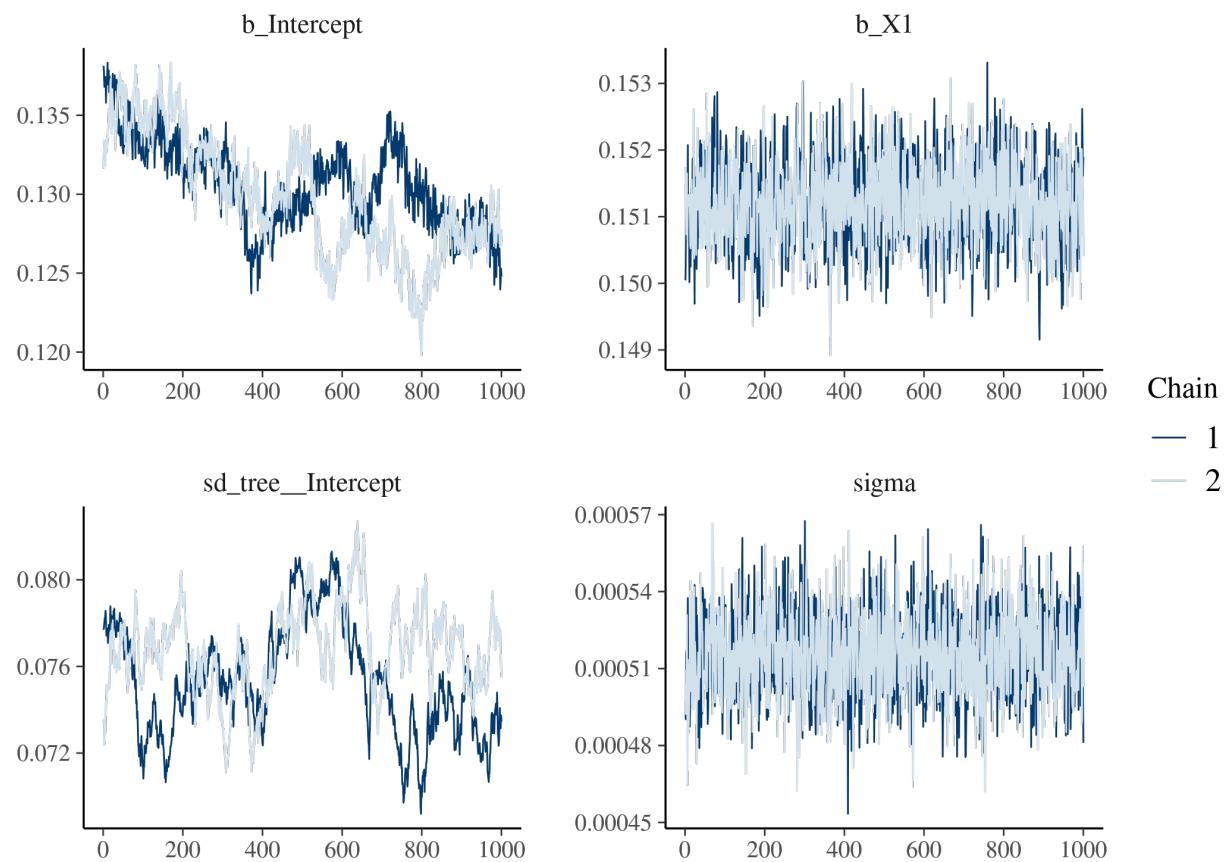


Figure 1.6: Trace of the posterior of the model for Species 2 without genetic IV

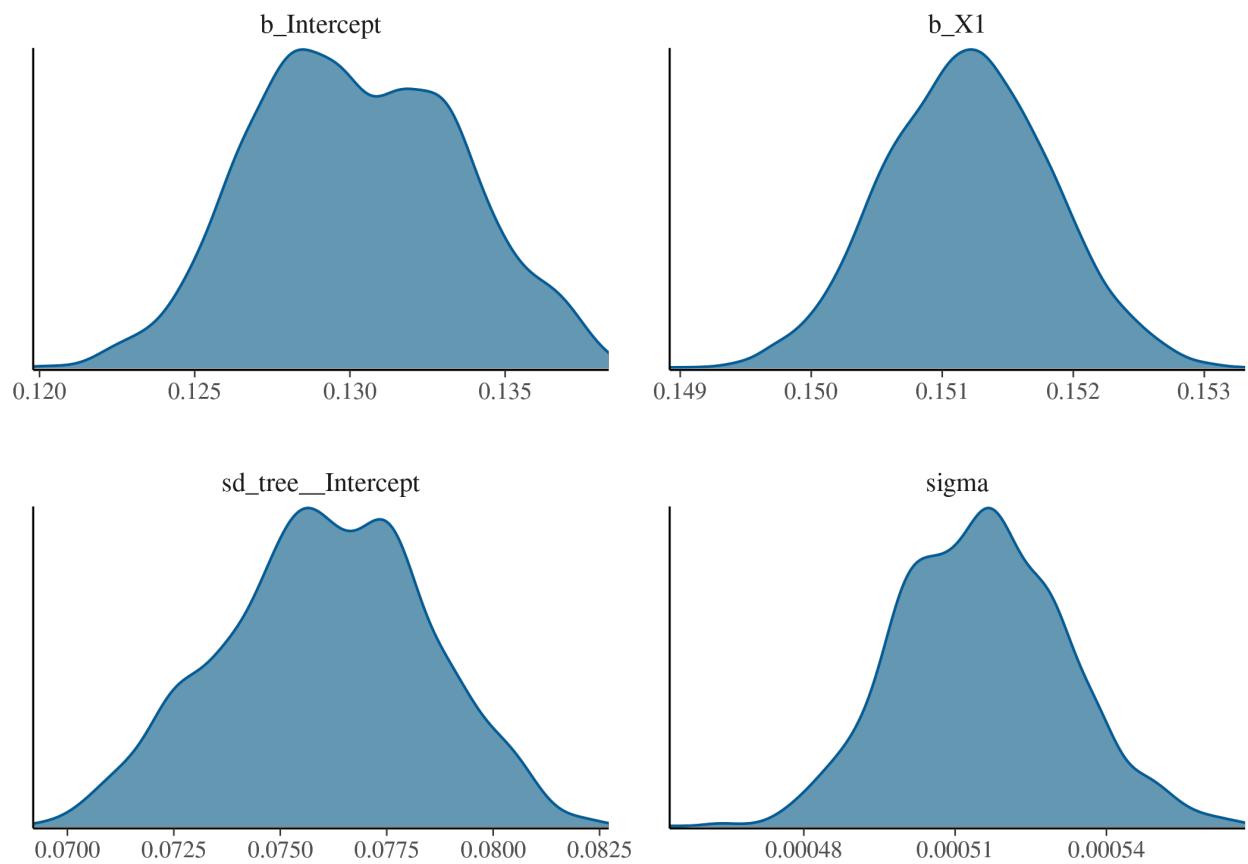


Figure 1.7: Density of the posterior of the model for Species 2 without genetic IV

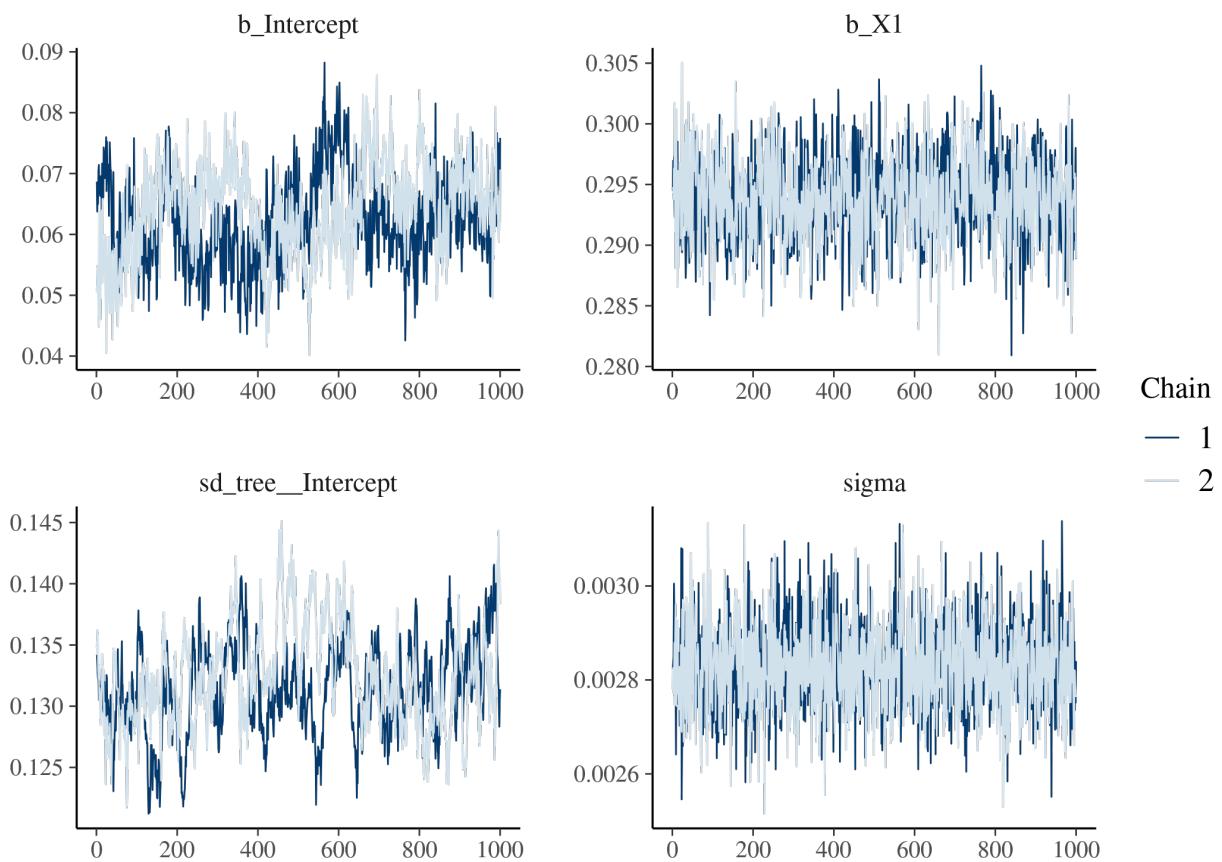


Figure 1.8: Trace of the posterior of the model for Species 1 with genetic IV

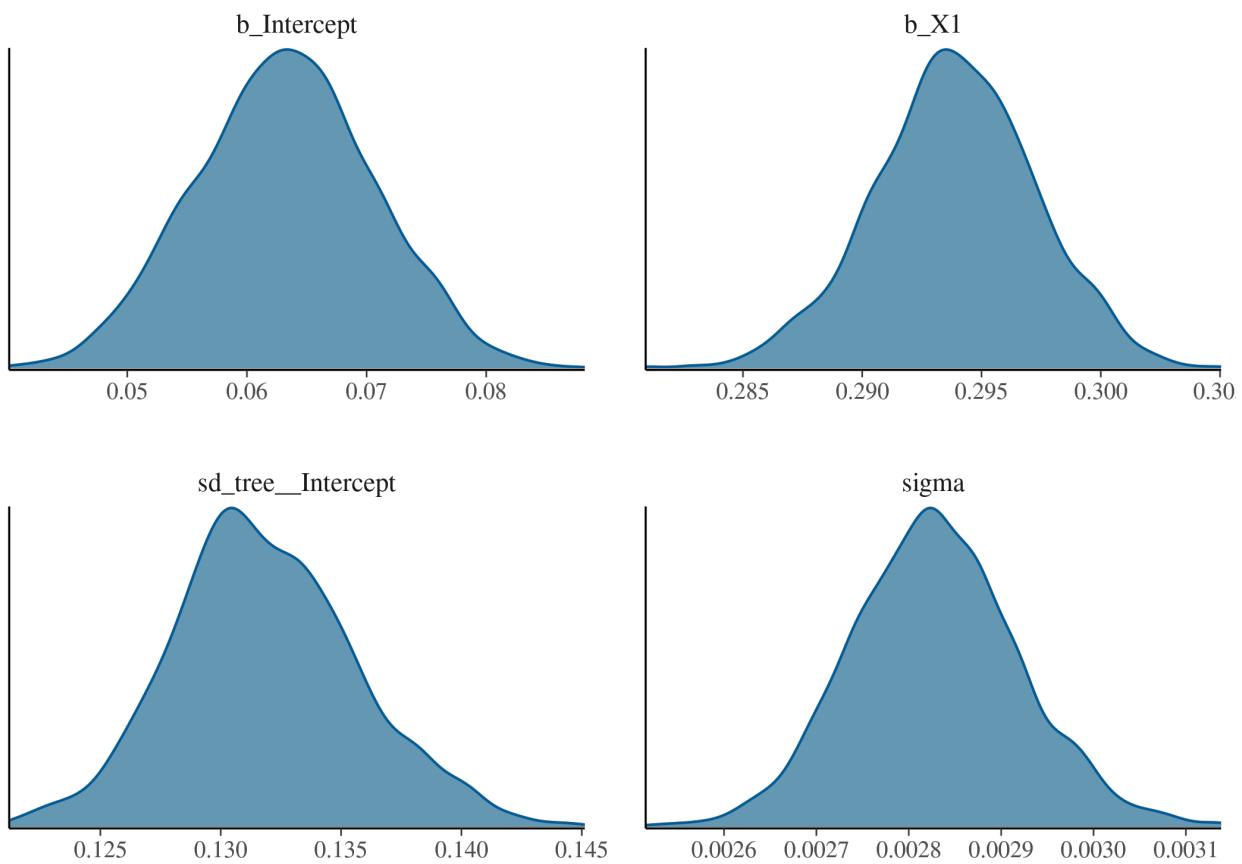


Figure 1.9: Density of the posterior of the model for Species 1 with genetic IV

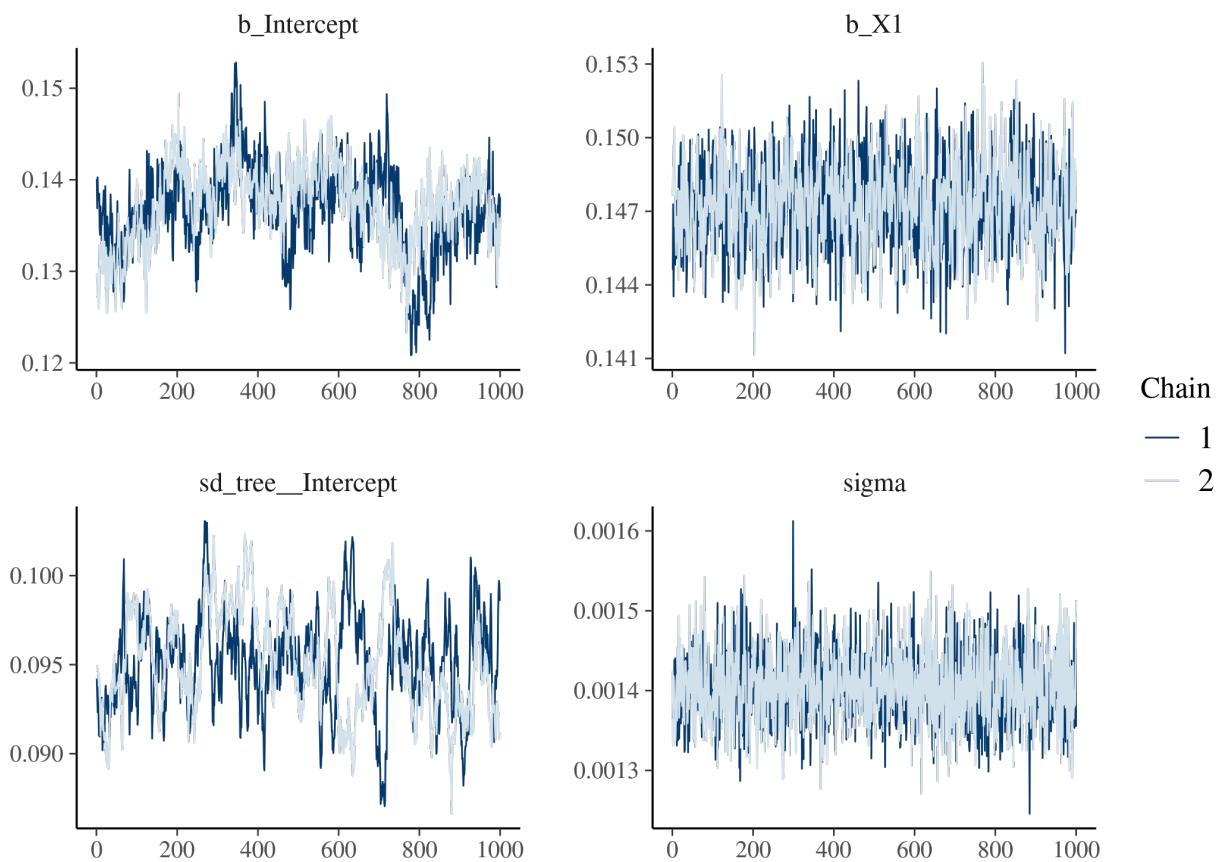


Figure 1.10: Trace of the posteriors of the model for Species 2 with genetic IV

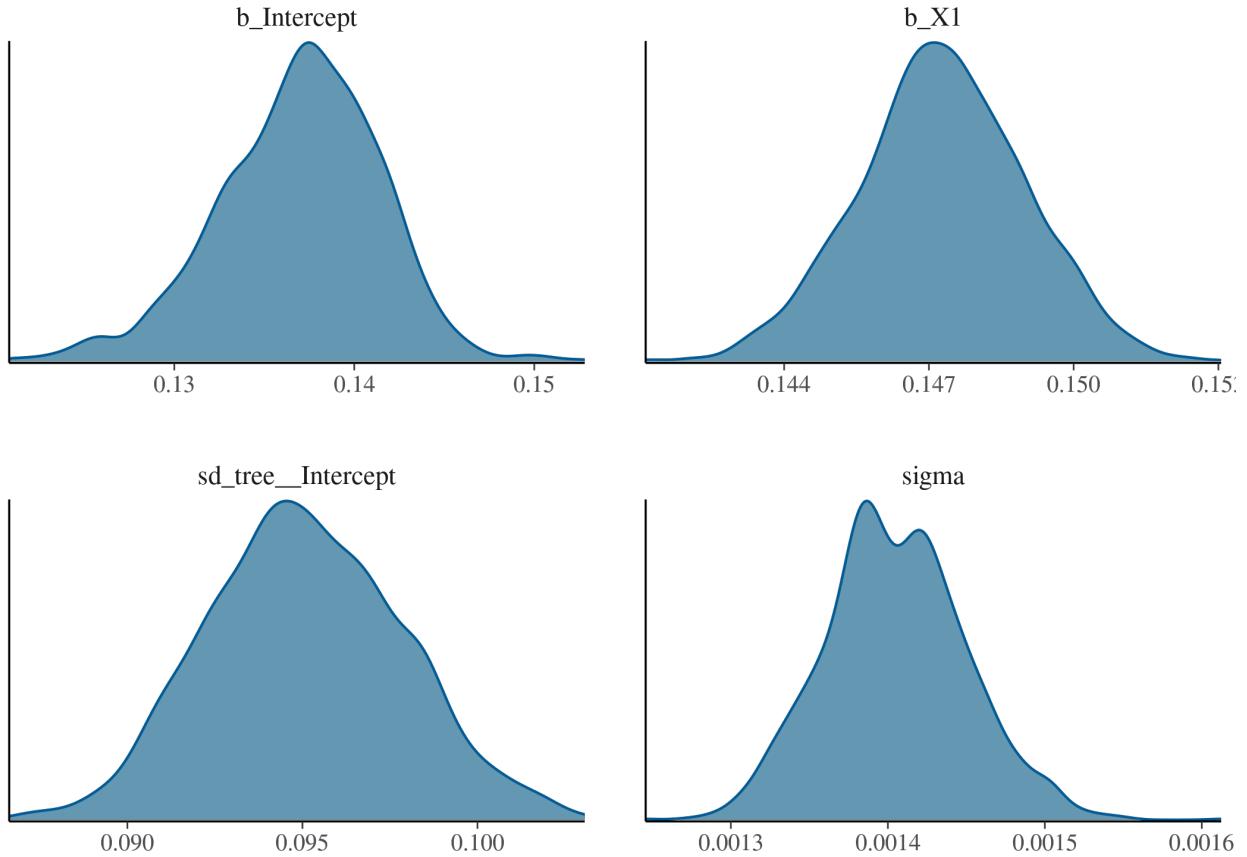


Figure 1.11: Density of the posteriors of the model for Species 2 with genetic IV

```
## Warning: Parts of the model have not converged (some Rhats are > 1.05). Be careful when analysing the
## results! We recommend running more iterations and/or setting stronger priors.
```

Table 1: Mean posteriors and their estimation error

	β_0	β_1	V_b	V
Species 1				
Estimate	6.4e-02	2.9e-01	9.5e-02	1.9e-03
Estimation error	4.7e-03	2.4e-03	3e-03	6.1e-05
Species 2				
Estimate	1.3e-01	1.5e-01	7.6e-02	5.2e-04
Estimation error	3.3e-03	6.3e-04	2.4e-03	1.7e-05

We inferred a high IV even in the absence of genetic IV (Table 1). Therefore, observed IV does not necessarily reveal intrinsic (mainly genetic) IV, but can also reveal hidden dimensions of the environment. We used the model parameters to plot the relationship between Y and $X1$ accounting for intraspecific variability.

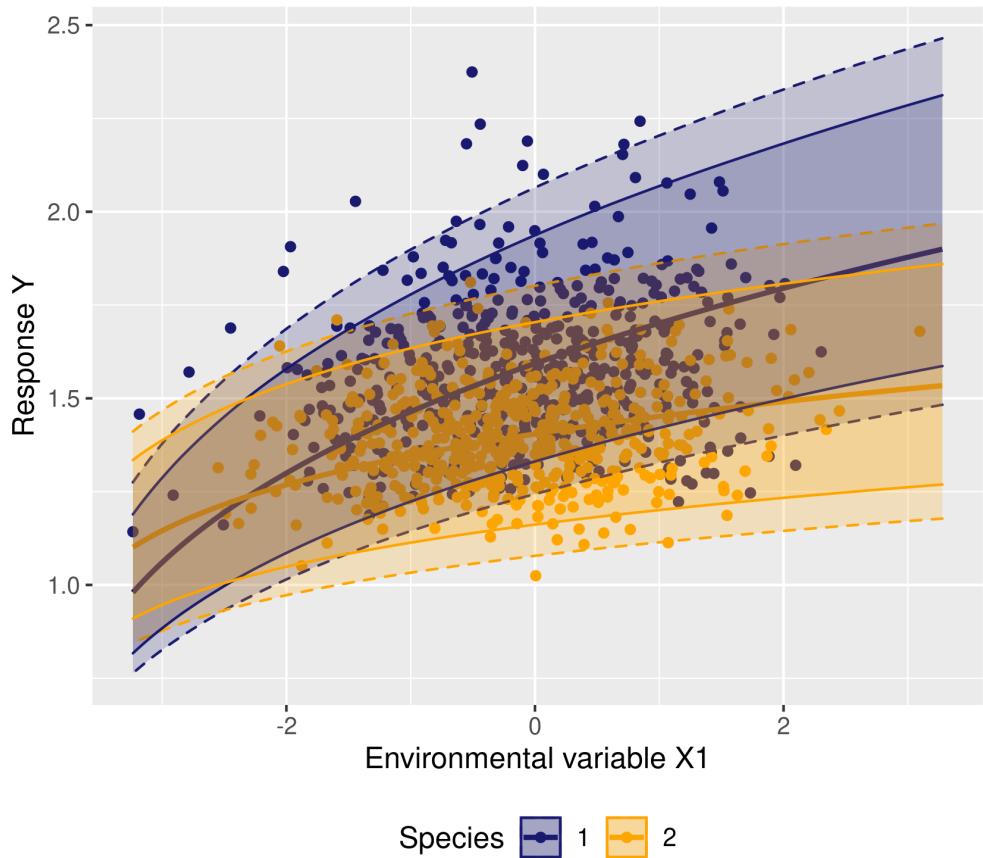


Figure 1.12: Plot of the real values - points - and estimated mean - bold line - and 95 % confidence interval - thin lines - of Y versus X1. The dotted lines correspond to the 95 % interval due to genetic IV.

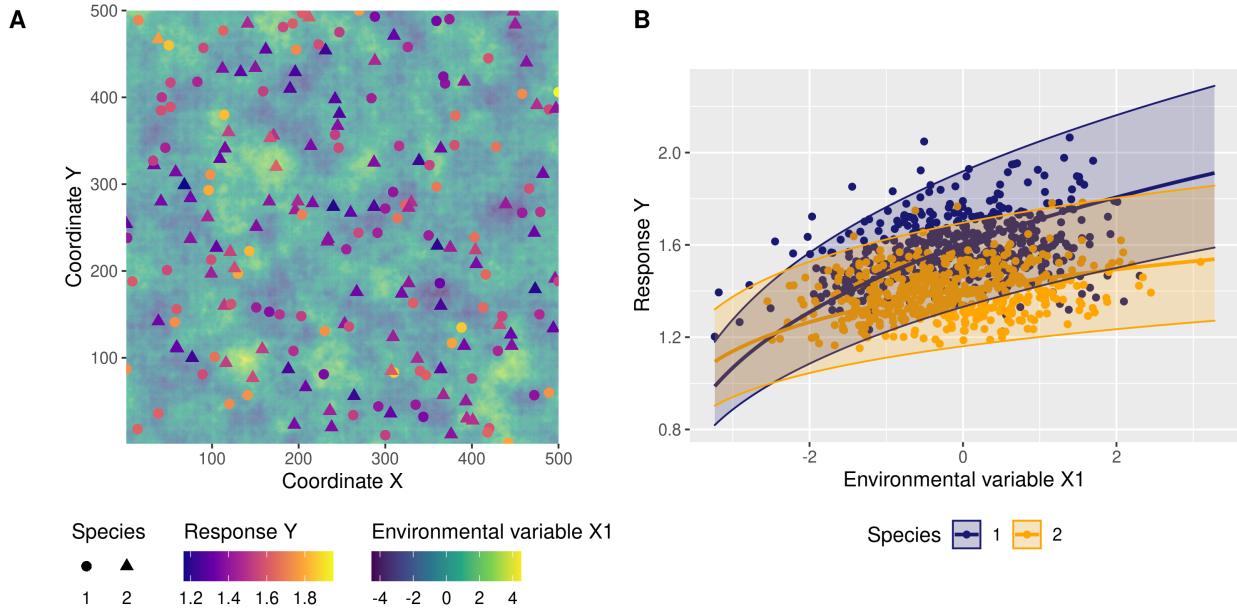


Figure 1.13: Response to an environmental variable and inferred intraspecific variability. (A) Positions of a sample of $I=600$ individuals from $J=2$ species in a landscape defined by a square grid of $C \times C$ cells ($C=500$). The background color indicates the value of the environmental variable X_1 (e.g., light) on each cell at date t . The response (e.g., growth) of each individual, which depends on the environment within each cell (Eq. I), is also indicated by a color scale. (B) Response as a function of the observed environmental variable X_1 for the two species. Points represent the data $\{Y_{ijt}, X_{1ijt}\}$. Thick lines represent the predictive posterior means for the two species. The envelopes delimited by two thin lines represent the 95% credible intervals of the predictive posterior marginalized over individuals (taking into account V_{bj}). The envelopes thus represent the intraspecific variability which is due to the $N-1$ unobserved environmental variables.

In Figure 1.12 and panel A of Figure 1.13, the solid and bold lines represent the mean rate of the response variable (e.g. growth) of Species 1 (blue) and Species 2 (orange) as computed with the parameters retrieved from the model with or without genetic variability respectively. The plain lines represent the 95% interval of the posteriors from the model without genetic variability and the dotted lines show the 95% confidence interval of the posteriors from the model with genetic variability. Figure 1.12 shows that genetic IV increases the overlap between the response of Species 1 and Species 2.

Figure 1.13 represents the virtual landscape of X_1 and the corresponding individual response and the plot of Y versus X_1 (without genetic variability) next to each other, helping to visualise this virtual experiment.

In the model without genetic IV, the unobserved variation in the environment resulted in an important inferred intraspecific variability. The performances of the two species can intersect, which means that the competitive hierarchy among the two species can be locally reversed depending on the micro-environmental variation, offering opportunities for the coexistence of the two species in a variable environment.

1.3 Spatial autocorrelation of individual response

We then analysed the spatial structure of the response variable at the individual scale. We hypothesised that as environmental variables are spatially autocorrelated, and the response variable is the result of these variables, then the response variable should also be spatially autocorrelated. To test this, we computed Moran's I test. It is computed with the Moran.I function of the ape R package (Paradis & Schliep 2019).

Table 2: Results of the Morans's I test for both species separately and together.

	Moran's I index	P-value
Species 1	6.7e-02	0e+00
Species 2	6.9e-02	0e+00
All individuals	4.3e-02	0e+00

Table 2 shows that the individual response variable is largely spatially autocorrelated. This is due to the spatial autocorrelation in the environmental variables. In this simple and controlled experiment, this seems natural. However, in a less controlled environment, detecting spatial autocorrelation in a response variable could be the sign of the spatial structure of the underlying environmental variables. Nonetheless, we acknowledge that the genetic structure of the population associated with limited dispersion for example could also lead to spatial autocorrelation in the response variable of the individuals, but this is likely to happen on a much larger geographical scale.

1.4 Similarity between conspecific individuals compared to heterospecific individuals and consequences for species coexistence.

Finally, we used semivariograms to visualise the spatial autocorrelation of the response variable and to test whether the individual growth was more similar within conspecifics than within heterospecifics. The semivariograms were computed and modelled with the variogram and the fit.variogram functions of the gstat R package (Gräler *et al.* (2016)) respectively. The variogram models were spherical.

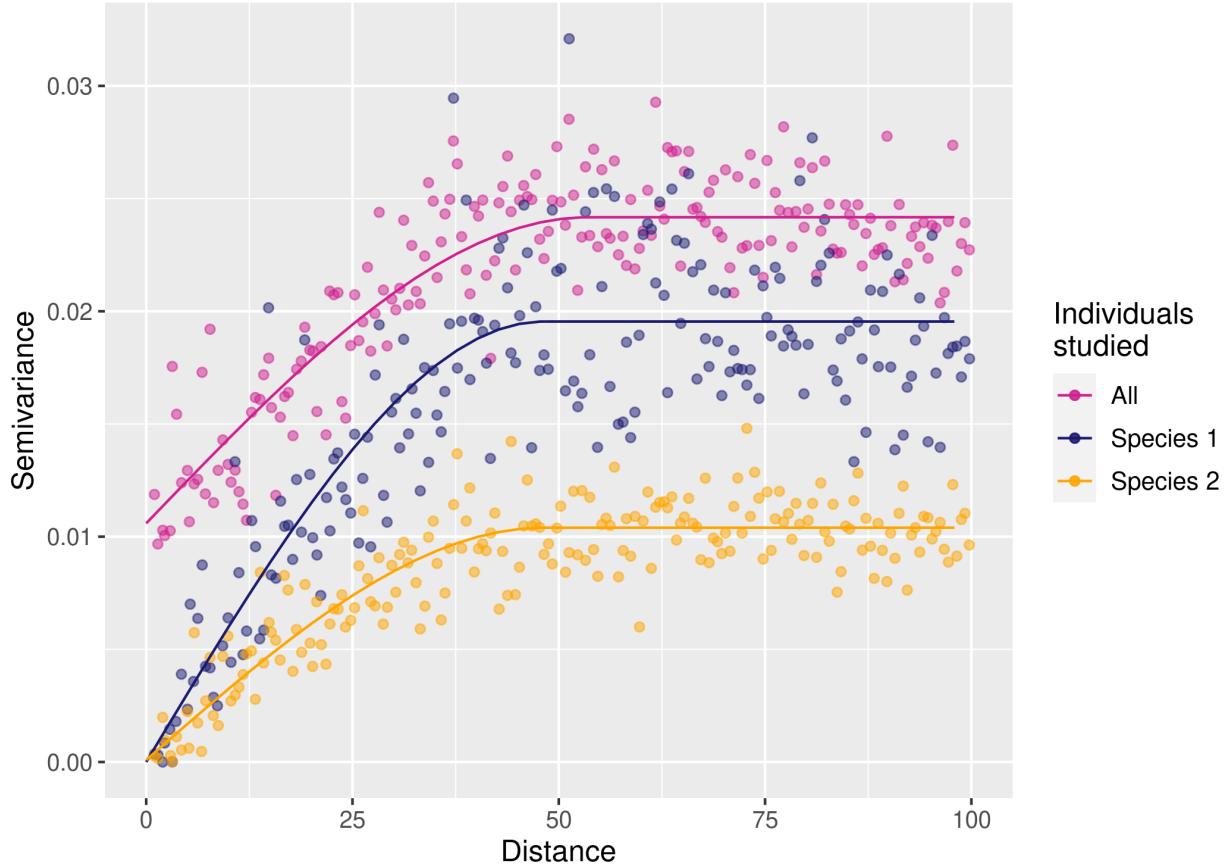


Figure 1.14: Spatial autocorrelation of response Y across individuals within and between species ($J=2$). This semivariogram represents the semivariance of the individual mean responses Y as a function of the distance between individuals. The increasing curves evidence spatial autocorrelation in Y (similar results using Moran's I test). The semivariance of all individuals taken together (purple curve) is higher than the semivariance of conspecifics for the two species (red and blue curves), which means that intraspecific variability is lower than interspecific variability.

As the semivariance between individuals of both species was higher than the semivariance between conspecifics (Figure 1.14), the individual response variable was more similar between conspecifics than heterospecifics. Considering this response variable as a proxy of the species performance, we argue that this would imply that intraspecific competition is greater than interspecific competition), which is the main condition for stable coexistence. Therefore, **stable** coexistence is possible even with high intraspecific variability, especially when this variability is not intrinsic but due to environmental heterogeneity that is structured in space.

2 Code implementation

The whole analysis is done using the R language (R Core Team 2021) in the Rstudio environment (RStudio Team 2021). The tables are made with the kableExtra package (Zhu 2021), the figures with the package ggplot2 (Wickham 2009), and the code uses other packages of the Tidyverse (Wickham *et al.* 2019) (dplyr (Wickham *et al.* 2021), magrittr (Bache & Wickham 2020)) and other R packages (here (Müller 2020), ggpubr (Kassambara 2020), sp (Pebesma & Bivand 2005; Bivand *et al.* 2013), ggnewscale (Campitelli 2021),

gstat Gräler *et al.* (2016)). The pdf and html documents are produced thanks to the R packages rmarkdown (Xie *et al.* 2019, 2020; Allaire *et al.* 2020), knitr (Xie 2014, 2015, 2021) and bookdown (Xie 2017).

References

- Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., *et al.* (2020). Rmarkdown: Dynamic documents for R.
- Bache, S.M. & Wickham, H. (2020). *Magrittr: A forward-pipe operator for r*.
- Bivand, R., Pebesma, E.J. & Gómez-Rubio, V. (2013). *Applied spatial data analysis with R*. Use R! second edition. Springer, New York Heidelberg Dordrecht London.
- Bürkner, P.-C. (2017). Brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80.
- Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10, 395.
- Campitelli, E. (2021). *Ggnewscale: Multiple fill and colour scales in 'ggplot2'*.
- Clark, J.S., Dietze, M., Chakraborty, S., Agarwal, P.K., Ibanez, I., LaDeau, S., *et al.* (2007). Resolving the biodiversity paradox. *Ecology Letters*, 10, 647–659.
- Gräler, B., Pebesma, E. & Heuvelink, G. (2016). Spatio-Temporal Interpolation using gstat. *The R Journal*, 8, 204.
- Kassambara, A. (2020). *Ggpubr: 'ggplot2' based publication ready plots*.
- Müller, K. (2020). *Here: A simpler way to find your files*.
- Paradis, E. & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528.
- Pebesma, E.J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences*, 30, 683–691.
- Pebesma, E.J. & Bivand, R.S. (2005). Classes and methods for spatial data in R. *R News*, 5, 9–13.
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- RStudio Team. (2021). *RStudio: Integrated development environment for r*. RStudio, PBC, Boston, MA.
- Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Use R! Springer, New York.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., *et al.* (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4, 1686.
- Wickham, H., François, R., Henry, L. & Müller, K. (2021). *Dplyr: A grammar of data manipulation*.
- Xie, Y. (2014). Knitr: A Comprehensive Tool for Reproducible Research in R. In: *Implementing reproducible research*, The R series (eds. Stodden, V., Leisch, F. & Peng, R.D.). CRC Press, Taylor & Francis Group, Boca Raton.
- Xie, Y. (2015). *Dynamic documents with R and Knitr*. Second edition. CRC Press/Taylor & Francis, Boca Raton.
- Xie, Y. (2017). *Bookdown: Authoring books and technical publications with R Markdown*. CRC Press, Boca Raton, FL.
- Xie, Y. (2021). *Knitr: A general-purpose package for dynamic report generation in r*.

- Xie, Y., Allaire, J.J. & Grolemund, G. (2019). *R Markdown: The definitive guide*. CRC Press, Taylor; Francis Group, Boca Raton.
- Xie, Y., Dervieux, C. & Riederer, E. (2020). *R markdown cookbook*. The R Series. 1st edn. Taylor; Francis, CRC Press, Boca Raton.
- Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax*.