

# Analysis of an Eucalyptus clonal plantation dataset

## Contents

<b>1</b>	<b>The dataset</b>	<b>1</b>
<b>2</b>	<b>Competition index</b>	<b>4</b>
<b>3</b>	<b>Statistical growth model</b>	<b>4</b>
3.1	Results of the model and variance partitioning . . . . .	5
<b>4</b>	<b>Conclusion</b>	<b>12</b>
<b>5</b>	<b>Code implementation</b>	<b>13</b>
	<b>References</b>	<b>13</b>

In this analysis we tested whether intraspecific variability (IV) in observed individual tree growth can emerge from the environment only. To this aim, we used a clonal experimental setup. The EUCFLUX common garden is a clonal trial where 14 Eucalyptus genotypes were grown in a replicated, statistically-sound design. One of its main goals is to determine and compare the productivity of each genotype. Our hypothesis was that IV in tree growth mainly results from responses to environmental factors, and not only from intrinsic genetic factors. Therefore, we used the EUCFLUX dataset in order to quantify IV (within genotypes) for growth, i.e. in a dataset where genetically-driven IV is nil. Following our hypothesis, we expected to detect IV in growth within single genotypes.

## 1 The dataset

The EUCFLUX experiment is located in Brazil, in the state of São Paulo. It includes 14 genotypes of 5 different species or hybrids, planted at the same date and grown in 10 replicated blocks of 100 trees each, which were monitored during 6 years. Each replicated block is distant by up to 1.5 km in a 200 ha stand showing some small changes in soil properties. The details of the setup can be found in (Maire et al. 2019). We used the DBH measured during 6 complete censuses in order to compute mean annual growth in mm/year at five points in time. We computed the mean annual growth of each tree in mm/y as the difference between two consecutive censuses divided by the time between the two censuses. In case of mortality of the tree between two censuses, the data was discarded. We computed the neperian logarithm of diameter and growth (with a constant for growth in order to avoid undefined values).

The dataset included 64125 growth estimates corresponding to 13531 trees.

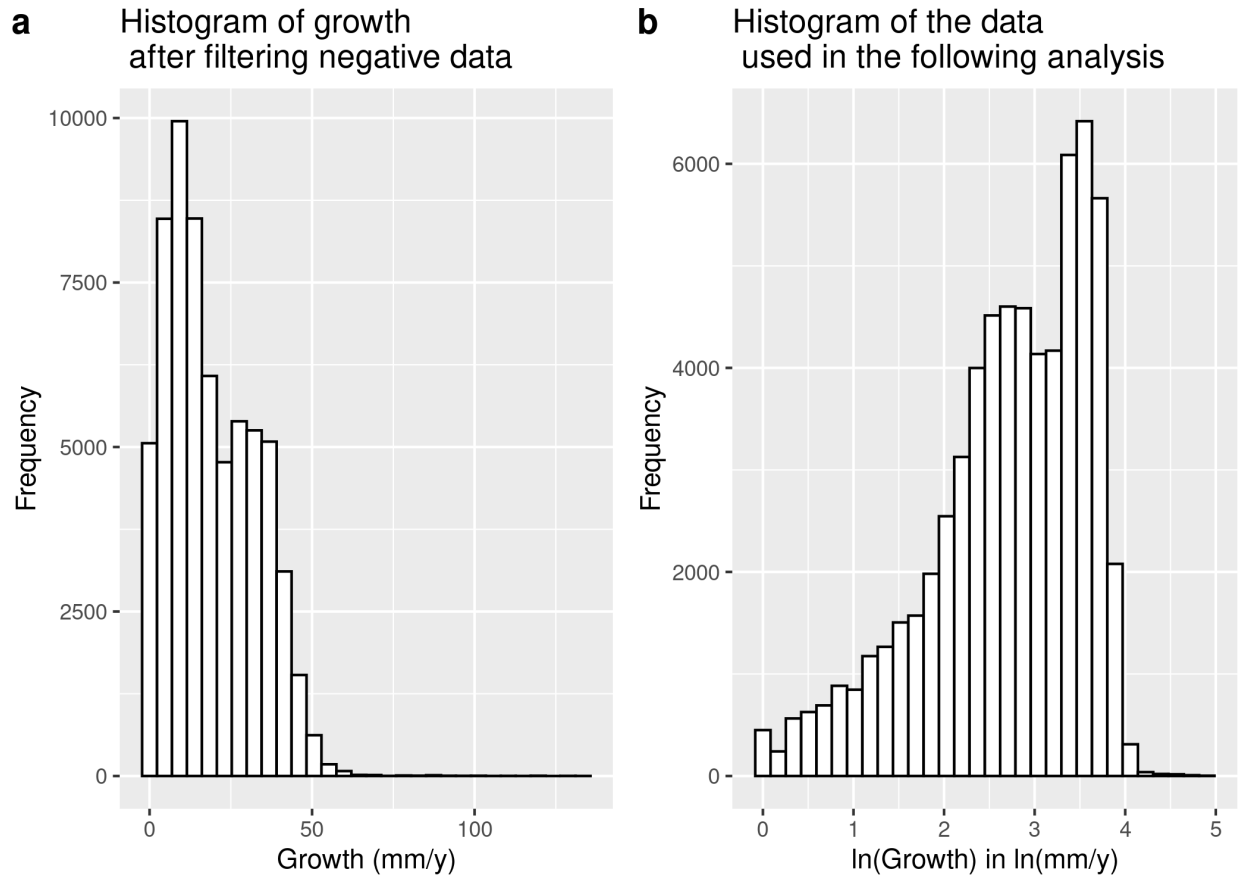


Figure 1.1: The original data without negative growth values (A) and the log-transformed growth (B).

Figure 1.1 shows the distribution of growth after removing negative values, and the latter with log-transformed values.

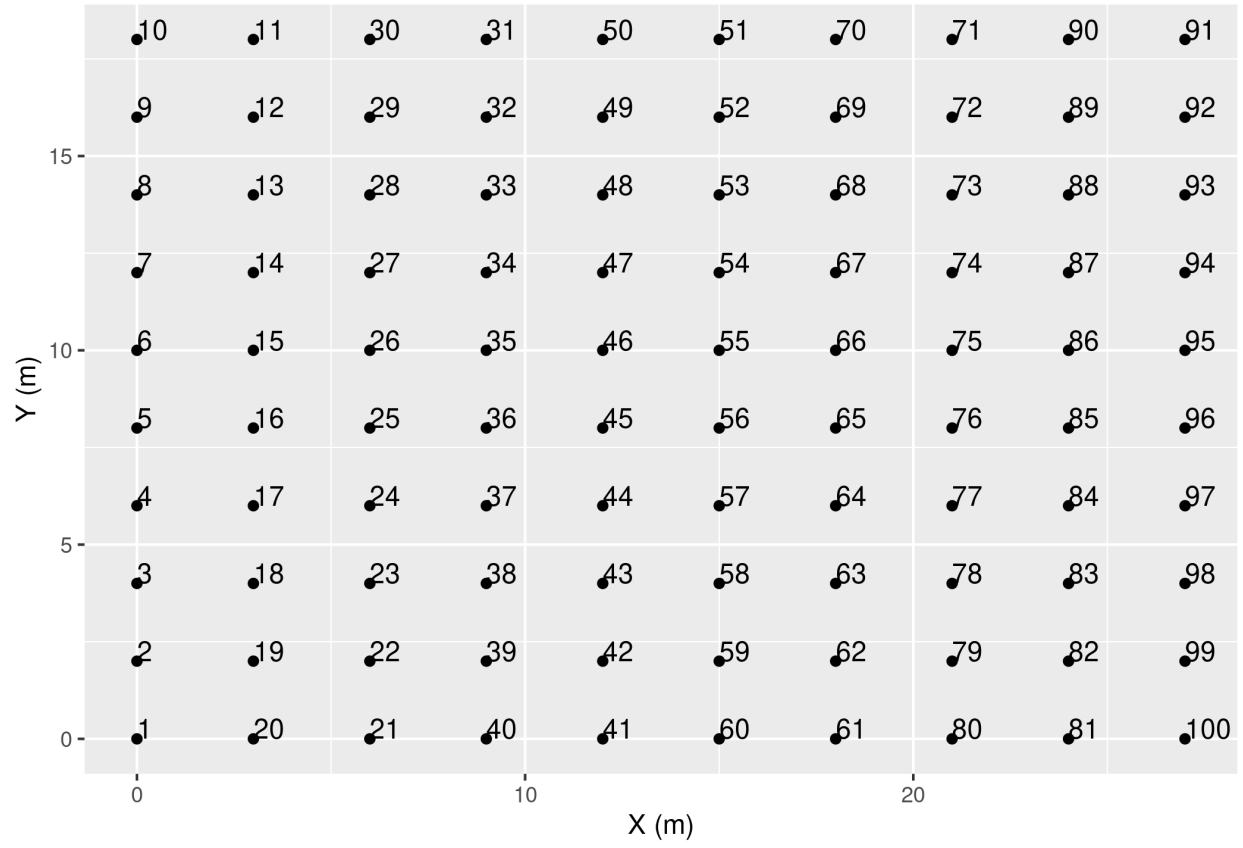


Figure 1.2: Design of a plot. Each point is a tree and the associated number is the tag of the tree.

Figure 1.2 shows the disposition of the trees in a single plot. There are 14 genotypes times 10 repetitions, so 140 plots with this same design.

## 2 Competition index

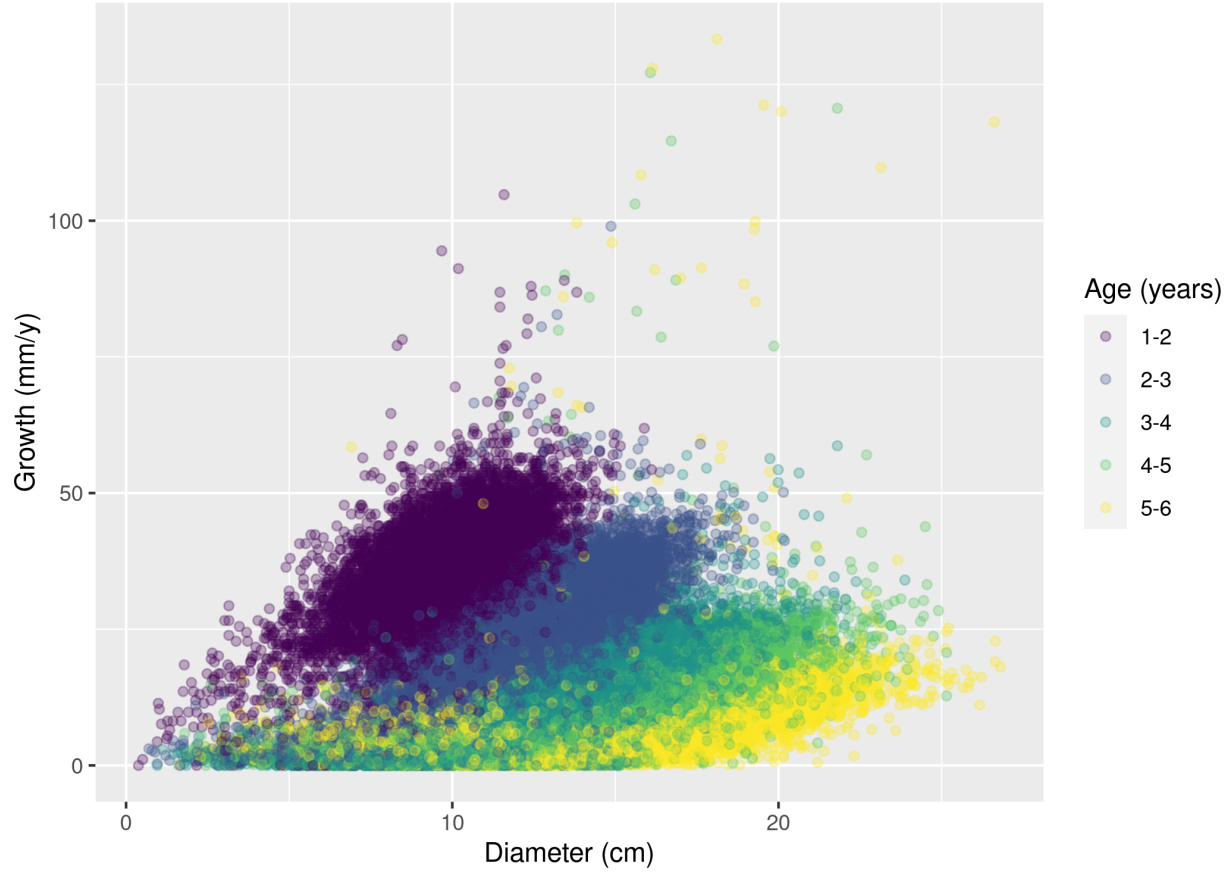


Figure 2.1: Plot of the growth versus the diameter. Each colour represents a tree age.

Figure 2.1 shows the age of the trees has a big influence on the values of growth but also on the relationship between growth and diameter : the slope is 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 computed a competition index  $C_{i,t}$  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 (BA) of the 8 direct neighbours divided by the area of the rectangle that comprises all the direct neighbours. It was then log-transformed. Dead neighbours are considered as having a null BA.

$$C_{i,t} = \sum BA_{neighbours(i,t)}$$

## 3 Statistical growth model

In order to partition the variance of individual growth data, we built a model incorporating a fixed effect on the intercept ( $\beta_0$ ), on the slope of diameter D ( $\beta_1$ ), and on the competition index C ( $\beta_2$ ) and several random effects, namely temporal (date of census,  $b_t$ ), individual (tree identifier,  $b_i$ ), spatial (block,  $b_b$ ), and genotype ( $b_g$ ).

This model was fitted in a hierarchical Bayesian framework using the brms package (Bürkner 2017, 2018 ) with 10000 iterations, a warming period of 5000 iterations and a thinning of 5. We obtained 1000 estimates per parameter.

Variables were scaled to help the convergence of the model.

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

Priors

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

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

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

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

$$b_b \sim \mathcal{N}(\text{mean} = 0, \text{var} = V_b), \text{iid}$$

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

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

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

Hyperpriors

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

### 3.1 Results of the model and variance partitioning

After convergence of the model (checked visually), we examined the proportion of the residual variance (variation of the response variables that is not explained by the covariates , i.e. in the unexplained part of the statistical model) related to each random effect, and this enabled us to perform a residual variance partitioning.

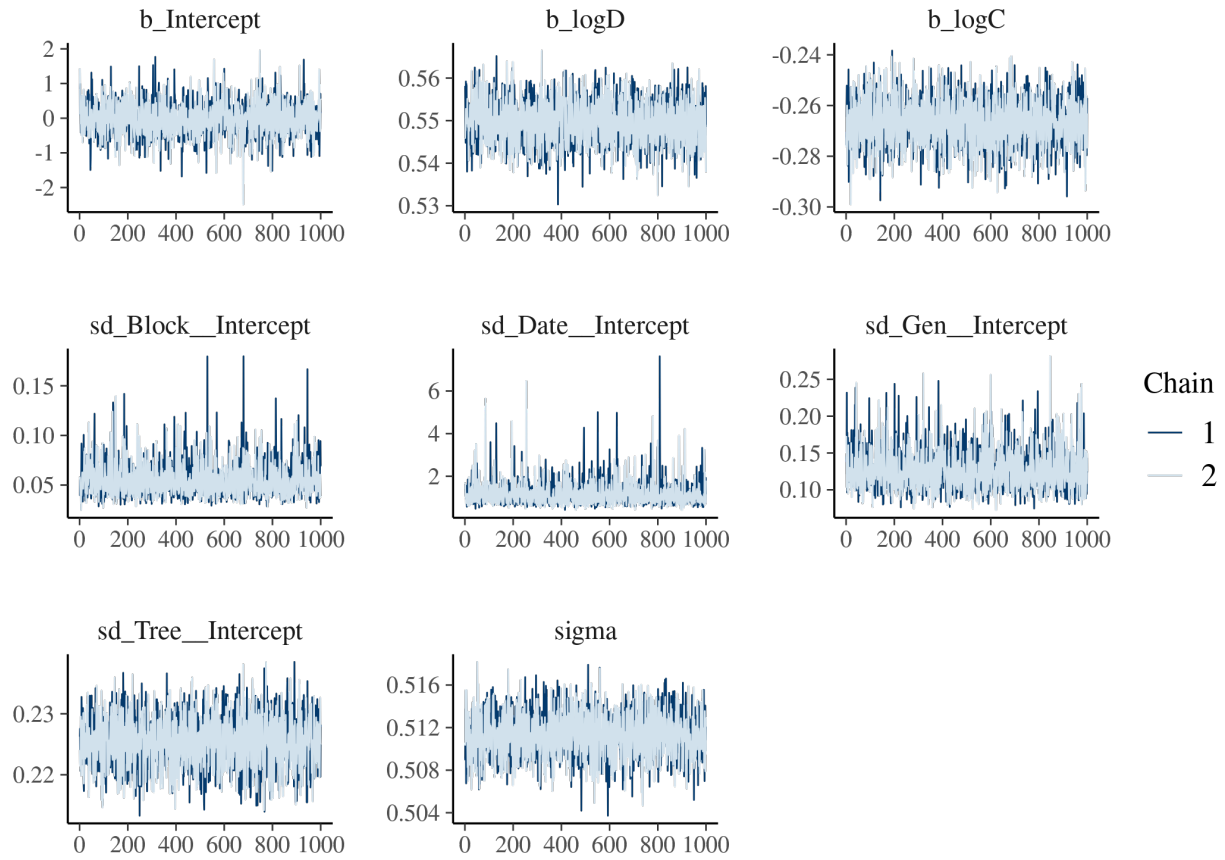


Figure 3.1: Trace of the posteriors of the inferred parameters

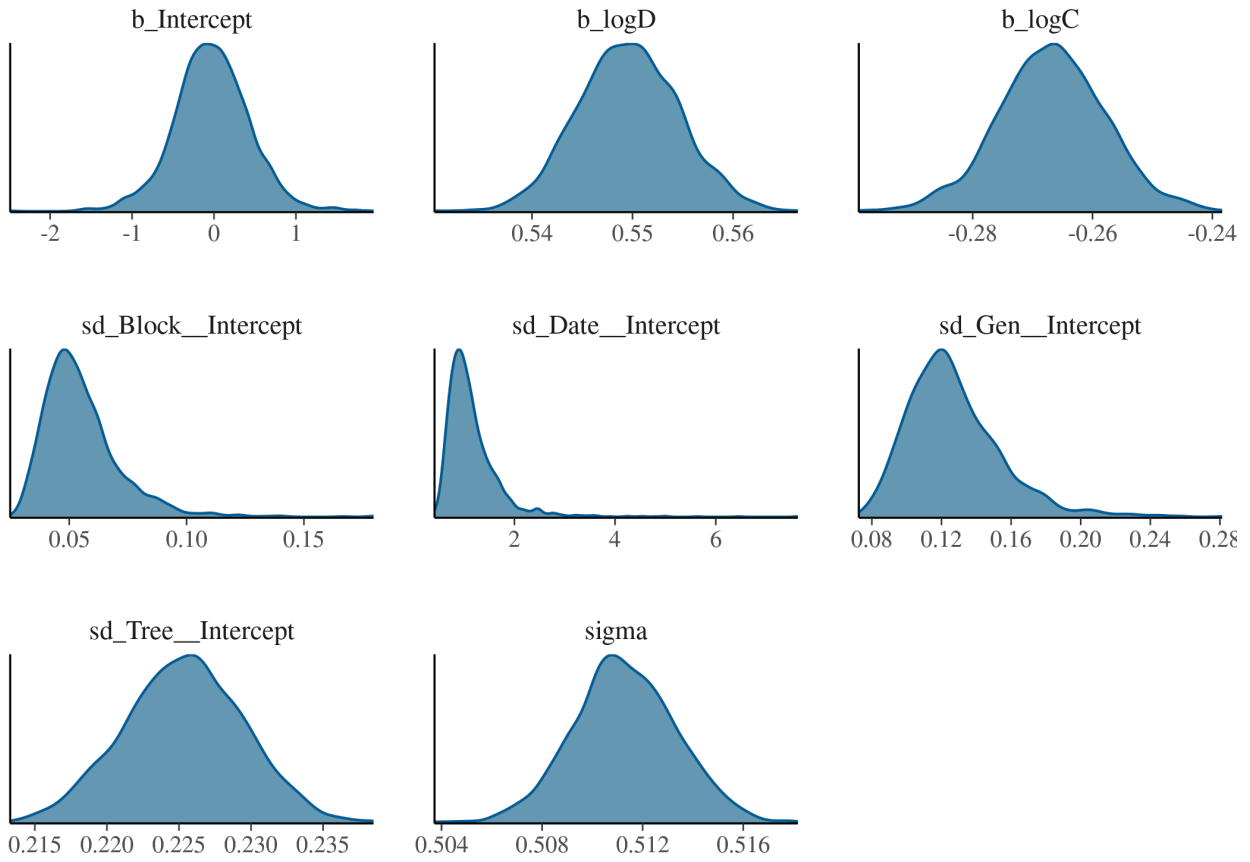


Figure 3.2: Density of the posteriors of the inferred parameters

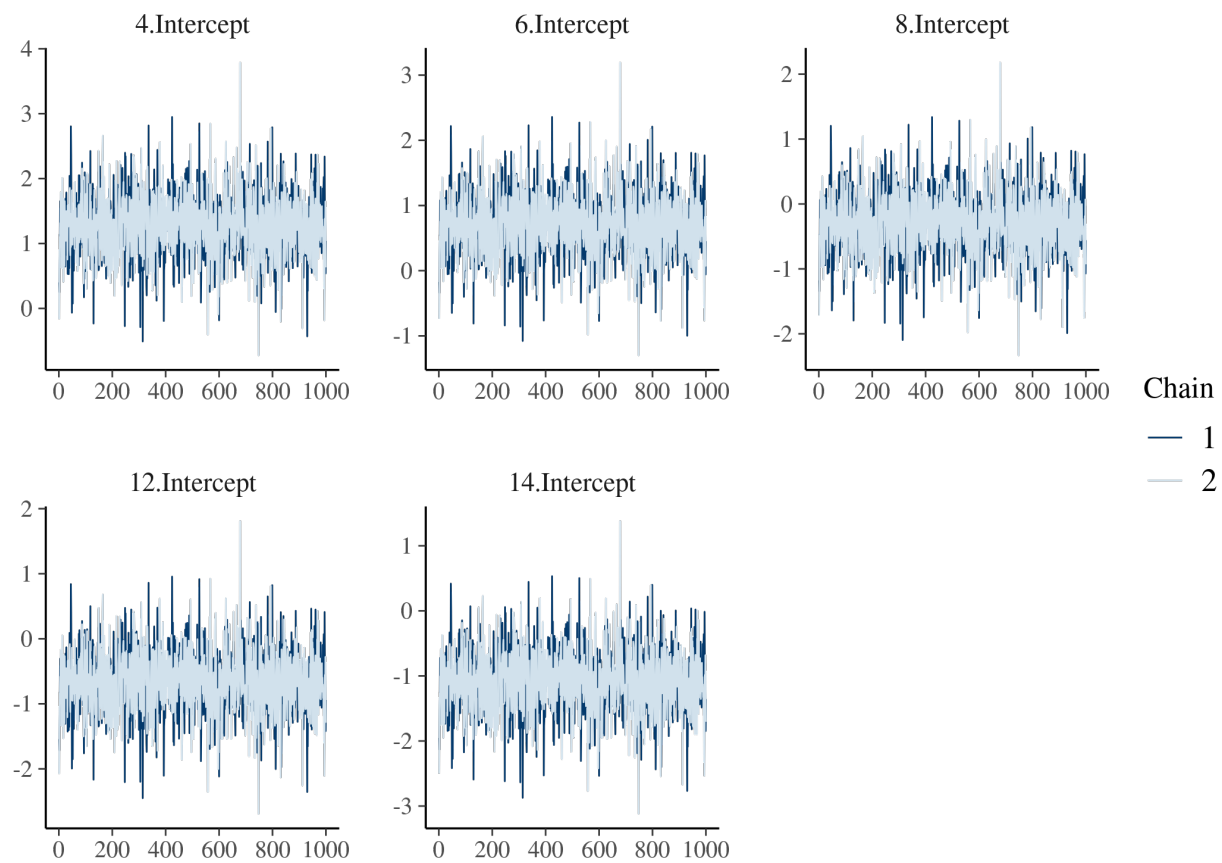


Figure 3.3: Trace of the temporal random effects



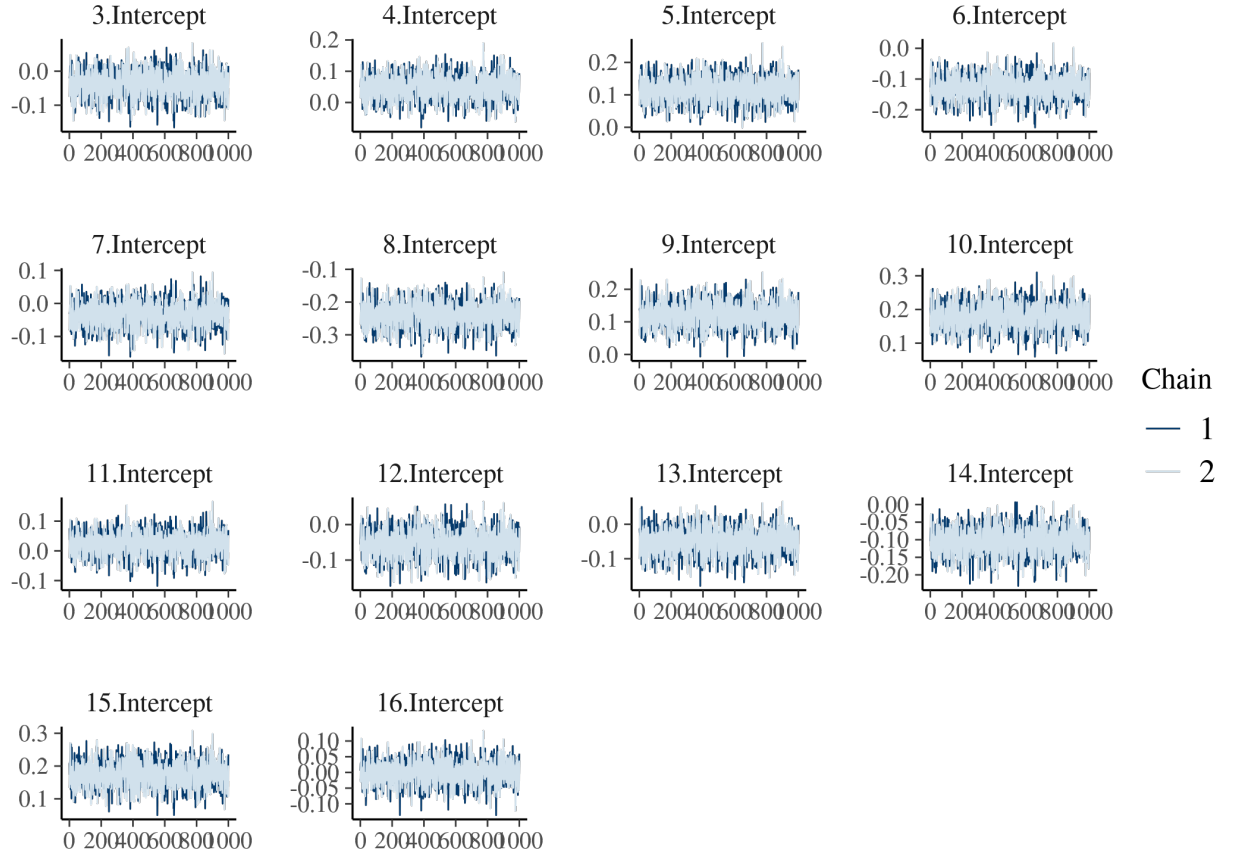


Figure 3.4: Trace of the genotype random effects

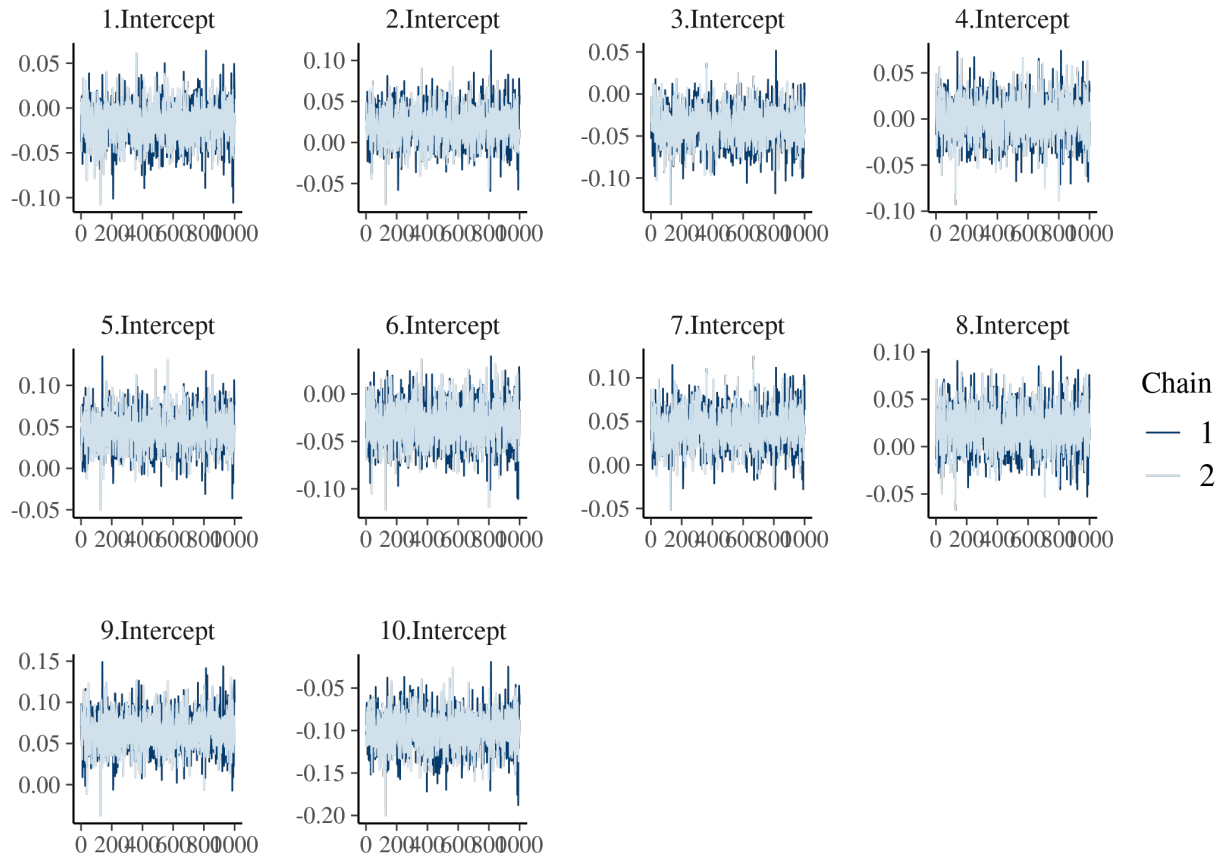
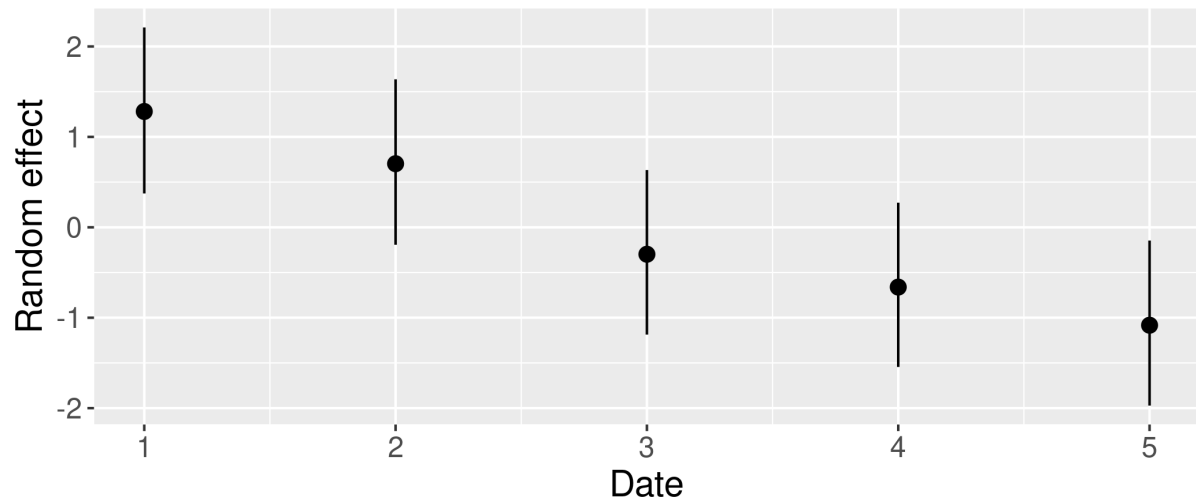
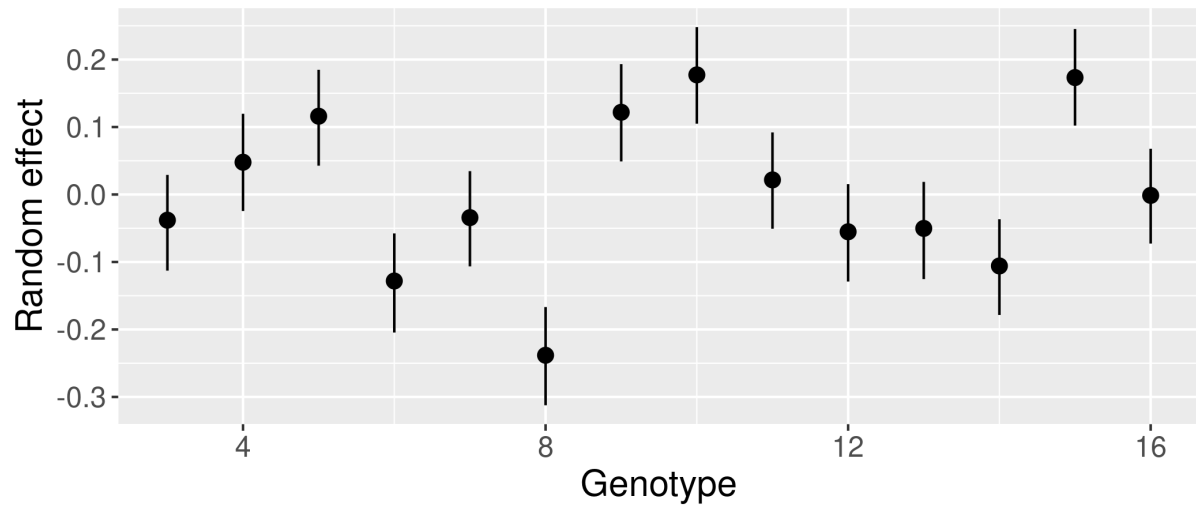


Figure 3.5: Trace of the spatial (block) random effects.

**a** Temporal random effects



**b** Genotype random effects



**c** Block random effects

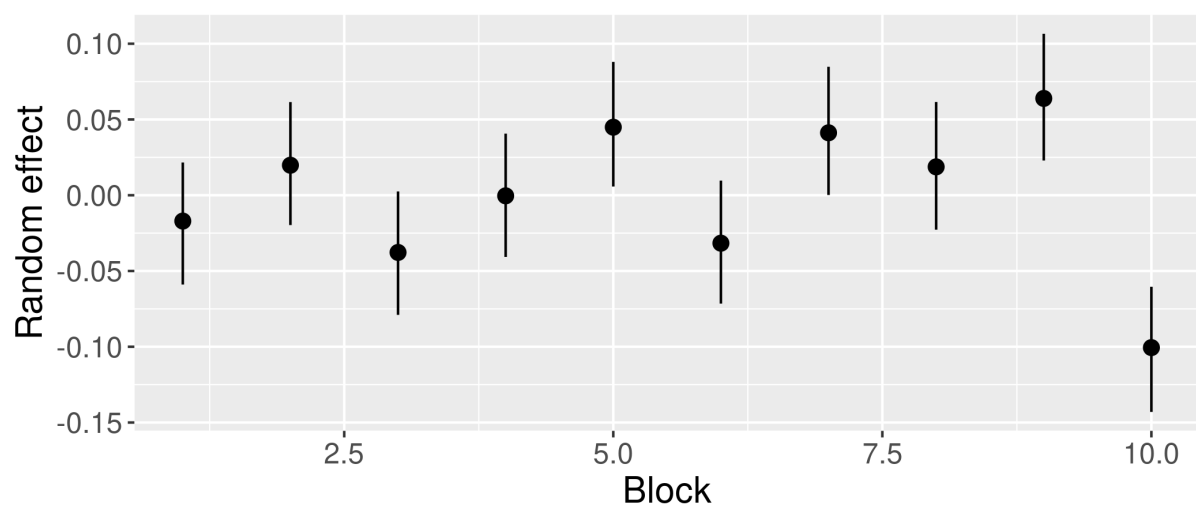


Figure 3.6: Mean values and 95% confidence interval of the temporal, genetic and spatial and random effects.

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

	Intercept ( $\beta_0$ )	Diameter ( $\beta_1$ )	Competition ( $\beta_2$ )	Individual variance ( $V_i$ )	Block variance ( $V_b$ )	Genetic variance ( $V_g$ )	Temporal variance ( $V_t$ )	Residuals variance ( $V$ )
Estimate	-2.5e-02	5.5e-01	-2.7e-01	2.3e-01	5.5e-02	1.3e-01	1.2e+00	5.1e-01
Estimation error	4.6e-01	5e-03	9e-03	4.1e-03	1.6e-02	2.7e-02	5.5e-01	2e-03

We found that the two most important contributors to variance were the date and individual identity. High estimation error for the intercept and the temporal random effect must be noted. The proportion of variance represented by each random effect and the residual variance were computed to visualise the variance partition.

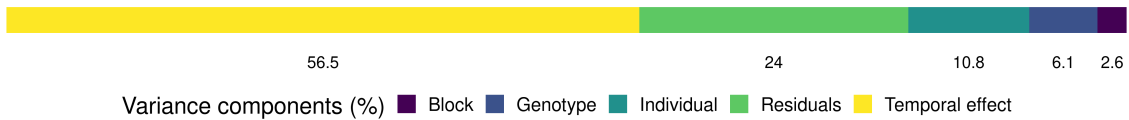


Figure 3.7: Proportion of each variance component of the unexplained variance.

The model showed that individual tree growth was a function of tree size and competition with neighbouring trees, and that variance around this model was mostly due to a temporal effect as well as an individual effect (Table 1, Figure 1.2). The effect of the genotype was quite small, and the effect of the block was even smaller. The temporal random effects declined with time (Figure 3.6, panel A), showing that the effect of the date on growth is negative (the older the trees become, the less they can grow). We attribute this tendency to competition. Therefore, the competition index C did not fully capture the effect of competition on growth. Another explanation is that the diameter slope was not able to fully capture the decrease of tree growth with size, maybe due to geometrical effects of distributing growth around increasing diameter or physiological constraints linked to height. The temporal effect explained the highest fraction of variance (Table 1, Figure 1.2). This could be due to the negative effect of competition for light, water, and/or nutrients on growth, which increases with the growth of the trees planted at high densities, and to physiological changes occurring with age.

Importantly, variability between individuals within genotypes (10.8%) is higher than between genotypes (6.1%). This shows that there is an observed individual variability within genotypes even if trees are clones. Estimated individual variability within genotypes can only be caused by exogenous factors in this particular case. Individual effects can be due to the micro-environment where each tree thrives in, but also to some individual history, such as seedling manipulation and plantation.

The block had the littlest impact with 2.5% of the variance explained. This is probably due to the fact that environmental conditions between blocks are quite homogeneous. As the experimental design aimed at minimizing environmental variations and selected productive genotypes able to accommodate several environmental conditions (Maire et al. 2019), this dataset is a strongly conservative test case for our hypothesis.

## 4 Conclusion

Overall, we found that there is IV within clonal tree plantations (Figure 4.1). This shows that IV can be due only to the environmental factors varying between individuals, proving that the source of IV is not necessarily intrinsic.

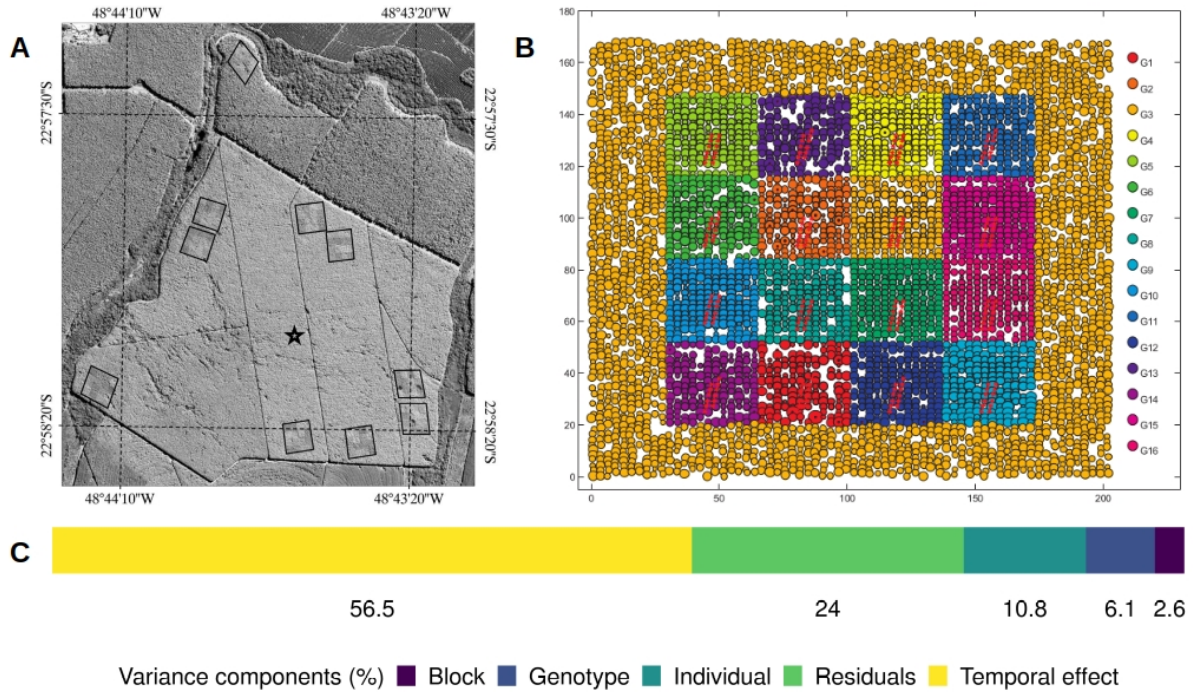


Figure 4.1: The EUCFLUX setup and the variance partitioning. Site (A; each square represents a block) and organisation of a block (B; each coloured square represents a genotype) in the EUCFLUX experiment. 16 clones are represented in B, but only 14 were used since the last two are from seed-origin and thus not genetically identical. (C) Variance partitioning of tree growth for a common garden experiment with various Eucalyptus clones

## 5 Code implementation

The whole analysis was conducted using the R language (R Core Team 2021) in the Rstudio environment (RStudio Team 2021). The tables were 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), lubridate (Grolemund and Wickham 2011), magrittr (Bache and Wickham 2020)) and other R packages (here (Müller 2020), bayesplot (Gabry and Mahr 2021)). The pdf and html documents were produced thanks to the R packages rmarkdown Xie et al. (2020), knitr Xie (2014) and bookdown (Xie 2017).

## References

- Allaire, J., Y. Xie, J. McPherson, J. Luraschi, K. Ushey, A. Atkins, H. Wickham, J. Cheng, W. Chang, and R. Iannone. 2020. Rmarkdown: Dynamic documents for R.
- Bache, S. M., and H. Wickham. 2020. Magrittr: A forward-pipe operator for r.
- 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.

- Gabry, J., and T. Mahr. 2021. Bayesplot: Plotting for bayesian models.
- Grolemund, G., and H. Wickham. 2011. Dates and Times Made Easy with **lubridate**. *Journal of Statistical Software* 40.
- Maire, G. le, J. Guillemot, O. C. Campoe, J.-L. Stape, J.-P. Laclau, and Y. Nouvellon. 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.
- Müller, K. 2020. Here: A simpler way to find your files.
- 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. Springer, New York.
- Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller, S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* 4:1686.
- Wickham, H., R. François, L. Henry, and K. Müller. 2021. Dplyr: A grammar of data manipulation.
- Xie, Y. 2014. Knitr: A Comprehensive Tool for Reproducible Research in R. *in* V. Stodden, F. Leisch, and R. D. Peng, editors. Implementing reproducible research. 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., J. J. Allaire, and G. Grolemund. 2019. R Markdown: The definitive guide. CRC Press, Taylor; Francis Group, Boca Raton.
- Xie, Y., C. Dervieux, and E. Riederer. 2020. R markdown cookbook. First edition. Taylor; Francis, CRC Press, Boca Raton.
- Zhu, H. 2021. kableExtra: Construct complex table with 'kable' and pipe syntax.