Analysis of an Eucalyptus clonal plantation dataset

Contents

In this analysis, following the conclusions of the theoretical experiment, we aim at testing if intraspecific variability (IV) in observed individual growth can emerge from the environment. Therefore, we use a clonal experimental setup. The EUCFLUX dataset is a clonal trial on Eucalyptus varieties. One of its main goals is to determine the productivity of each clone. One hypothesis on IV in growth is that it mainly derives from environmental factors, and not only from genetic, intrinsic factors. Therefore, we can use the Eucflux dataset in order to ask if we can we detect some IV of growth within clones. As we suppose growth to be mainly the result of environmental variables, we expect to detect IV of growth within clones (or single genotypes), although we also expect the genotype to influence the growth response.

1 The dataset

The EUCFLUX setup is located in Brasilia, in the state of Sao Paulo. It has 16 genotypes (14 clones and two seed-origin genotypes) of 5 different species or hybrids, which grow on 10 sites and were monitored over 6 years. The detail of the setup is detailed in [1]. We used the DBH measured during 5 complete censuses in order to compute annual growth in mm/year : we kept only the dates of measurements for which the code is "1" for complete DBH measure [1].

The raw data of the experiment was manually rearranged into six files with LibreOffice Calc, each file corresponding to a complete census. The seed-origin individuals were removed from the data since they contain genetic individual variability.

We compute annualised growth in mm/y as the difference between two consecutive censuses divided by the time between the two censuses and we remove all negative growth values. We compute the neperian logarithm of diameter and of growth (with a constant for growth in order to avoid undefined values).

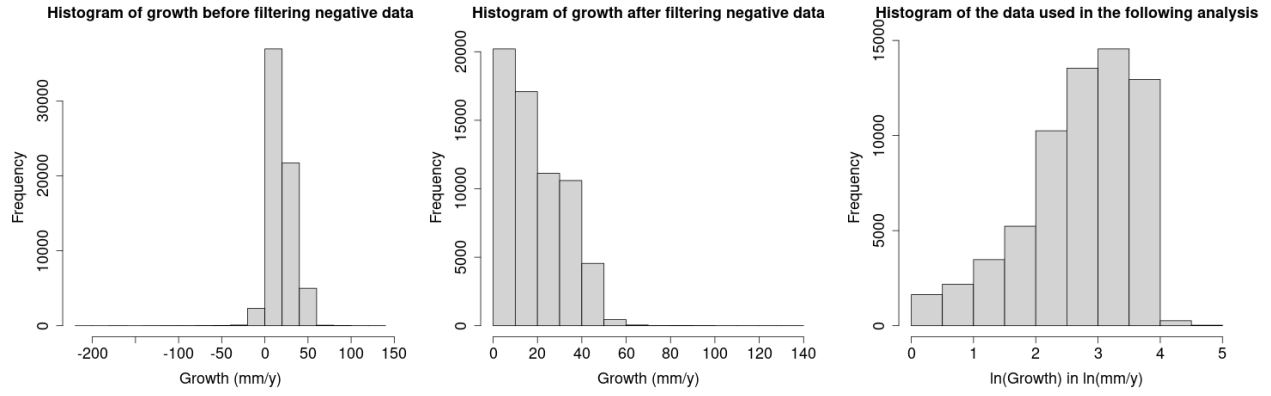The studied dataset finally encompasses 13531 trees.

Figure 1.1: The original data, the data without negative growth values and the loged growth.

Figure 1.1 shows the distribution of growth before and after removing negative values, and the latter with logged values.
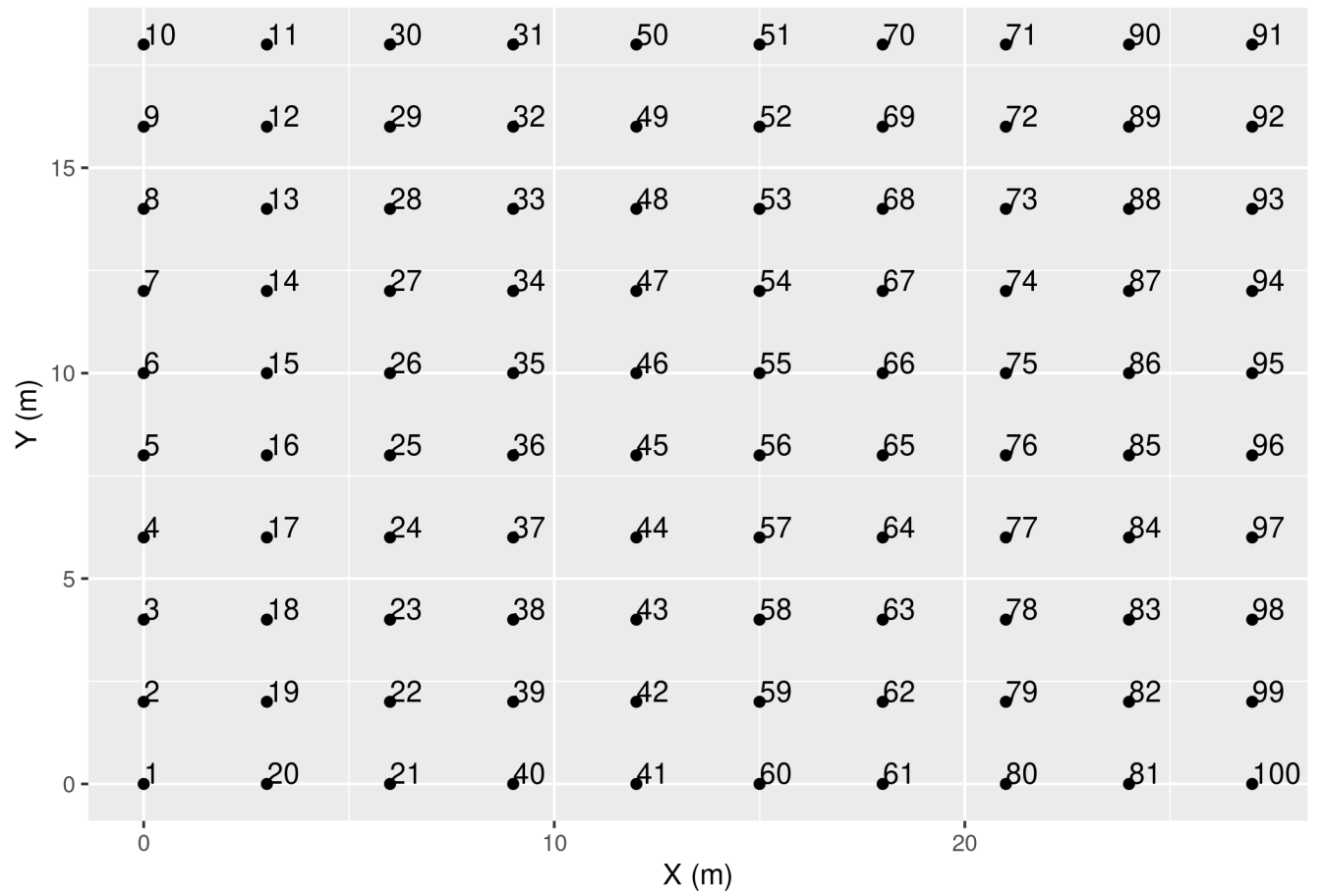

Figure 1.2: Design of a plot. Each point is a tree and the associated number is the tag of the tree.

Figure 1.2 shows the disposition of the trees in a single plot.
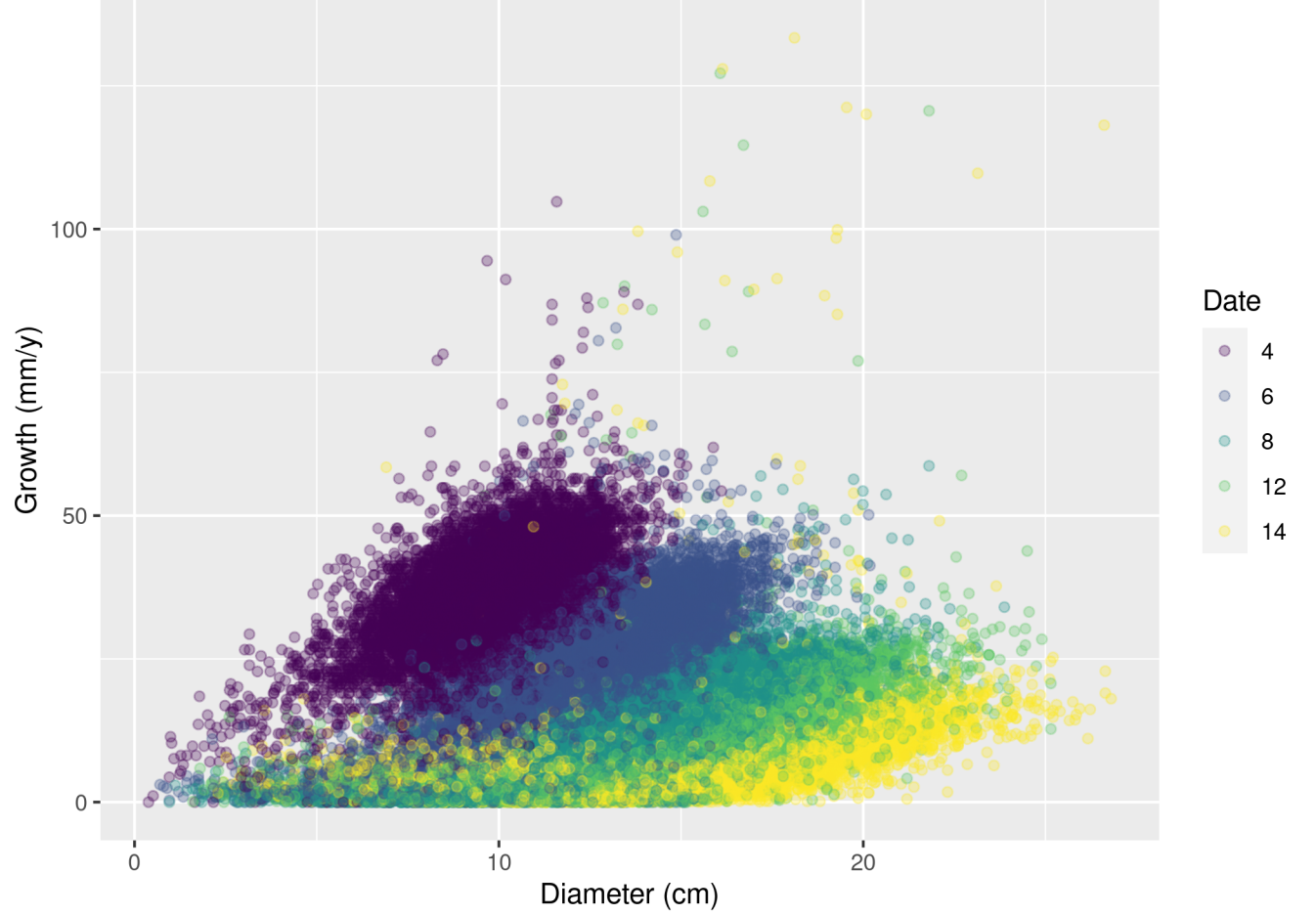
2   Competition index

Figure 2.1: Plot of the growth versus the diameter. Each colour represents a different census.

Figure 2.1 shows that the date of measure has a big influence on the values of growth but also on the relationship between growth and diameter : the slope seems to be smaller with time, indicating that for the same diameter, growth is slower through time. This is likely an effect of competition for light and possibly underground resources, since as the trees grow their capacity to capture resources increases.

Therefore, we compute a competition index to integrate this effect in the growth model. The competition index is computed for each tree which is not on the edge of a plot. It is the sum of the basal areas of the direct neighbours divided by the area of the rectangle that comprises all the direct neighbours. It is then logged.

$C_{i,t} = \frac{\sum BA_{neighbours(i,t)}}{A}$, A is constant.

3   Statistical growth model

In order to partition the variance of individual growth data, we built a hierarchical Bayesian model and used Stan and the package brms [2,3] in Rstudio to implement it in R. We used the following parameters : n.adapt = 1000 ; n.burn = 1000 ; n.iter = 5000 ; n.thin = 5.

Our model incorporated a fixed effect on the intercept (β_0), on the slope of diameter D (β_1),and on the competition index C (β_2) and several random effects, namely temporal (date of census, b_t), individual (tree identifier, b_i), spatial (site, b_s), and genotype (b_g).

Variables are scaled to help the convergence of the model.

$$ln(G_{it} + 1) = (\beta_0 + b_i + b_s + b_g + b_t) + \beta_1 \times ln(D_{it}) + \beta_2 \times ln(C_{it}) + \epsilon_{it}$$

Priors

$$\beta_0 \sim \mathcal{N}(mean = 0, var = 1), iid$$

$$\beta_1 \sim \mathcal{N}(mean = 0, var = 1), iid$$

$$\beta_2 \sim \mathcal{N}(mean = 0, var = 1), iid$$

$$b_i \sim \mathcal{N}(mean = 0, var = V_i), iid$$

$$b_s \sim \mathcal{N}(mean = 0, var = V_s), iid$$

$$b_g \sim \mathcal{N}(mean = 0, var = V_g), iid$$

$$b_t \sim \mathcal{N}(mean = 0, var = V_t), iid$$

$$\epsilon_{it} \sim \mathcal{N}(mean = 0, var = V), iid$$

Hyperpriors

$$V_i \sim \mathcal{IG}(shape = 10^{-3}, rate = 10^{-3}), iid$$

4   Results of the model and variance partitioning

After convergence of the model, we examined the variance of each random effect, and this enabled us to perform a variance partitioning.
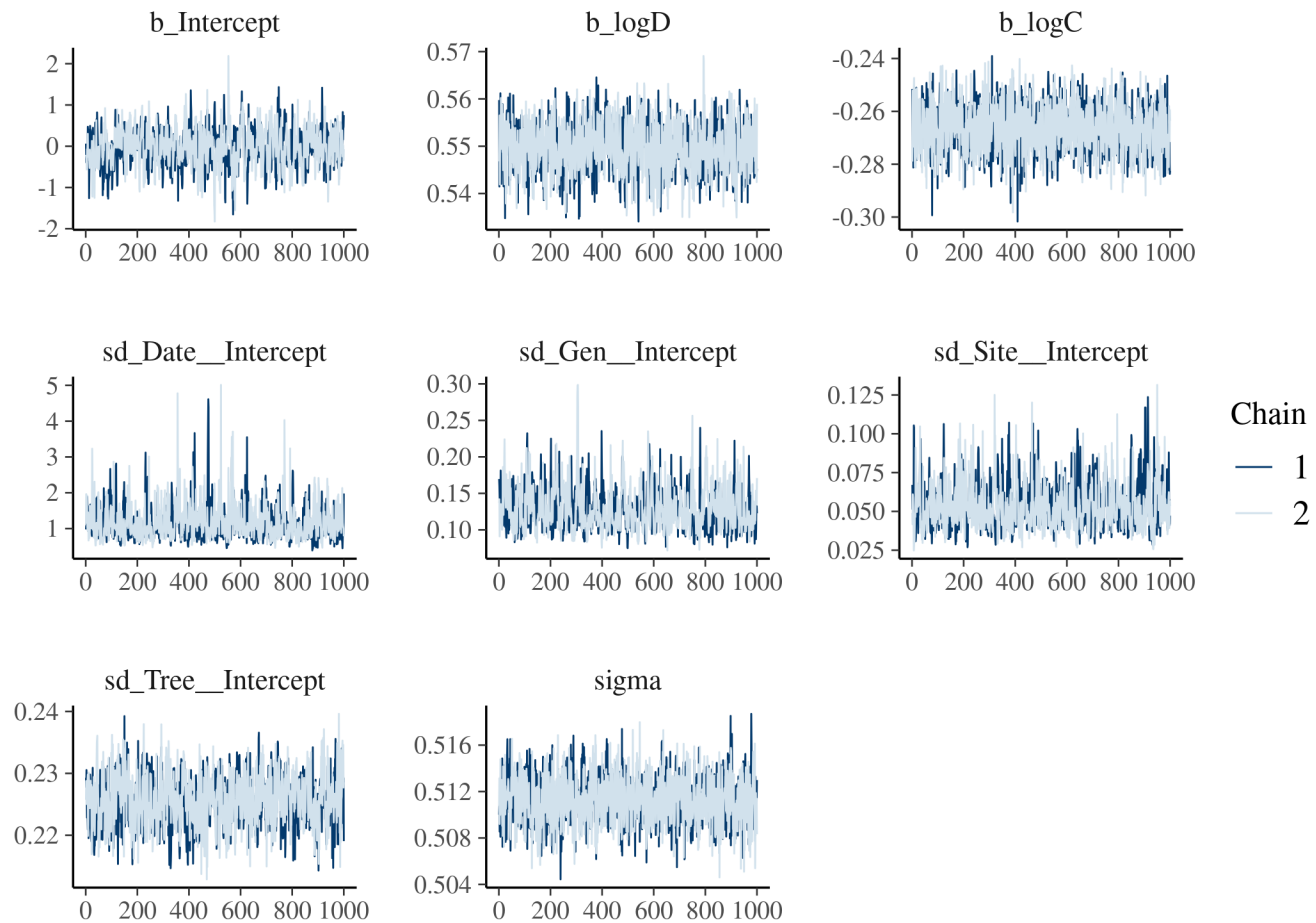
```
## No divergences to plot.
```
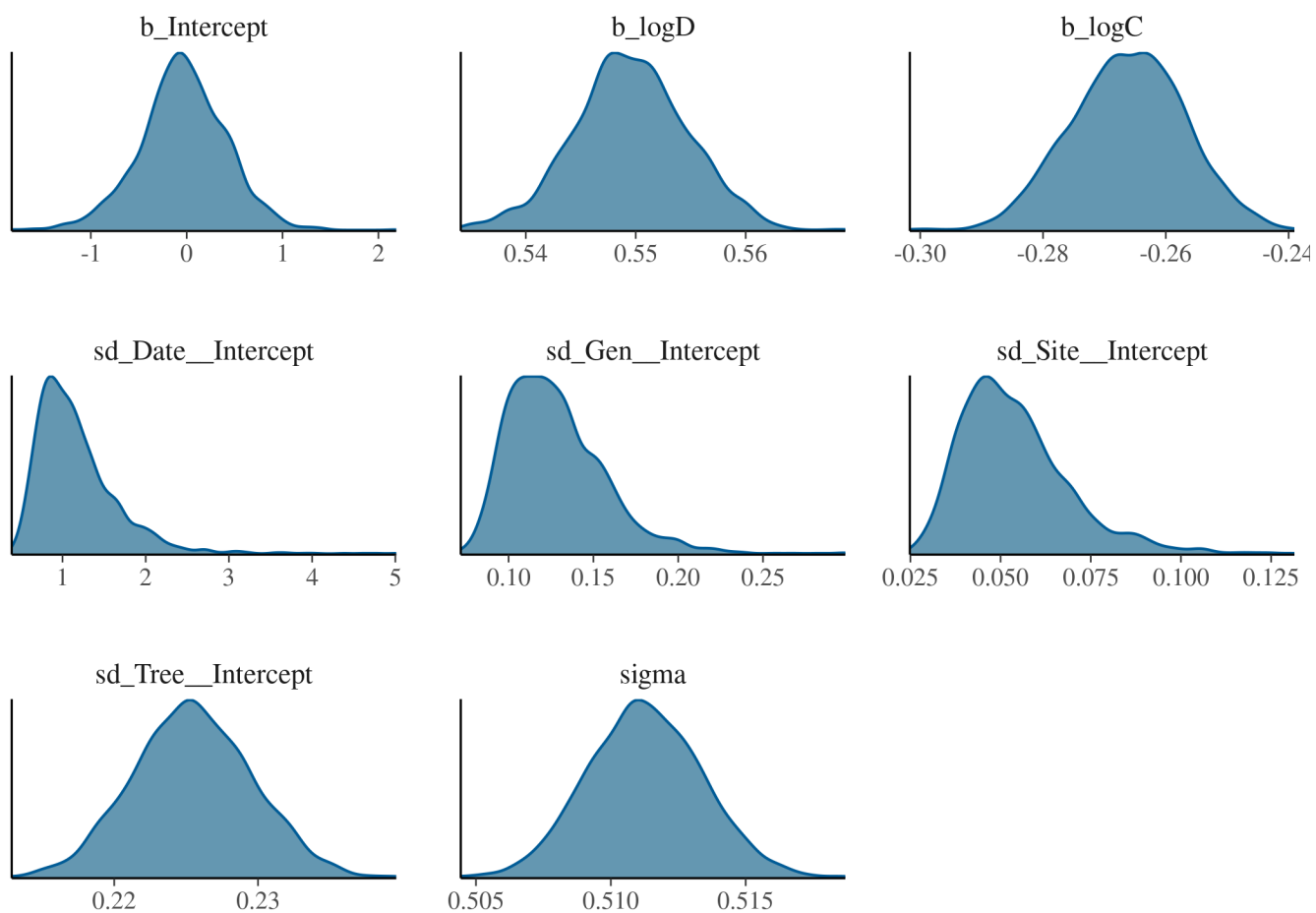
Figure 4.1: Trace of the posteriors

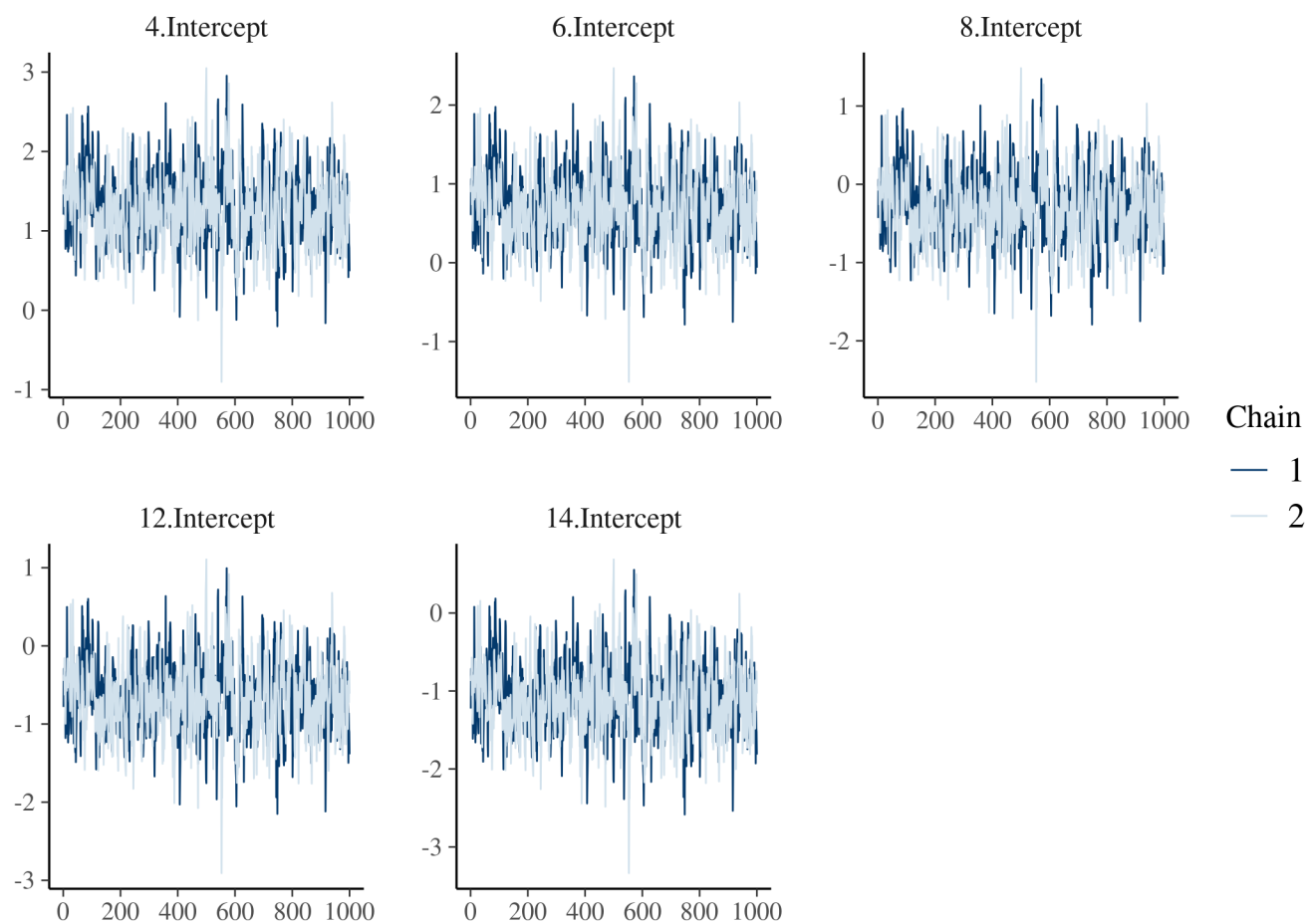Figure 4.2: Density of the posteriors

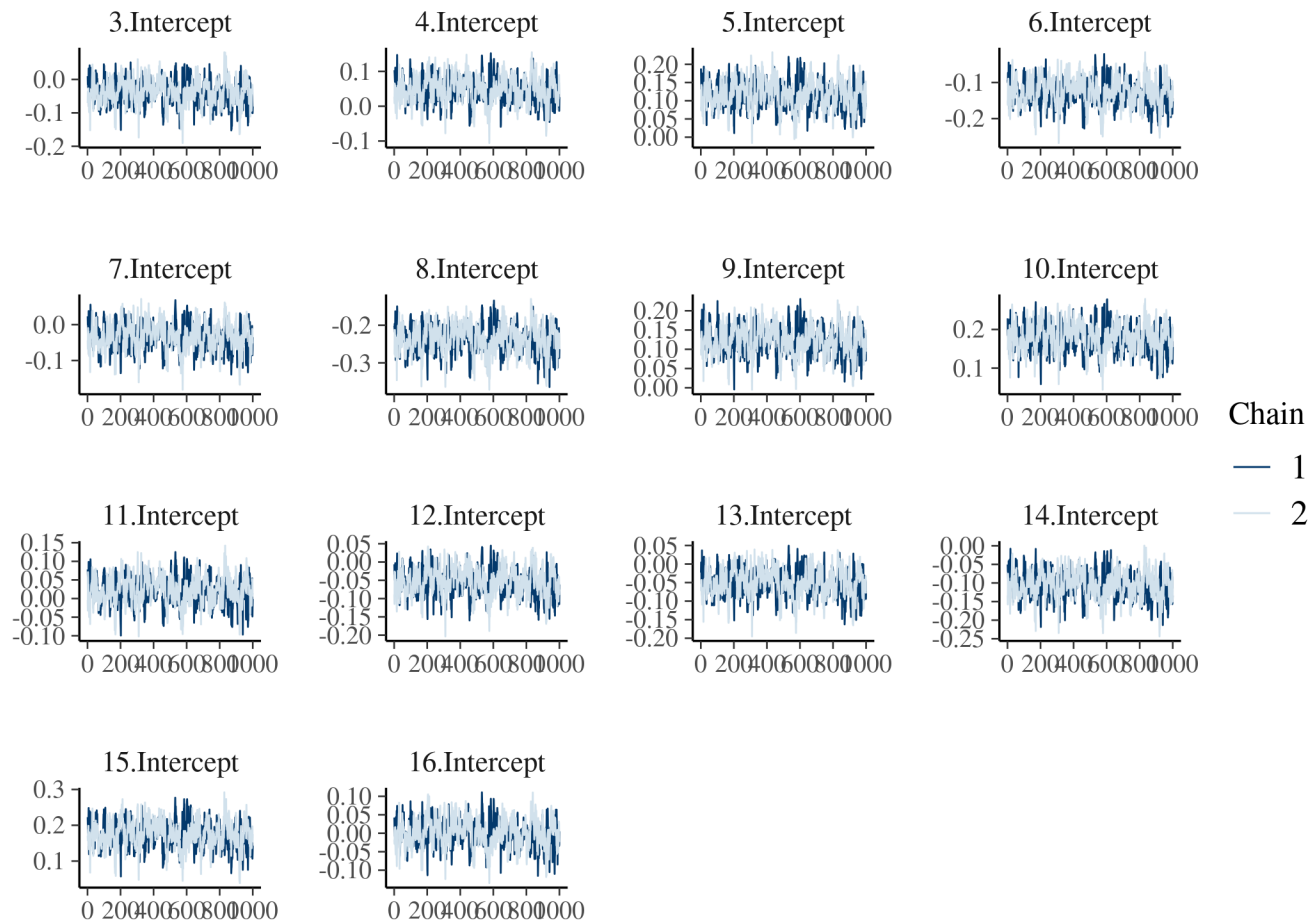Figure 4.3: Trace of the temporal random effects

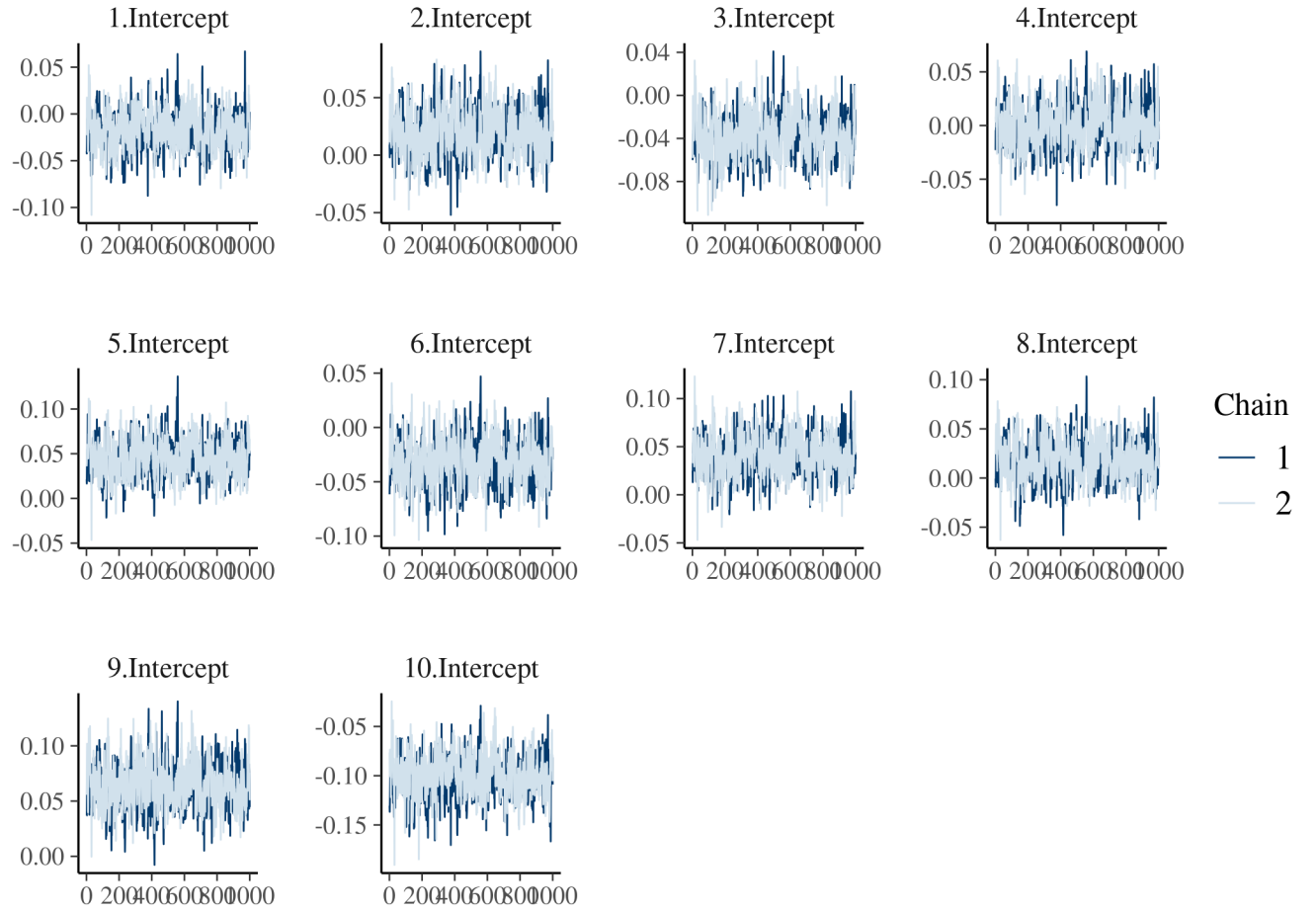Figure 4.4: Trace of the genotype random effects

Figure 4.5: Trace of the site random effects.
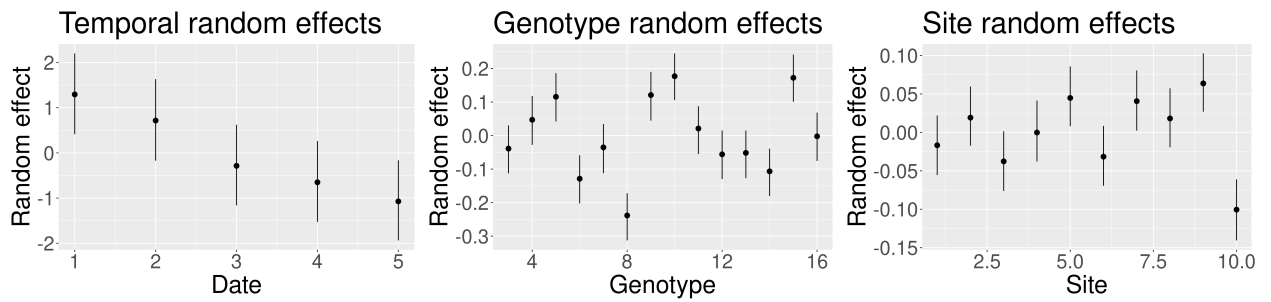
Figure 4.6: Values of the random effects.

We obtained a good model convergence, and found that the two most important contributors to variance (apart from the residuals) were the date and individual identity. The effect of the genotype is quite small, and the effect of site is even smaller. The temporal random effects declined with the date, showing that the effect of the date on growth is negative (the older the trees become, the less they can grow). Therefore, the competition index C did not fully capture the effect of competition on growth.

Table 1: Mean posteriors of the model and their estimation errors.

	Intercept (β_0)	Diameter (β_1)	Competition (β_2)	Individual variance (V_i)	Site variance (V_s)	Genetic variance (V_g)	Temporal variance (V_t)	Residuals variance (V)
Estimate	-3.6e-02	5.5e-01	-2.7e-01	2.3e-01	5.4e-02	1.3e-01	1.2e+00	5.1e-01
Estimation error	4.5e-01	5.1e-03	9.2e-03	4.1e-03	1.5e-02	2.8e-02	4.9e-01	2.1e-03

In order to visualise what part of variation is represented by which effect, we compute the proportion of unexplained variance represented by each random effect and the residual variance.
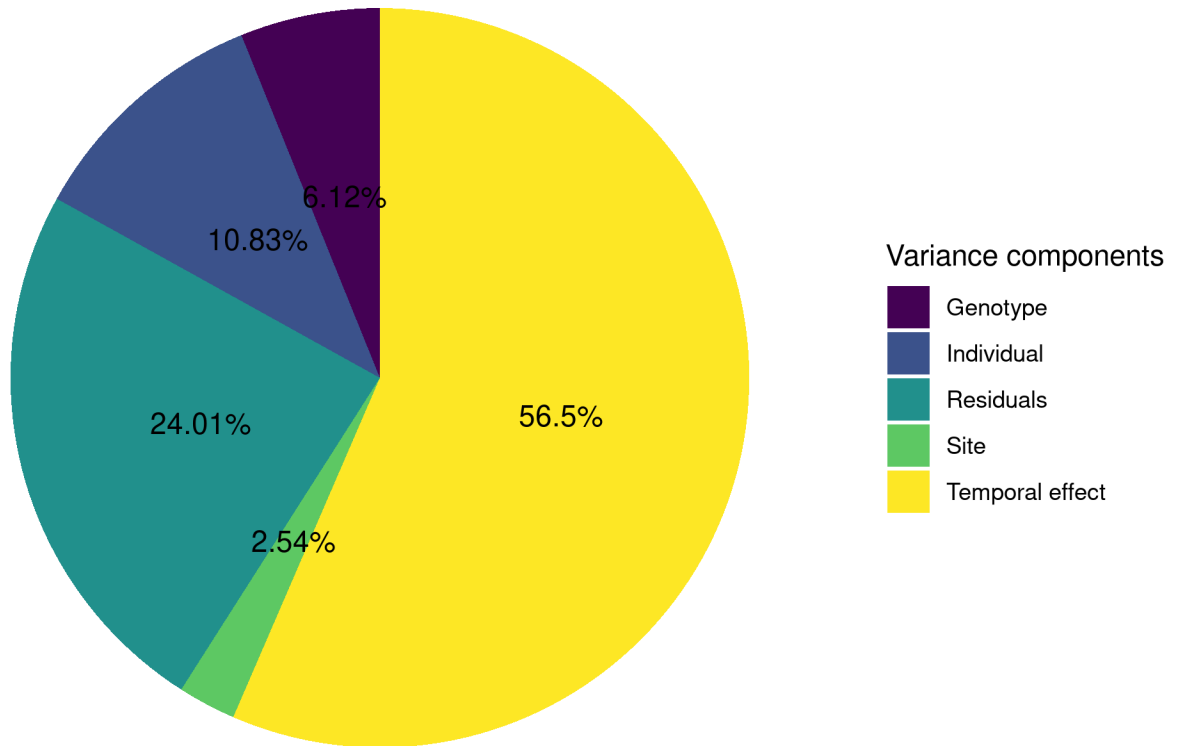

Figure 4.7: Proportion of each variance component of the unexplained variance.

The highest part of variance is explained by the census date. This is due to the negative effect of competition for light on growth, which increases with the growth of the trees. This means that our competition index does not completely capture competition.

The next random effect that explains most variance (apart from residuals) is the individual effect. This shows that there is individual variability even if the trees are clones and therefore that individual variability can be

due to exogenous causes. This individual effect can be due to the micro-environment where it is positioned, but also to some individual history, here seedling manipulation and plantation. As the experimental design aimed at minimizing environmental variations and selected productive genotypes able to accommodate several environmental conditions [1], this dataset is a strongly conservative test case for our hypothesis.

The genotype explained 6.1% of the variance, which is surprising as some striking differences were denoted between clones beforehand. Therefore the impact of individual identity on growth is stronger than the effect of genotype. We did not find any relevant explanation for the mean random effect of each genotype.

Finally, the site has the littlest impact with 2.5% of the variance explained, this means that the physical environment between sites is quite homogeneous, or that the physical environment does not play a big role in growth. We know that at least three sites stand out : two for having clayey soil (8 and 9), and one for having been struck by lightning (10), leading to high mortality. Looking at the mean random site effects, we can see that the site 10 indeed has a lower random effect, and that sites 8 and 9 have a positive random effect. However, site 8 does not really depart from other sites. We can detect some site effects, but it is not always clear.

High estimation error for the intercept and the date random effect must be noted.

5   Conclusion

Overall, there is IV within a clone. This shows that the environmental factors (in a broad sense : not genetic) have an impact on growth and that IV can indeed emerge within a clone.

6   Code implementation

The whole analysis is done using the R language [4] in the Rstudio environment [5]. The tables are made with the kableExtra package [6], the figures with the package ggplot2 [7], and the code uses other packages of the Tidyverse [8] (dplyr [9], lubridate [10], magrittr [11]) and other R packages (here [12], bayesplot [13]). The pdf and html documents are produced thanks to the R packages rmarkdown [16], knitr [19] and bookdown [20].

References

1   Maire, G. le *et al.* (2019) Light absorption, light use efficiency and productivity of 16 contrasted genotypes of several Eucalyptus species along a 6-year rotation in Brazil. *Forest Ecology and Management* 449, 117443

2   Bürkner, P.-C. (2017) Brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80,

3   Bürkner, P.-C. (2018) Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal* 10, 395

4   R Core Team (2021) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing.

5   RStudio Team (2021) *RStudio: Integrated development environment for r*, RStudio, PBC.

6   Zhu, H. (2021) *kableExtra: Construct complex table with 'kable' and pipe syntax*,

7   Wickham, H. (2009) *Ggplot2: Elegant graphics for data analysis*, Springer.

8   Wickham, H. *et al.* (2019) Welcome to the Tidyverse. *Journal of Open Source Software* 4, 1686

9   Wickham, H. *et al.* (2021) *Dplyr: A grammar of data manipulation*,

10   Grolemund, G. and Wickham, H. (2011) Dates and Times Made Easy with **lubridate**. *Journal of Statistical Software* 40,

11   Bache, S.M. and Wickham, H. (2020) *Magrittr: A forward-pipe operator for r*,

12   Müller, K. (2020) *Here: A simpler way to find your files*,

13   Gabry, J. and Mahr, T. Bayesplot: Plotting for bayesian models. (2021)

14   Allaire, J. *et al.* Rmarkdown: Dynamic documents for R. (2020)

15   Xie, Y. *et al.* (2019) *R Markdown: The definitive guide*, CRC Press, Taylor; Francis Group.

16   Xie, Y. *et al.* (2020) *R markdown cookbook*, (1st edn) Taylor; Francis, CRC Press.

17   Xie, Y. (2021) *Knitr: A general-purpose package for dynamic report generation in r*,

18   Xie, Y. (2015) *Dynamic documents with R and Knitr*, Second edition.CRC Press/Taylor & Francis.

19   Xie, Y. (2014) Knitr: A Comprehensive Tool for Reproducible Research in R. In *Implementing reproducible research* (Stodden, V. et al., eds), CRC Press, Taylor & Francis Group

20   Xie, Y. (2017) *Bookdown: Authoring books and technical publications with R Markdown*, CRC Press.