

Theoretical model

Intraspecific variability is often seen as an unstructured, intrinsic property of individuals, mostly genetic. Here, we rather investigate the effect of spatial variations of the environment on individuals. Our hypothesis is that the individuals can be clones, i.e. have no genetic differences, and still have different responses because the environment in which they strive varies. As the response of an individual is a result of all the environmental conditions it has encountered during its life, the individual level is a tool to describe the environment more acutely. This tool may incorporate some genetic differences, but is not able to discriminate the genetic effect from the effect of local environmental variation, or microsite effect. Moreover, the differences between individuals due to environmental variation does not imply a broader overlap of the species niches. Therefore, this type of intraspecific variability has no effect on species coexistence directly.

We designed and conducted a virtual experiment that aims at illustrating that intraspecific variability, or “individual effects”, can result from unobserved variation in some environmental variables [1], therefore accounting for multidimensional species differences which cannot be observed on a few niche axes.

To do so, we first consider a model that depicts the response of a process, e.g. growth, for individual clones of two different species to all the environmental variables that influence it. Henceforth denoted the “Perfect knowledge model”, it 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 is switched to the natural logarithm to represent a log-log relationship as would be the relationship between tree growth and light.

We fix 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 are chosen randomly between -0.05 and 0.05 and β_7 to β_{10} are chosen randomly between 0 and 0.05.

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

The difference between the species is imposed by those parameters. Here, the species 1 is more performant on average thanks to its higher intercept (β_0) and reaction 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 light for growth for instance.

To account for genetic variability in our generated data, we add an intraspecific variability 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 of such a forest plot, 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 consider that all environmental variables are independent. Each variable is simulated by using the gstat package, enabling to create autocorrelated random fields. A spherical semivariogram model is used for

each of the ten environmental variables, with a mean of 0 for each explanatory variable ($\beta = 0$), a sill of 1 ($\text{psill} = 1$) for each and a range of 50 ($\text{range} = 50$).

We then consider that Y has been measured at two times, t_0 and t_1 , for each individual as it would be under periodic forest plot censuses for instance. To each of the 500 individuals is randomly assigned a pair of coordinates within this spatialised environment. We consider that some of the environmental variables (light, temperature, humidity, nutrient availability for instance) changed between t_0 and t_1 and others did not (slope, altitude for instance). For the first environmental variable which has the stronger impact on Y values (like light for growth for instance) and another randomly drawn environmental variable, we compute value at t_1 as $t_0 + \epsilon$, $\epsilon \sim \mathcal{N}(0, 0.1)$. 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 leads to two repeated measurements of Y for each individual of each species, i.e. 2000 values of $Y_{i,j,t}$, with $i = [1; 2]$; $j = [1, 50]$; $t = [t_0; t_1]$.

These two virtual datasets (with and without genetic variability) is then analysed with a “Partial knowledge model”, which represents the process understanding ecologists can derive using these datasets and from their incomplete characterisation of the environment : only a few variables (here only one, X_1 , e.g. light) 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 :

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

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

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

Second level priors :

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

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

This model includes a random individual effect on the intercept (b_{0i}) allowing to account for any variability among individuals within species regarding this parameter.

We ran this model twice, with the datasets generated with and without intraspecific genetic variability. We then acquire the species parameters and the intraspecific variability generated with genetic intraspecific variability and get only intraspecific variability from the model with clones.

These models were fitted with bayesian inference thanks to Stan, with the brms package in Rstudio, with 100000 iterations, a warming period of 50000 iterations and a thinning of 50, thus we finally obtain 1000 estimates per parameter.

Following are some plots to visualise the data, and a plot helping to visualise the link between the virtual landscape and the growth values (without genetic variability).

On this figure, we can already see that microsite effects, which is the effect of the local multidimensional environment on the observed variable, can result in local reversals of the competitive hierarchy between species, i.e. 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. Doing this, microsite effects foster the coexistence of Species 1 and Species 2.

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

No divergences to plot.

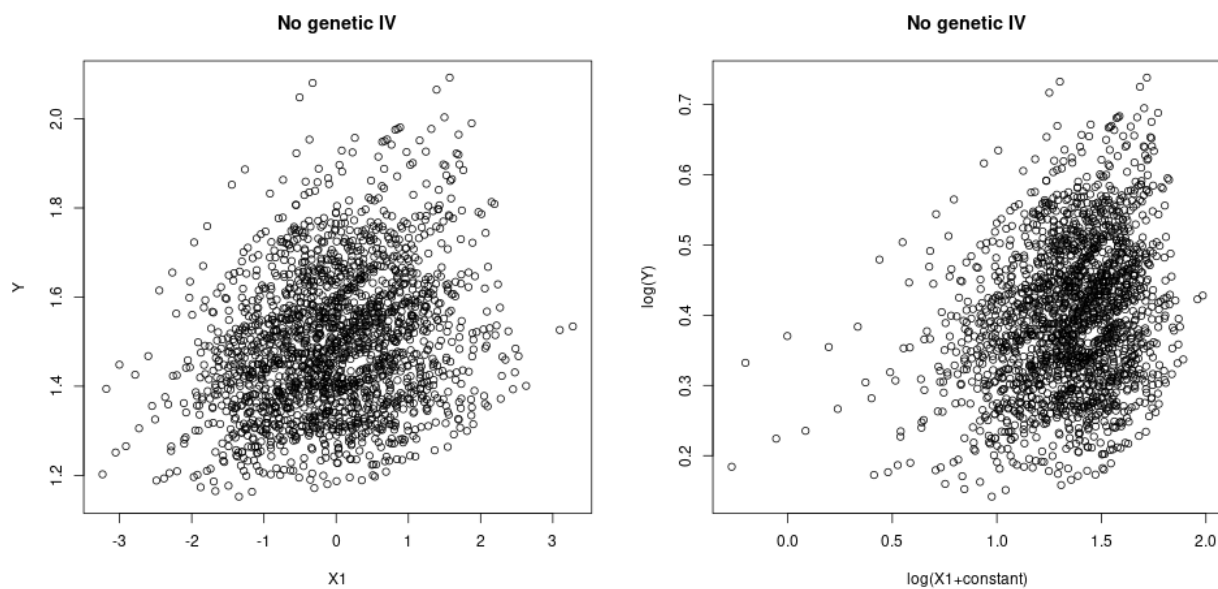


Figure 1: Raw and logged Y versus X_1 plots without genetic IV

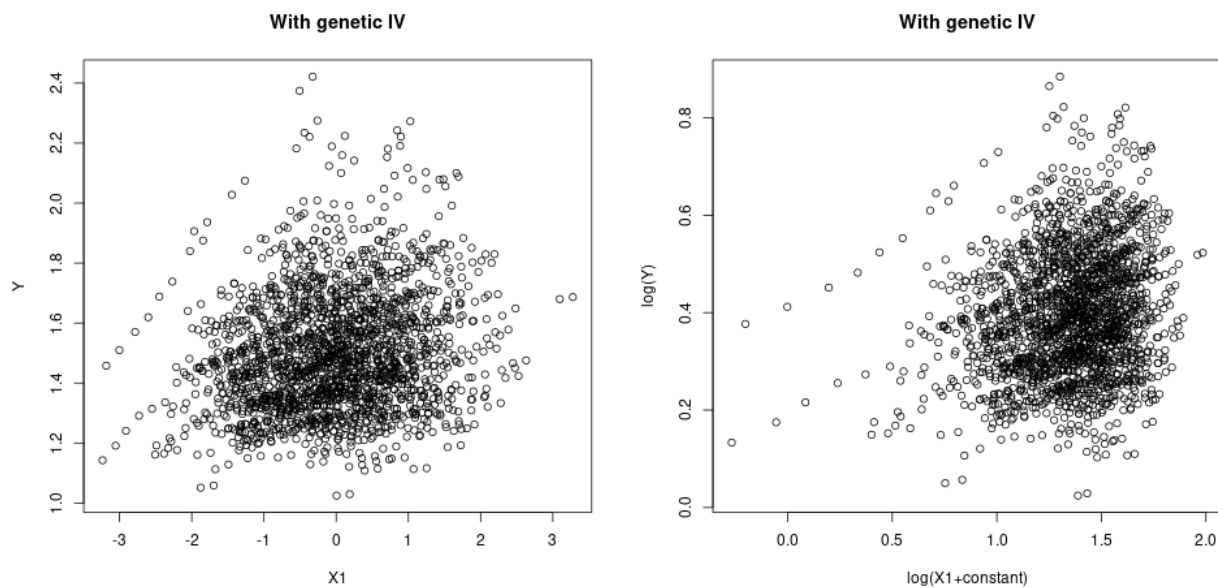


Figure 2: Raw and logged Y versus X_1 plots with genetic IV

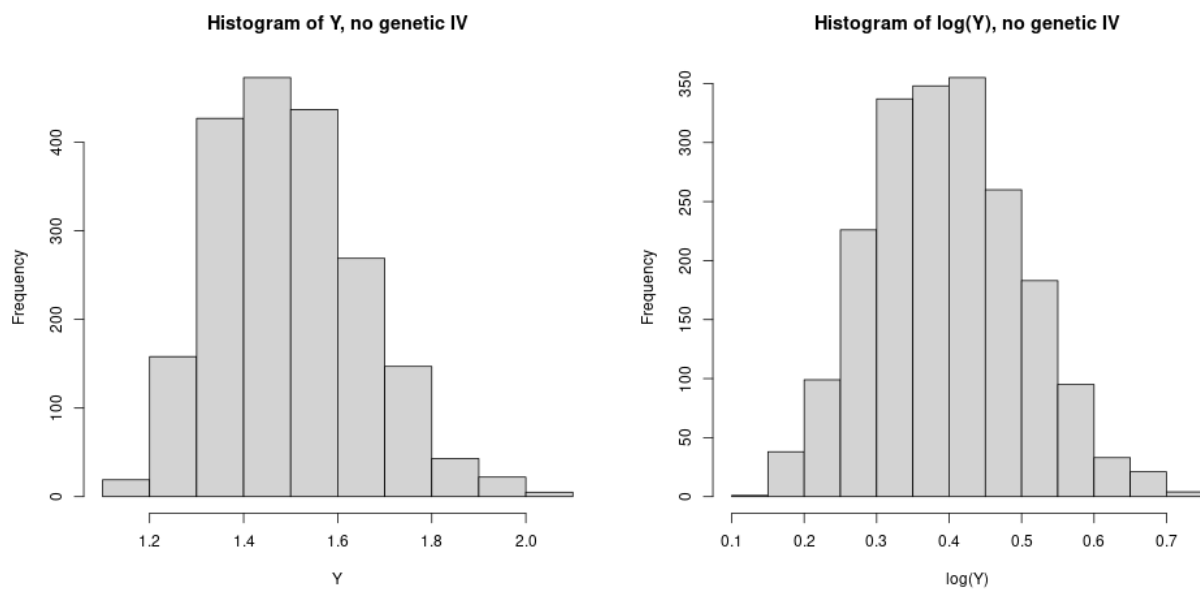


Figure 3: Histogram of Y and log(Y) without genetic IV

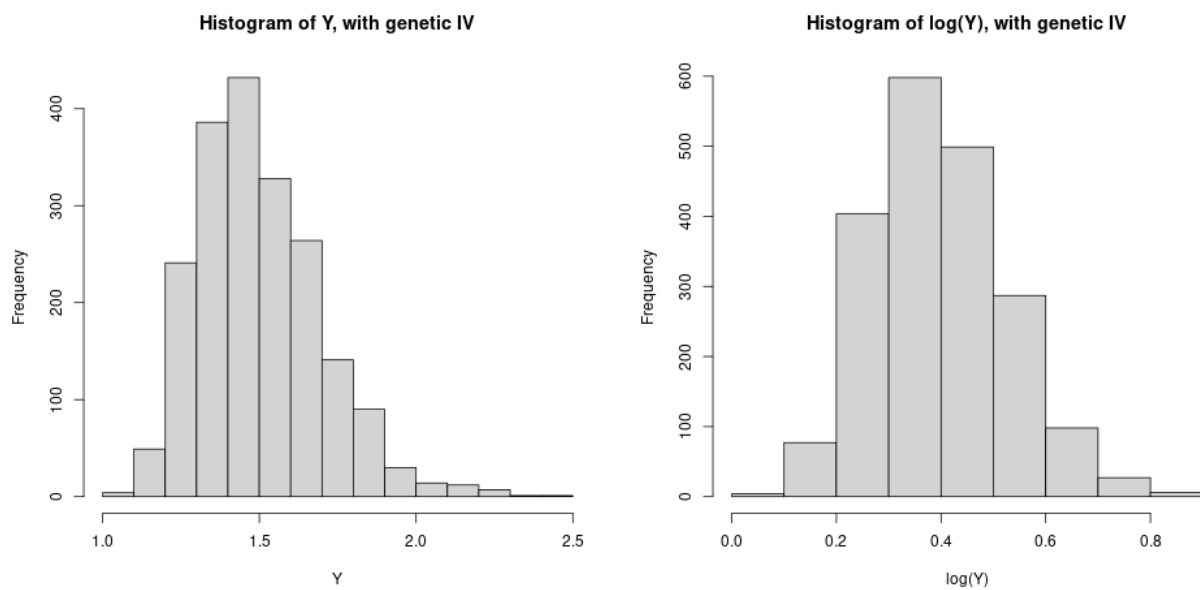


Figure 4: Histogram of Y and log(Y) with genetic IV

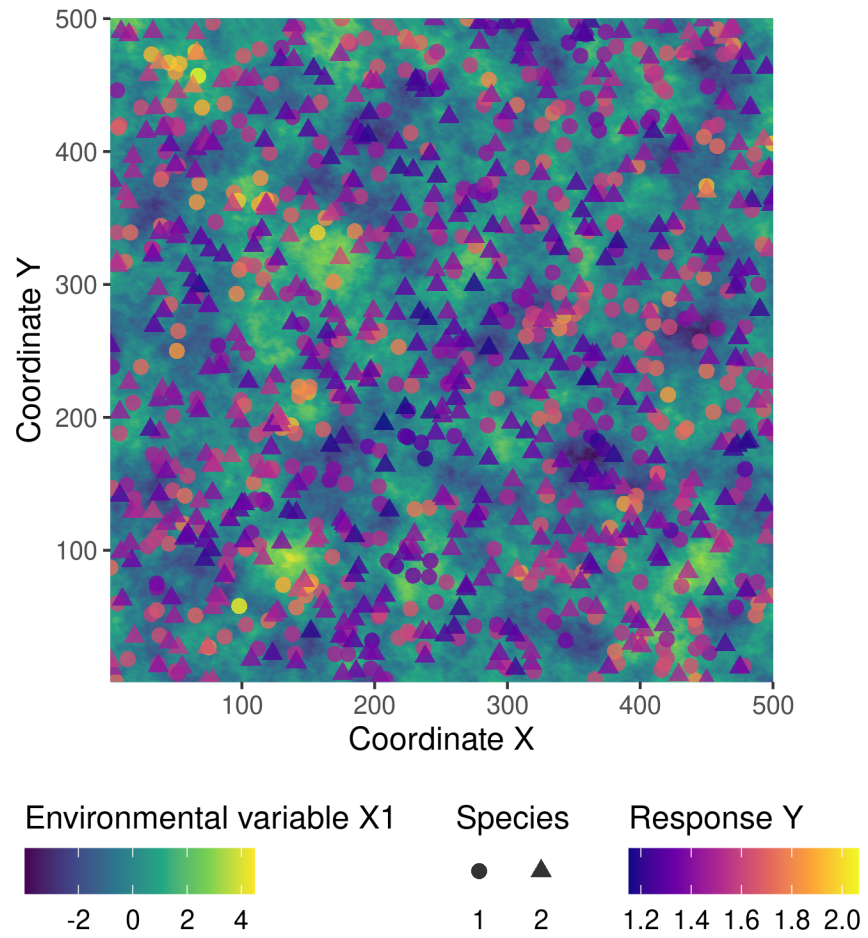
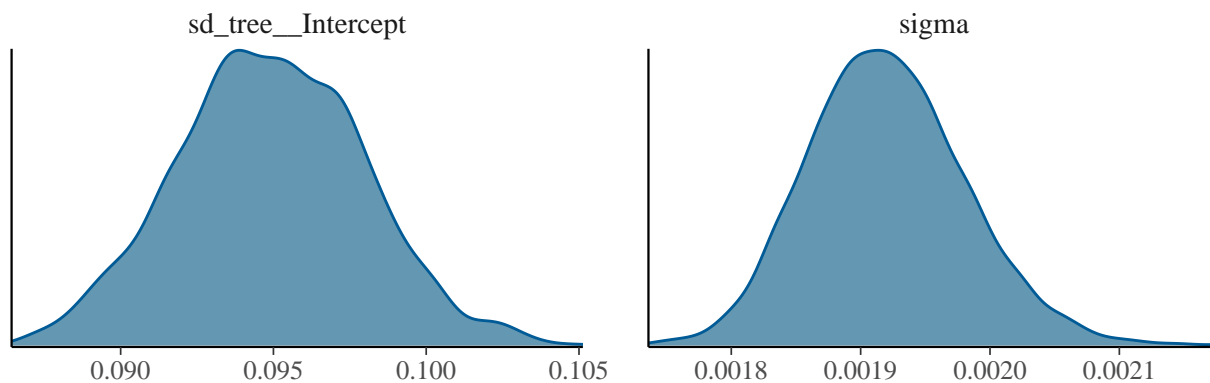
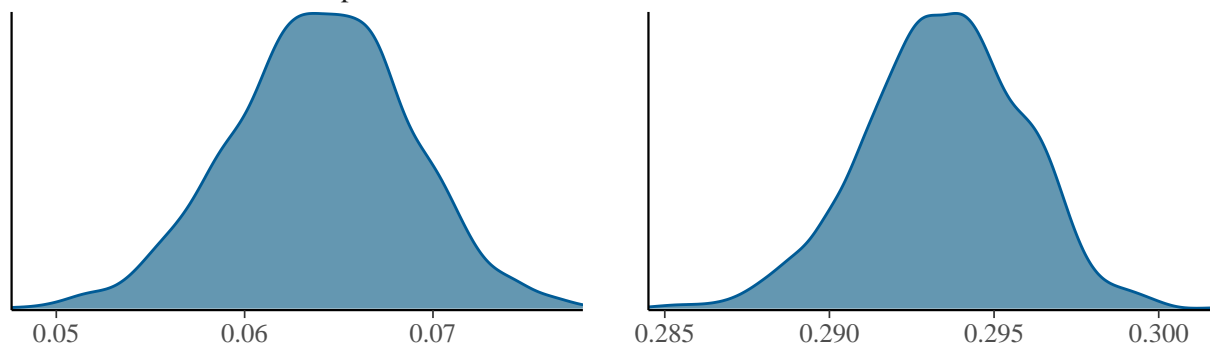
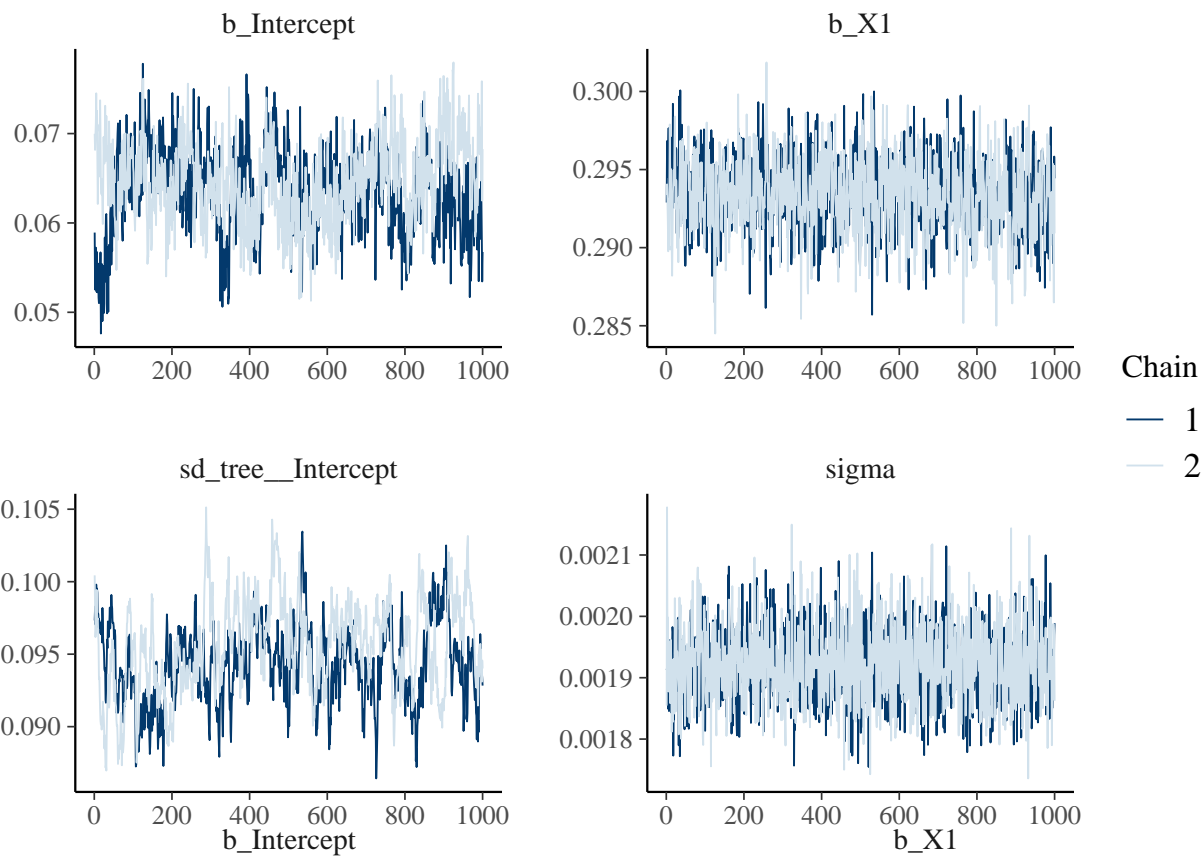


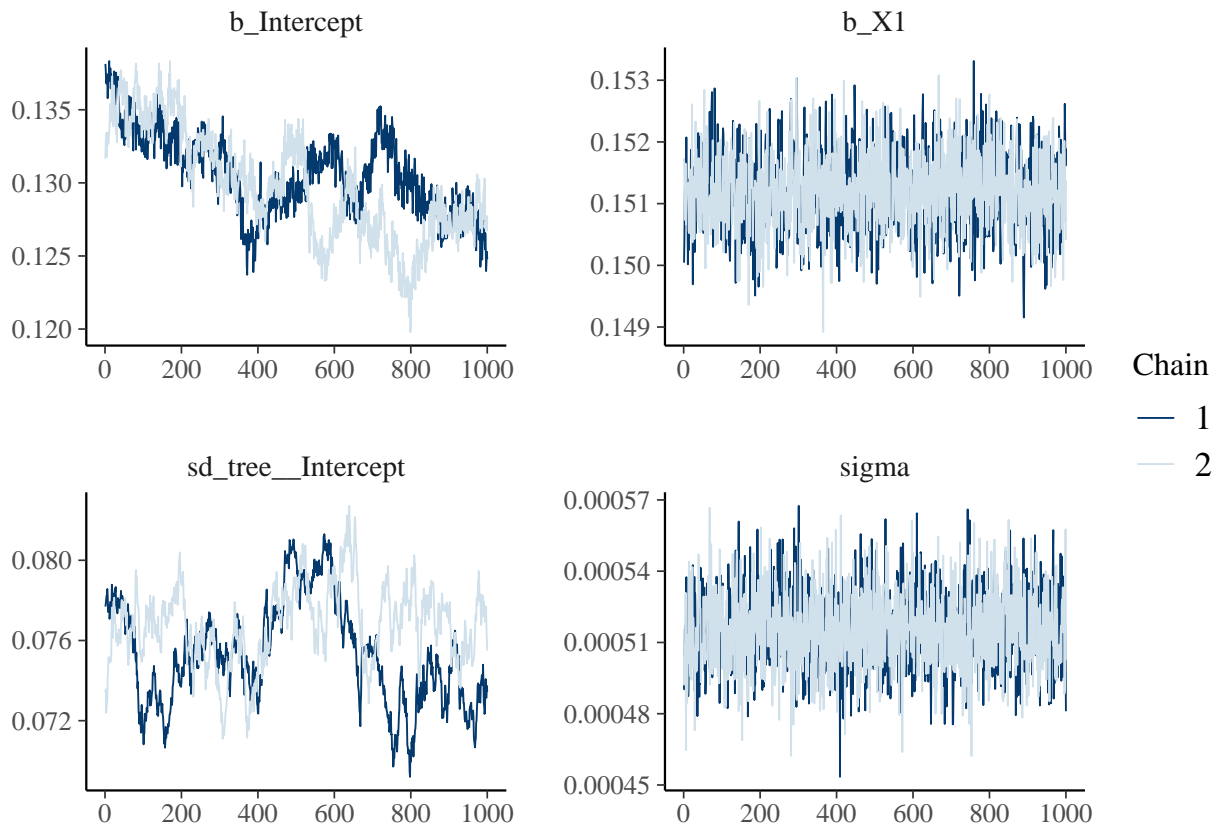
Figure 5: Virtual landscape representing X1 and the individual with their respective Y value.

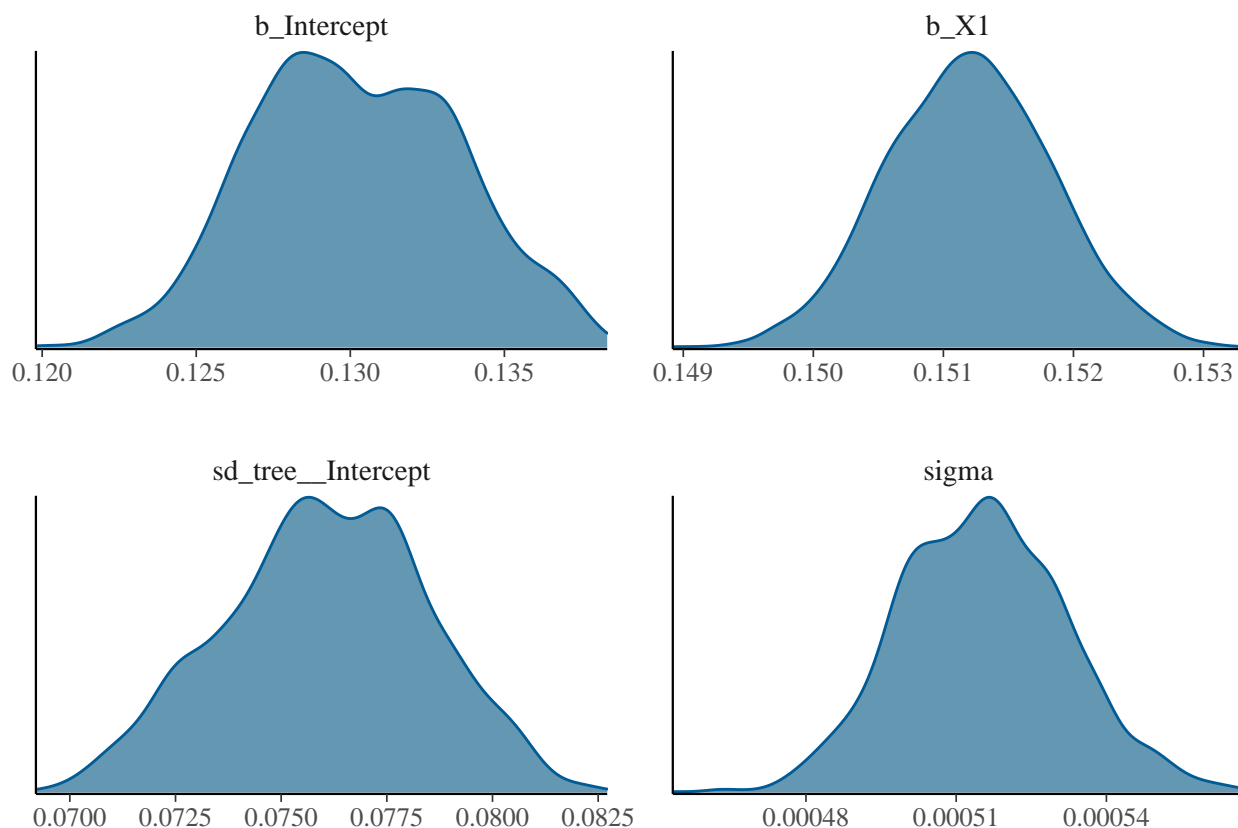


```
## Family: gaussian
## Links: mu = identity; sigma = identity
```

```
## Formula: Y ~ 1 + X1 + (1 | tree)
## Data: Data_brms_sp_1 (Number of observations: 1000)
## Samples: 2 chains, each with iter = 1e+05; warmup = 50000; thin = 50;
## total post-warmup samples = 2000
##
## Group-Level Effects:
## ~tree (Number of levels: 500)
##      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)    0.09     0.00    0.09    0.10 1.04      60     189
##
## Population-Level Effects:
##      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept      0.06     0.00    0.05    0.07 1.02      78     180
## X1             0.29     0.00    0.29    0.30 1.00    1062    1392
##
## Family Specific Parameters:
##      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma      0.00     0.00    0.00    0.00 1.00     1784    1697
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

No divergences to plot.

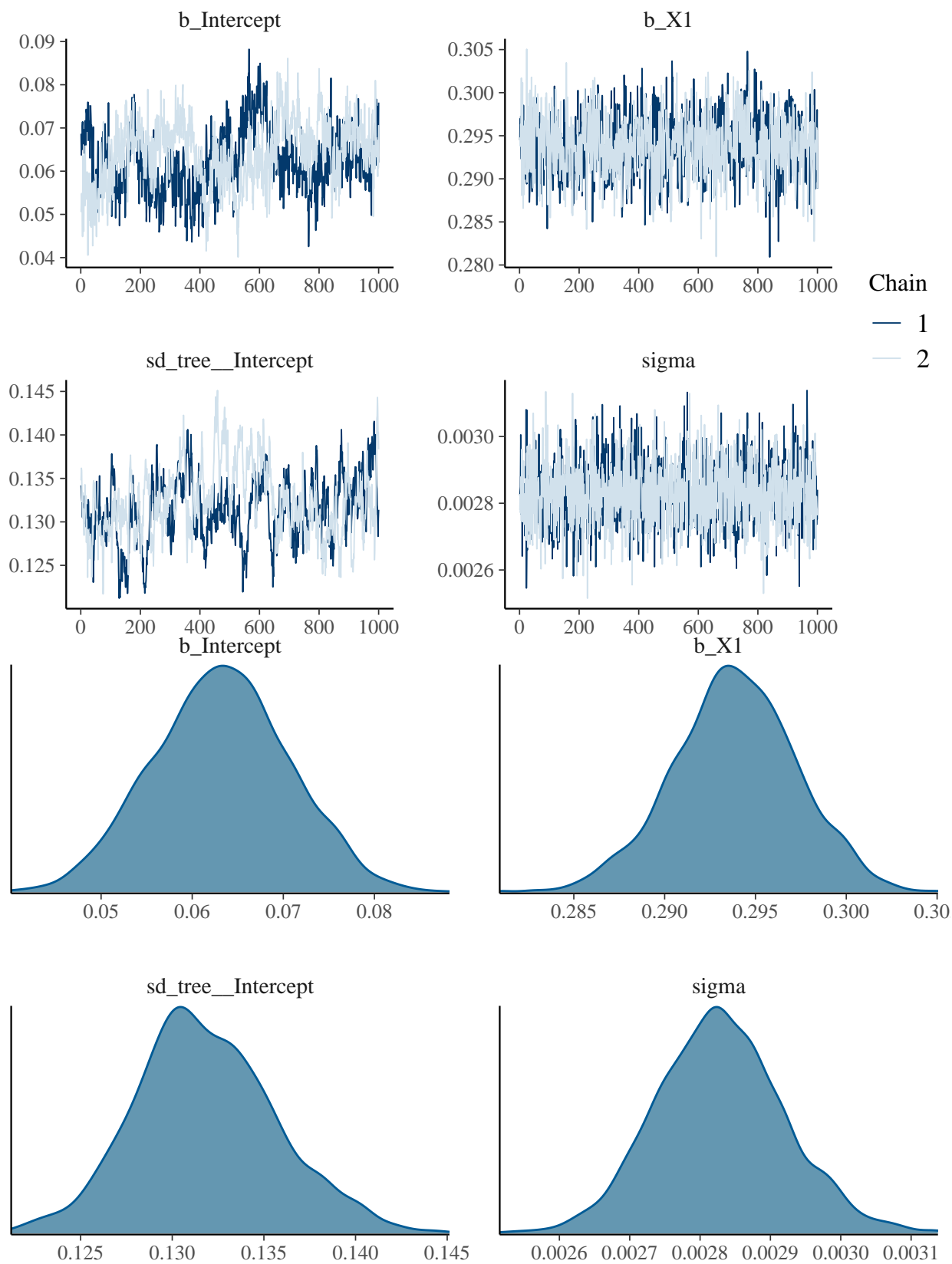




Warning: Parts of the model have not converged (some Rhats are > 1.05). Be careful when
analysing the results! We recommend running more iterations and/or setting stronger priors.

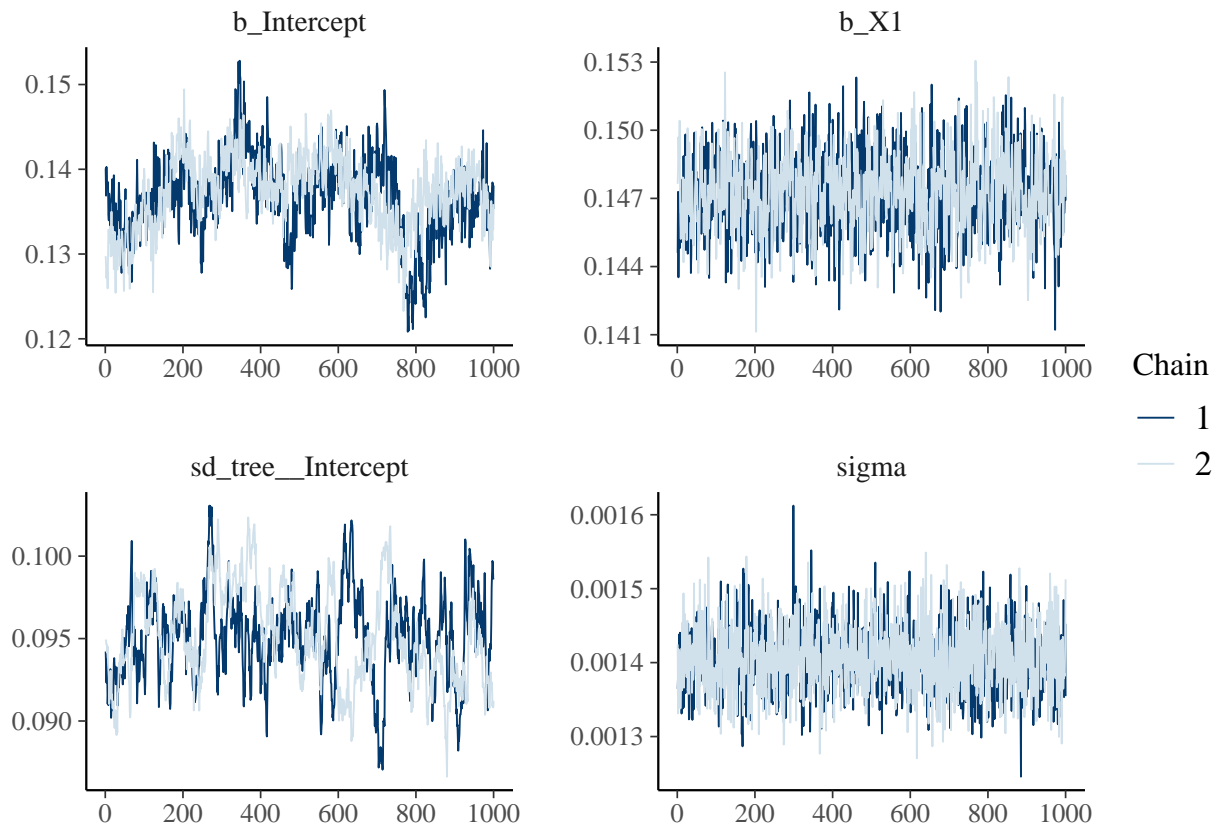
```
## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: Y ~ 1 + X1 + (1 | tree)
## Data: Data_brms_sp_2 (Number of observations: 1000)
## Samples: 2 chains, each with iter = 1e+05; warmup = 50000; thin = 50;
##           total post-warmup samples = 2000
##
## Group-Level Effects:
## ~tree (Number of levels: 500)
##           Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)    0.08      0.00    0.07    0.08 1.10      16      80
##
## Population-Level Effects:
##           Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept      0.13      0.00    0.12    0.14 1.38         4      39
## X1              0.15      0.00    0.15    0.15 1.00     1298     1636
##
## Family Specific Parameters:
##           Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma         0.00      0.00    0.00    0.00 1.00     1809     1964
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

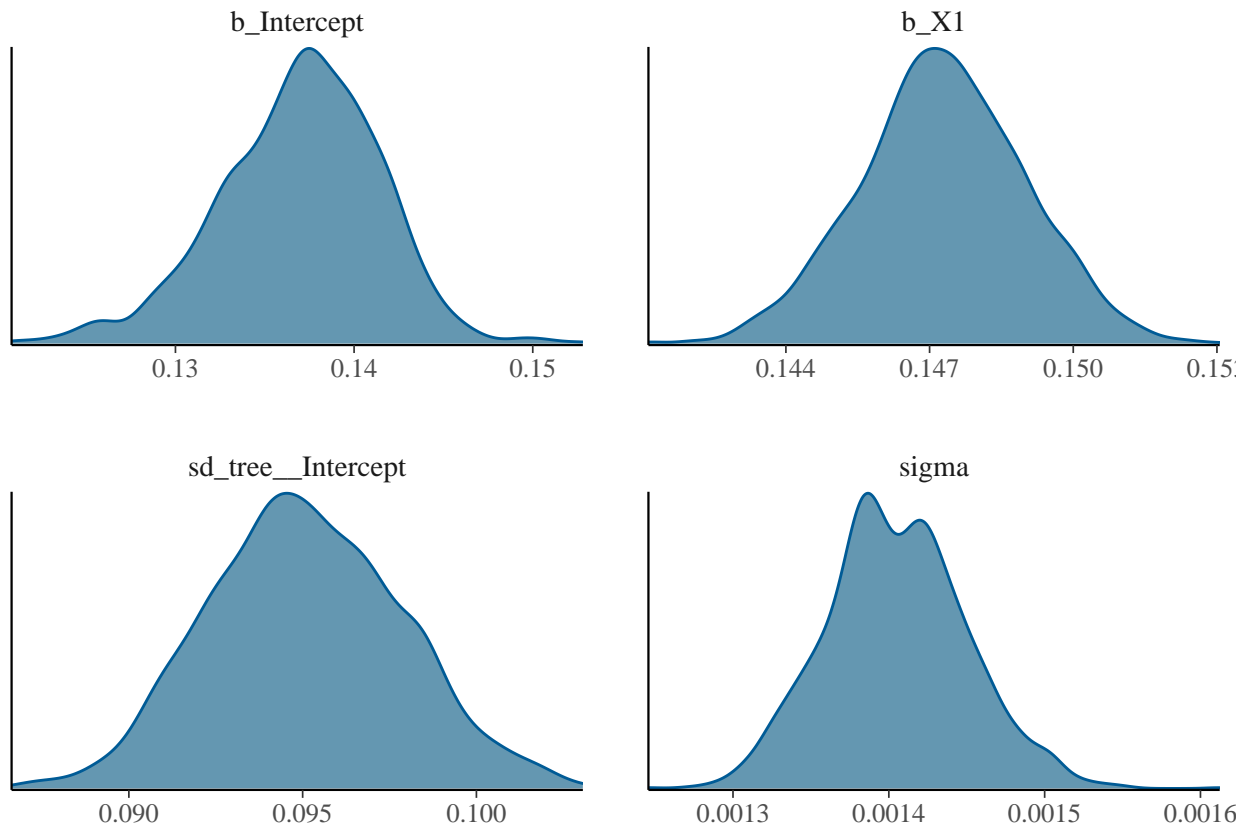

No divergences to plot.



```
## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: Y ~ 1 + X1 + (1 | tree)
## Data: Data_brms_sp_1_IV (Number of observations: 1000)
## Samples: 2 chains, each with iter = 1e+05; warmup = 50000; thin = 50;
##           total post-warmup samples = 2000
##
## Group-Level Effects:
## ~tree (Number of levels: 500)
##           Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)    0.13      0.00    0.12    0.14 1.02      46      136
##
## Population-Level Effects:
##           Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept        0.06      0.01    0.05    0.08 1.05      39      182
## X1                0.29      0.00    0.29    0.30 1.00     1075     1474
##
## Family Specific Parameters:
##           Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma           0.00      0.00    0.00    0.00 1.00     1847     1803
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

No divergences to plot.





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.

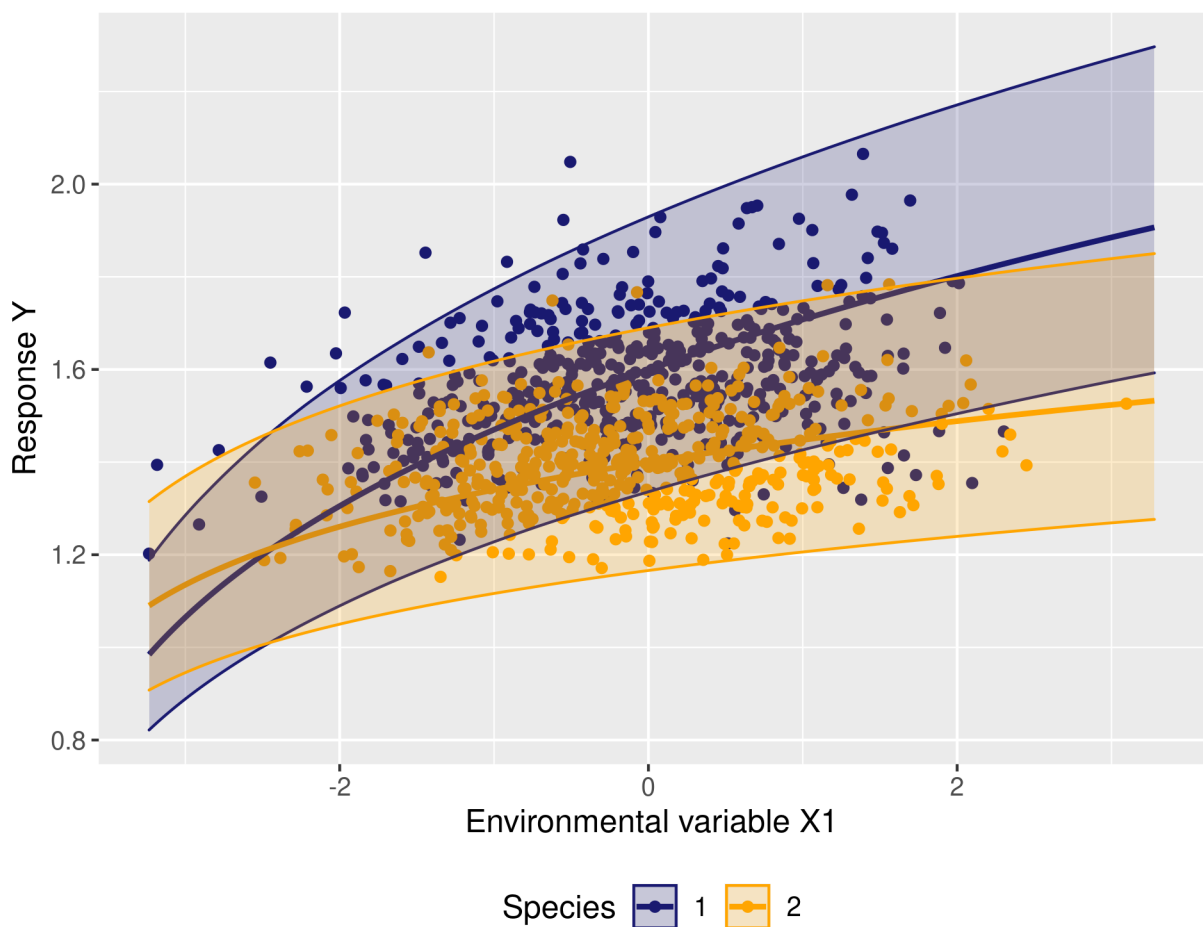
```
## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: Y ~ 1 + X1 + (1 | tree)
## Data: Data_brms_sp_2_IV (Number of observations: 1000)
## Samples: 2 chains, each with iter = 1e+05; warmup = 50000; thin = 50;
##           total post-warmup samples = 2000
##
## Group-Level Effects:
## ~tree (Number of levels: 500)
##           Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)    0.10      0.00    0.09    0.10 1.05      64      148
##
## Population-Level Effects:
##           Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept        0.14      0.00    0.13    0.14 1.01      38       74
## X1                0.15      0.00    0.14    0.15 1.00     1074     1366
##
## Family Specific Parameters:
##           Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma          0.00      0.00    0.00    0.00 1.00     1899     1547
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

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.

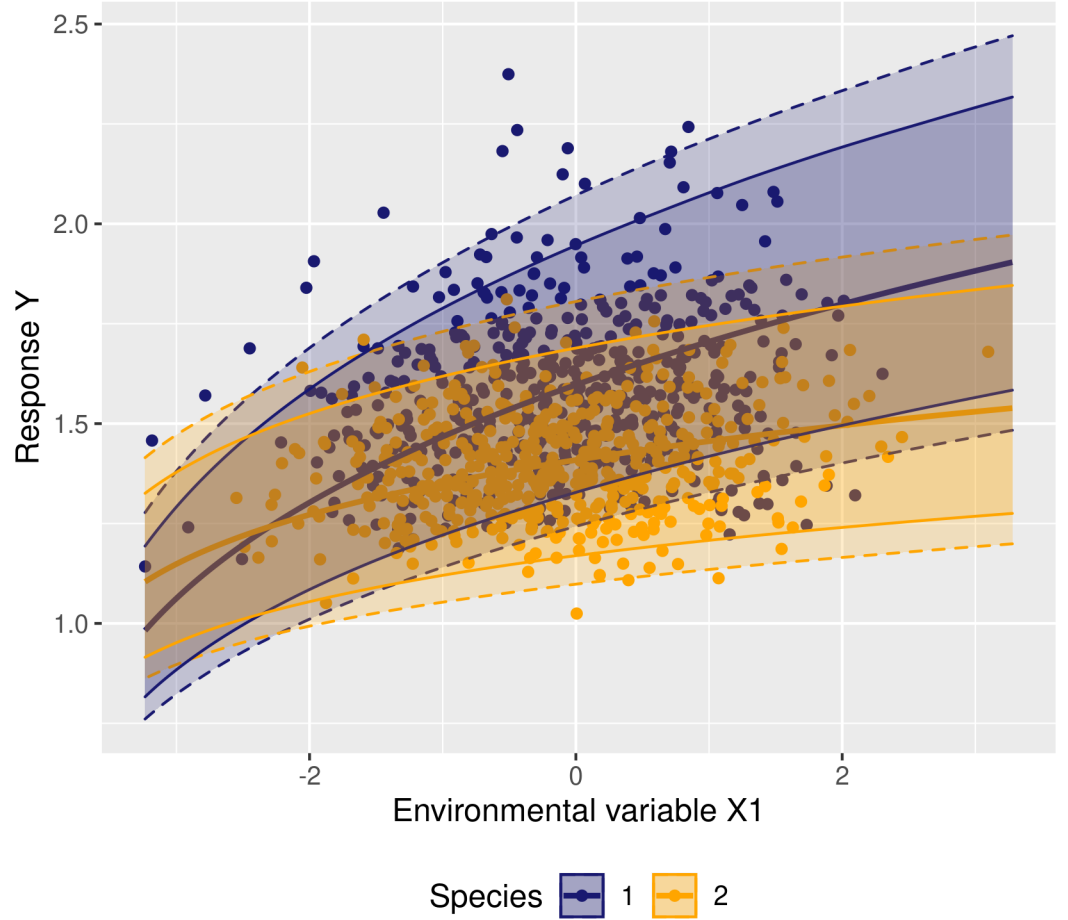
	β_0	β_1	V_b	V
Species 1				
Estimate	6.4e-02	2.9e-01	9.5e-02	1.9e-03
Estimation error	4.7e-03	2.4e-03	3e-03	6.1e-05
Species 2				
Estimate	1.3e-01	1.5e-01	7.6e-02	5.2e-04
Estimation error	3.3e-03	6.3e-04	2.4e-03	1.7e-05

We infer a high intraspecific variability even in the absence of genetic intraspecific variability. Therefore, observed intraspecific variability does not necessarily reveal intrinsic (mainly genetic) intraspecific variability, but can also reveal hidden dimensions of the environment.

The mean and quantiles of the results of the models are used to visualise the inferred link between Y and X_1 . To do so, we create a sequence of explanatory variable values and compute the response variable with the mean and quantiles of the parameters inferred with the models.



\begin{figure}
\caption{Plot of the real values (points) and estimated mean (bold line) and 95% confidence interval (thin lines and shaded area) of Y versus X_1 .} \end{figure}



\begin{figure}
\caption{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.} \end{figure}

In these figures, the solid and bold lines represent the mean growth rate 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.

Putting the virtual landscape of X_1 and the corresponding individual growth and the plot of growth versus light next to each other help to visualise this virtual experiment.

In the model without genetic intraspecific variability, the unobserved variation in the environment results in an important “individual effect”. 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. In that sense, incorporating the individual level in statistical models can help to explain the coexistence of species through a better integration of niche multidimensionality in models.

We then analyse the spatial structure of individual growth.

To do so, we compute Moran’s I test, and we plot the semivariograms of all individuals and of individuals of each species separately. Morans’s I is computed with the Moran.I function of the ape R package.

These results show that individual growth is largely spatially autocorrelated. This is due to the spatial autocorrelation in the environmental variables. In this simple, completely controlled experiment, this is

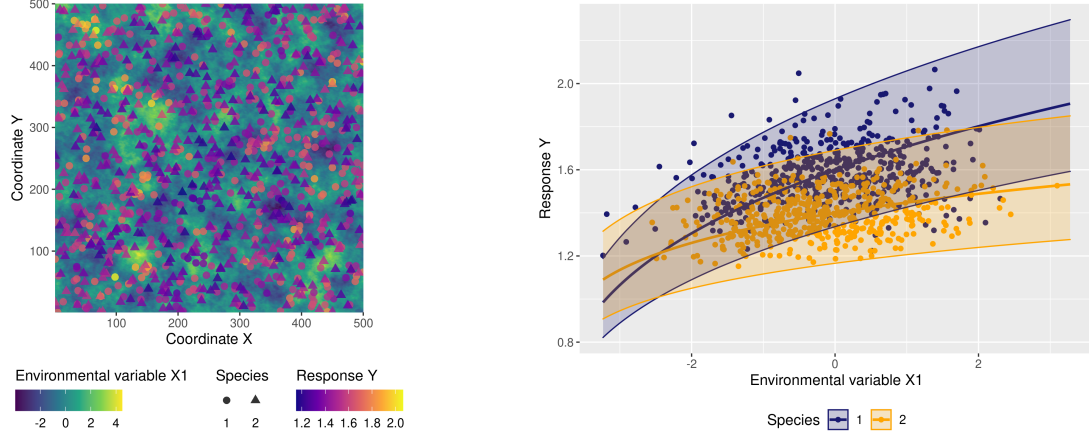


Figure 6: The virtual landscape of X1 values and of Y values and the corresponding estimated relationship between X1 and Y.

Table 1: Results of the Morans’s I test for both species separatly 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

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.

However, other factors like the genetic structure could lead to spatial autocorrelation in growth.

Finally, we can visualise this spatial autocorrelation with semivariograms, but also control if the individual growth is more similar within conspecifics than heterospecifics. The semivariograms are computed and modelled with the variogram and the fit.variogram functions of the gstat R package respectively. The variogram models are spherical.

As the semivariance between individuals of both species is higher than the semivariance between conspecifics, individual growth is more similar between conspecifics than heterospecifics. Considering growth as a proxy of fitness, and linking fitness to competition, we argue that in a Lotka-Volterra model, this would translate into $\alpha_{i,i} > \alpha_{i,j}$, 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.

References

- 1 Clark, J.S. *et al.* (2007) Resolving the biodiversity paradox. *Ecology Letters* 10, 647–659

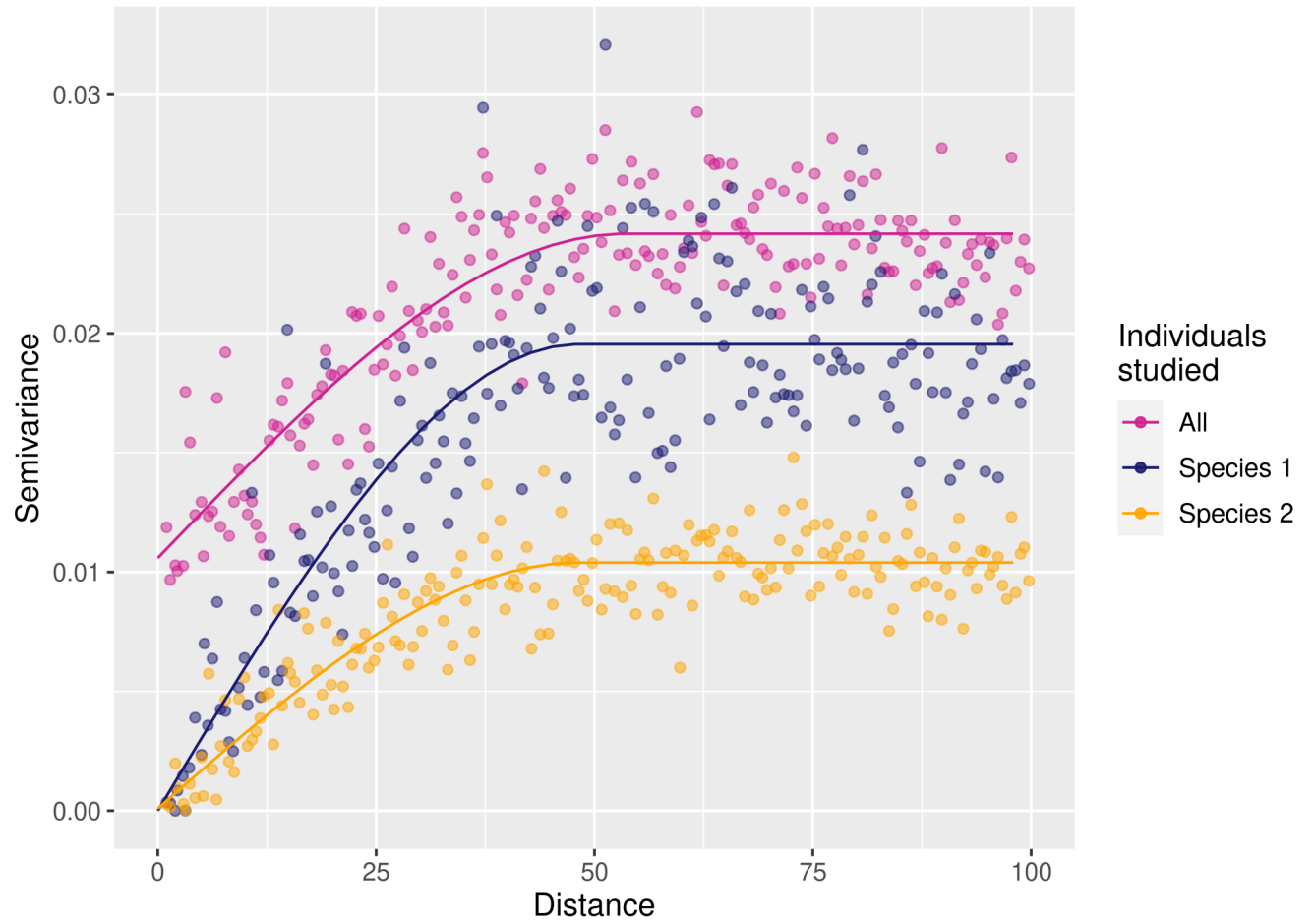


Figure 7: Estimated and empirical semivariograms of Y for each species and both species together.