Analysis of tropical forest inventory data

Contents

After a theoretical model showed that observed intraspecific variability can emerge from environmental heterogeneity and an experimental clonal dataset showed that a significant part of intraspecific variability in growth is not due to the genotype, we here want to examine the features of three natural tropical forest. One is located in French Guiana (Paracou), another is located in the Whestern Ghats in India (Uppangala) and the last one is located in Panama (Barro Colorado island, denoted BCI). The analyses will follow the ones

done on the virtual dataset generated with the theoretical model. Indeed, this will help making connexions between theory and observations. First, we will estimate the intraspecific variance of annualised growth in the three datasets using a hierachical bayesian model. Then, we will test if we can detect spatial structure in individual mean growth. Finally, we will test if intraspecific variability is higher than interspecific variability.

1 Datasets

1.1 Paracou

The Paracou forest is located in French Guiana. It is one of the best-studied lowland tropical forests in the Guiana Shield region. It belongs to the Caesalpiniaceae facies and has amongst the highest alpha-diversity in the Guiana shield with 150-200 tree species per hectare in inventories of trees with diameter at breast height (DBH) $\geq$ 10 cm [2]. The studied area gathers 768 species on 93,75 ha corresponding to 254 genus of 65 plant families. The Guiana Shield is characterized by Pre-Cambrian granitic and metamorphic geological formations, highly eroded. It is associated with gently undulating landscapes and a very dense hydrographic system. Paracou forest lies in a hilly area, on a formation called "série Armina" characterized by schists and sandstones and locally crossed by veins of pegmatite, aplite and quartz. The topography of the site consists of small hills separated by narrow ($<$ 5 m wide) sandy waterbeds. The altitude varies from 5 to about 45 m above sea level [2]. The mean annual temperature is of 26 °C and winds are generally weak. There is a well marked dry season (from mid-August to mid-November) and a long rain season with a short drier period between March and April (mean annual rainfall of 3,041 mm). Different study programs have been led at the Paracou site, which is managed by the CIRAD but is open to the scientific community. Here, we have used a silvicultural experiment called the "disturbance experiment" under which 15 plots of 6,25 ha were exposed to four different logging conditions between 1986 and 1988. Since then, cartesian coordinates, DBH, species identity and survival for each tree with a DBH $\geq$ 10 cm has been collected every one or two years, during the dry season (mid-August to mid-November). This represents about 46000 trees in total.

1.2 Uppangala

Uppangala Permanent Sample Plot (UPSP) is located in South-East Asia, in the Western Ghats of India, established in 1089 by the French Institute of Pondicherry in the Kadamakal Reserve Forest, in the Pushpagiri Wildlife Sanctuary, in Karnataka state, India [3]. It is a low altitude (500-600 m) wet evergreen monsoon Dipterocarp forest [4]. The studied area, of 5.07 ha, is quite steep, with a mean slope angle of about 30–35°. The plots are five north–south oriented transects, 20 m wide, 180 to 370 m long, and 100 m apart center to center and three rectangular plots which overlap the transects. The transects gather data from 1990 to 2011 and the rectangular plots from 1993 to 2011. The trees with GBH (Girth at Breast Height) $\geq$ 30 cm (equivalent to about 9.5 cm DBH) were measured every 3 to 5 years. The raw dataset contains 3870 trees and 102 species (including 2 morphospecies). This forest is considered as one of the rare undisturbed tropical forests in the world [5].

1.3 BCI

The Barro Colorado Island site is located in central America, in Panama, covered by a lowland tropical moist forest. The zone became an island after a valley was flooded in order to build the Panama Canal, in 1913. It is nowadays considered as the most intensively studied tropical forest in the world. The studied site is a 50 ha plot (500 × 1000 m). It has an elevation of 120 m and is quite flat (most slopes are gentler than 10°). Complete censuses of all trees with DBH $\geq$ 1 cm were performed every 5 years since 1980. In the raw data, there are 328 tree species and 423617 trees and when only taking into account trees with DBH $\geq$ 10 cm, there are 255 species and 37224 trees. The data is available in [6].

2 Data preparation

2.1 Paracou

1) Charge the original data (plots 1 to 15)

Here is the head of the dataset :

Forest	Plot	PlotArea	SubPlot	TreeFieldNum	idTree	Xfield
Paracou	1	6.25	1	1	75475	5
Paracou	1	6.25	1	1	75475	5
Paracou	1	6.25	1	1	75475	5
Paracou	1	6.25	1	1	75475	5
Paracou	1	6.25	1	1	75475	5
Paracou	1	6.25	1	1	75475	5

Yfield	Xutm	Yutm	Lat	Lon	Family	Genus
125.5	285109.6	582983.2	5.271257	-52.93895	Euphorbiaceae	Conceveiba
125.5	285109.6	582983.2	5.271257	-52.93895	Euphorbiaceae	Conceveiba
125.5	285109.6	582983.2	5.271257	-52.93895	Euphorbiaceae	Conceveiba
125.5	285109.6	582983.2	5.271257	-52.93895	Euphorbiaceae	Conceveiba
125.5	285109.6	582983.2	5.271257	-52.93895	Euphorbiaceae	Conceveiba
125.5	285109.6	582983.2	5.271257	-52.93895	Euphorbiaceae	Conceveiba

Species	BotaSource	BotaCertainty	idVern	VernName	CommercialSp	CensusYear
guianensis	Bota	4	601	kusisi	0	1984
guianensis	Bota	4	601	kusisi	0	1985
guianensis	Bota	4	601	kusisi	0	1986
guianensis	Bota	4	601	kusisi	0	1987
guianensis	Bota	4	601	kusisi	0	1988
guianensis	Bota	4	601	kusisi	0	1989

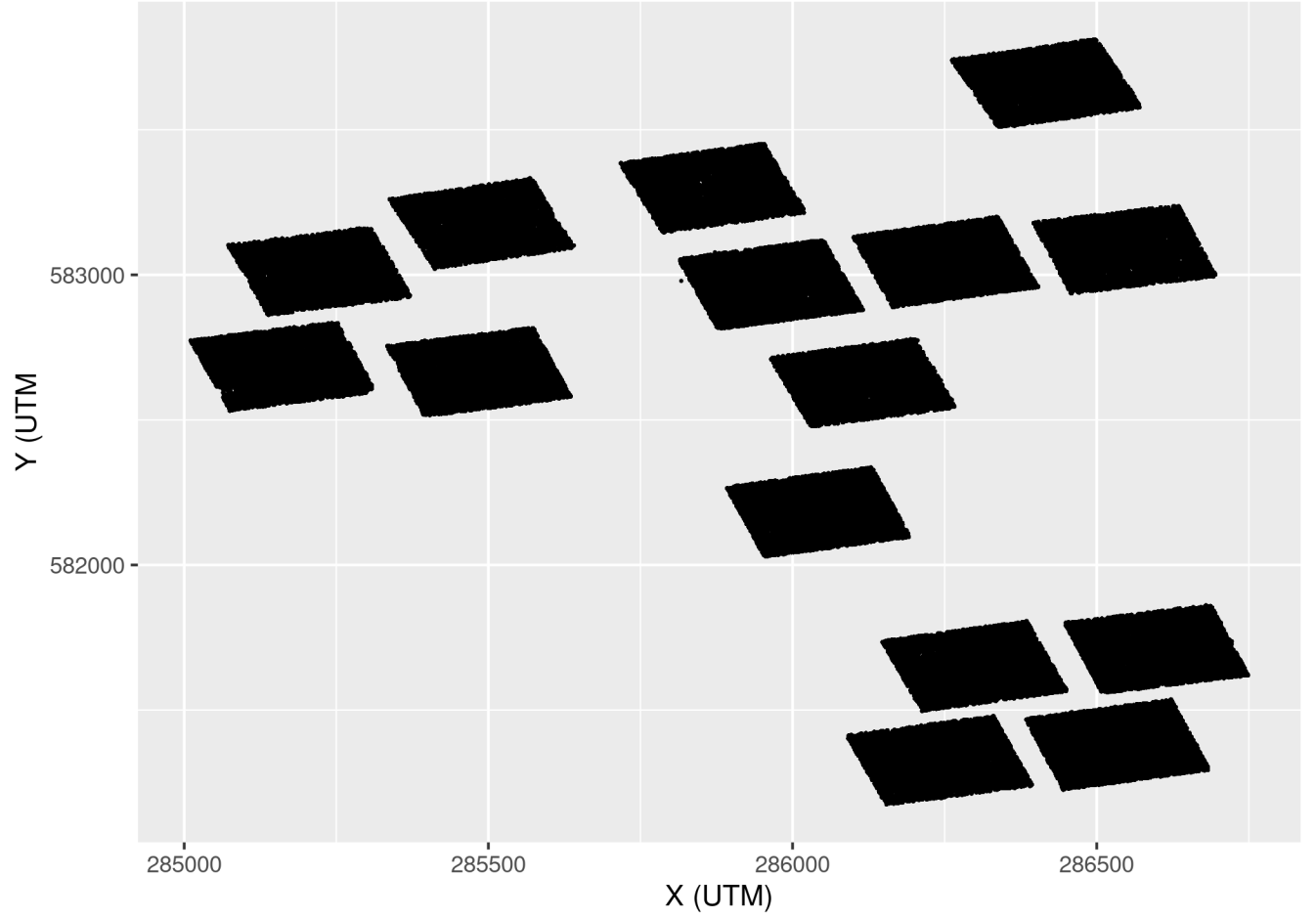| CensusDate	CensusDateCertainty	CodeAlive	MeasCode	Circ	CircCorr	CorrCode
1984-09-01	1	1	0	50.5	50.5	0
1985-11-01	1	1	0	51.5	51.5	0
1986-10-01	1	1	0	51.0	51.0	0
1987-10-01	1	1	0	51.0	51.0	0
1988-09-01	1	1	0	51.0	51.0	0
1989-10-01	1	1	0	51.0	51.0	0

Figure 2.1: Plots at the Paracou site. Each point is a tree.

2) Compute annualised growth between two censuses and rename all indeterminate species with the same name "Indet_sp." Growth is computed as the difference of DBH (Diameter at Breast Height) between two consecutive censuses, divided by the time period between those two censuses.

There are 760 species, 90262 trees and 1448802 observations in this dataset.

Here are the new columns created :

D_tp1	D_tp1_corr	D_t	D_t_corr	Date_tp1	Date_t	DiffD	DiffD_corr
16.39296	16.39296	16.07465	16.07465	1985-11-01	1984-09-01	0.3183099	0.3183099
16.23380	16.23380	16.39296	16.39296	1986-10-01	1985-11-01	-0.1591549	-0.1591549
16.23380	16.23380	16.23380	16.23380	1987-10-01	1986-10-01	0.0000000	0.0000000
16.23380	16.23380	16.23380	16.23380	1988-09-01	1987-10-01	0.0000000	0.0000000
16.23380	16.23380	16.23380	16.23380	1989-10-01	1988-09-01	0.0000000	0.0000000
16.23380	16.23380	16.23380	16.23380	1990-10-01	1989-10-01	0.0000000	0.0000000

DiffDate	G_tp1	G_tp1_corr	Tree	Sp	IdentSpecies
426 days	2.727303	2.727303	P1T75475	Conceveiba_guianensis	86
334 days	-1.739268	-1.739268	P1T75475	Conceveiba_guianensis	86
365 days	0.000000	0.000000	P1T75475	Conceveiba_guianensis	86
336 days	0.000000	0.000000	P1T75475	Conceveiba_guianensis	86
395 days	0.000000	0.000000	P1T75475	Conceveiba_guianensis	86
365 days	0.000000	0.000000	P1T75475	Conceveiba_guianensis	86

3) Remove the observations where growth < -2 mm/an or > 100 mm/an and remove the years before the perturbations were performed and the biodiversity plots were added (1985-1991).

There are 615 species and 78434 and 1053241 trees at this step.

4) Compute mean annual growth. Mean growth is computed as the difference of DBH between the first and the last time a tree was measured, divided by the period between those two measures.

Tree	Sp	IdentSpecies	Mean_G_Tree
P10T120923	Protium_surinamense	461	1.150808
P10T120924	Licania_alba	297	1.029670
P10T120925	Indet_sp	225	4.300386
P10T120926	Indet_sp	225	2.665028
P10T120927	Licania_alba	297	1.029670
P10T120928	Indet_sp	225	4.542661

2.2 Uppangala

1) Charge the original data.

	TreeID	CensusDate	Girth	Recruited	GirthError	GirthNote	SpCode	Xsite
2	A001	1994-11	73.78	NA	NA	NA	vain	650.05
4	A001	1990-11	71.98	NA	G3	new belt fixed	vain	650.05
5	A001	1991-11	73.16	NA	NA	NA	vain	650.05
8	A001	2000-12	73.76	NA	NA	NA	vain	650.05
9	A001	2001-12	74.02	NA	NA	NA	vain	650.05
10	A001	1993-11	73.62	NA	NA	NA	vain	650.05

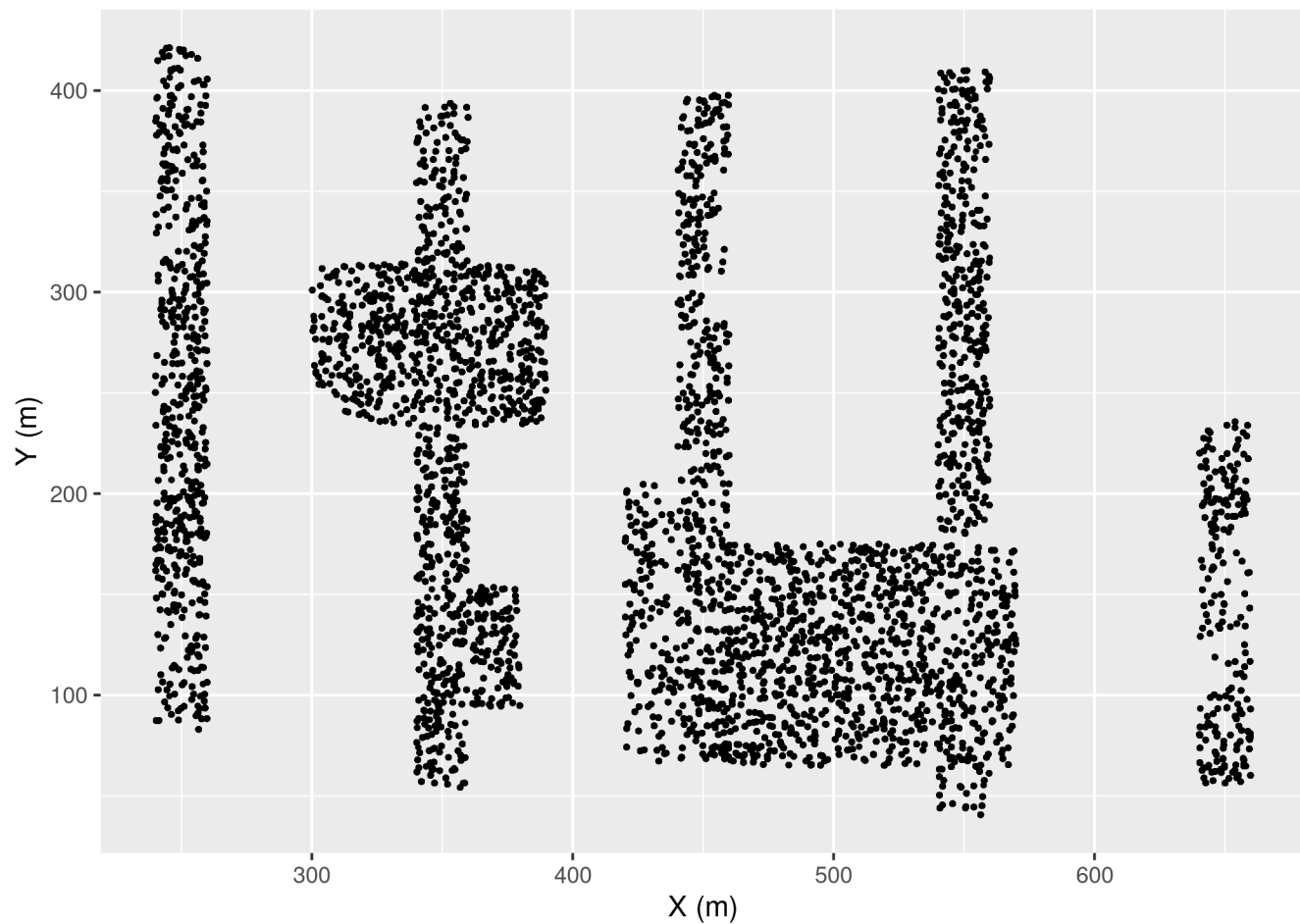|	Ysite	slope	aspect	PotStrata	death.cause	date	Plot	SubPlot
2	56.35	36.13096	67.08345	0	NA	1704	A	Aa01
4	56.35	36.13096	67.08345	0	NA	244	A	Aa01
5	56.35	36.13096	67.08345	0	NA	609	A	Aa01
8	56.35	36.13096	67.08345	0	NA	3926	A	Aa01
9	56.35	36.13096	67.08345	0	NA	4292	A	Aa01
10	56.35	36.13096	67.08345	0	NA	1339	A	Aa01

Figure 2.2: Plots at Uppangala site. Each point is a tree.

There are 102 species, 3870 and 69459 observations in this dataset.

We remove the census dates which were not common for all plots.

2) Compute annualised growth.

3) Remove lines with a growth higher than 100 mm/y or inferior to -2 mm/y.

Thus there are 102 species, 3725 trees and 57921 observations after annual growth computation.

4) Compute mean annual growth.

Here are the new columns :

Date_tp1	D_t	Date_t	DiffD	DiffDate	G_tp1	Mean_G_Tree	Tree
1995-11-01	23.48490	1994-11-01	0.0127324	365 days	0.1273240	0.3466920	A001
1995-11-01	65.72463	1994-11-01	-0.0063662	365 days	-0.0636620	-0.3339966	A002
1995-11-01	31.72913	1994-11-01	0.5602254	365 days	5.6022540	2.3360440	A003
1995-11-01	28.11313	1994-11-01	0.5220282	365 days	5.2202821	2.2204800	A004
1995-11-01	14.51493	1994-11-01	0.0763944	365 days	0.7639437	0.5945981	A005
1995-11-01	10.52969	1994-11-01	0.0572958	365 days	0.5729578	0.4127286	A006

2.3   BCI

1) Load the original data.

There are 328 species, 423617 trees and 3388936 observations in this dataset.

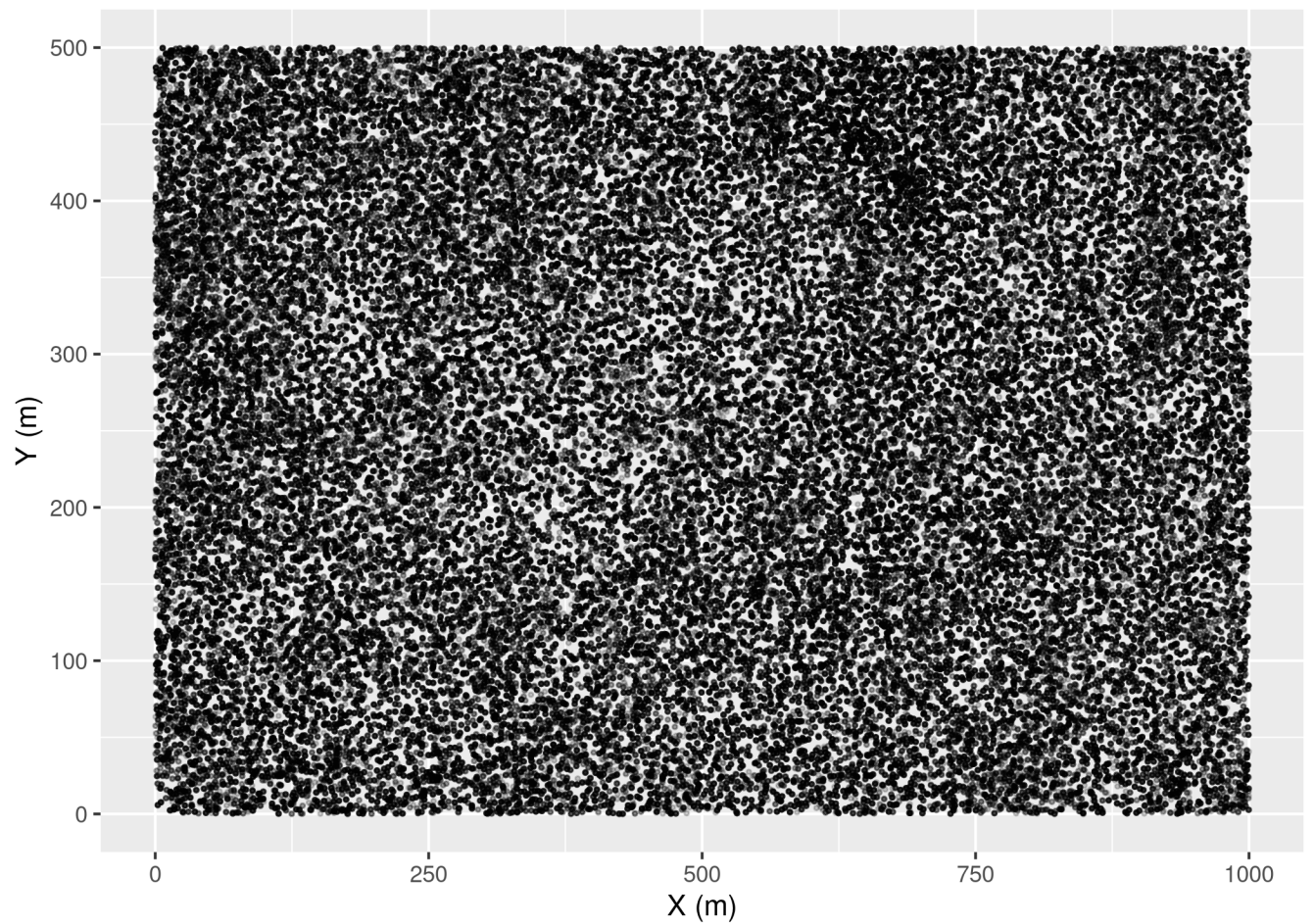2) Remove lines with missing DBH measures and trees with DBH < 10 cm for consistence with the other datasets.

Figure 2.3: The 50 ha plot of the Barro Colorado Island site. Each point is a tree.

3) Compute annualised growth.

4) Remove lines with a growth higher than 100 mm/y or inferior to -2 mm/y.

There are now 244 species and 30386 trees and 167618 observations.

5) Compute mean annual growth.

Finally, homogenise the three datasets.

3   Abundance diagrams

These diagrams can help to visualise how diverse the tropical forest is, but also that there are few dominant species and many rare species (long threshold).
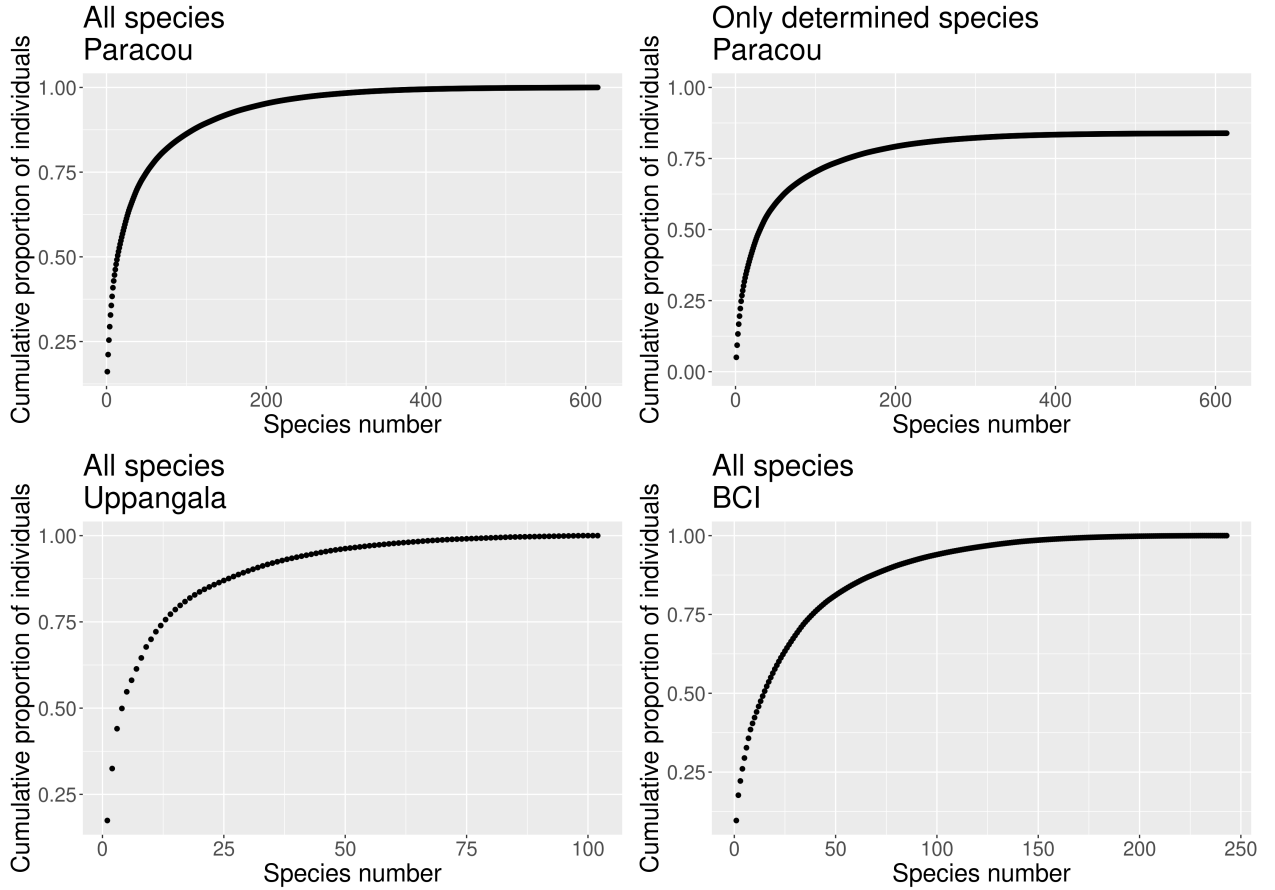

Figure 3.1: Abundance diagrams of each site

4   Estimation of IV

We estimate the importance of intraspecific variability compared to interspecific variability. To do so, we use Stan to run a Bayesian hierarchical model, with a species random effect and an individual random effect

on the intercept. In order to approach more classic growth models, and estimate an intraspecific variability within diameter classes, we also added an intercept and a diameter fixed effect. We use the brms package [7,8] in Rstudio to implement the model. We scale the data to improve convergence time.

$$ln(G_{ijt} + 2) = [\beta_0 + b_{0j} + d_{0i}] + \beta_1 \times ln(D_{ijt}) + \epsilon_{ijt},$$

Priors

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

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

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

$$d_{0j} \sim \mathcal{N}(mean = 0, var = V_d), iid$$

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

Hyperpriors

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

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

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

Where i is the individual and j the species. The model includes two fixed effects (β_0 and β_1), a random species effect b_{0j} on the intercept and a random individual effect d_{0i} on the intercept. V_d is the intraspecific variance and V_b is the interspecific variance.
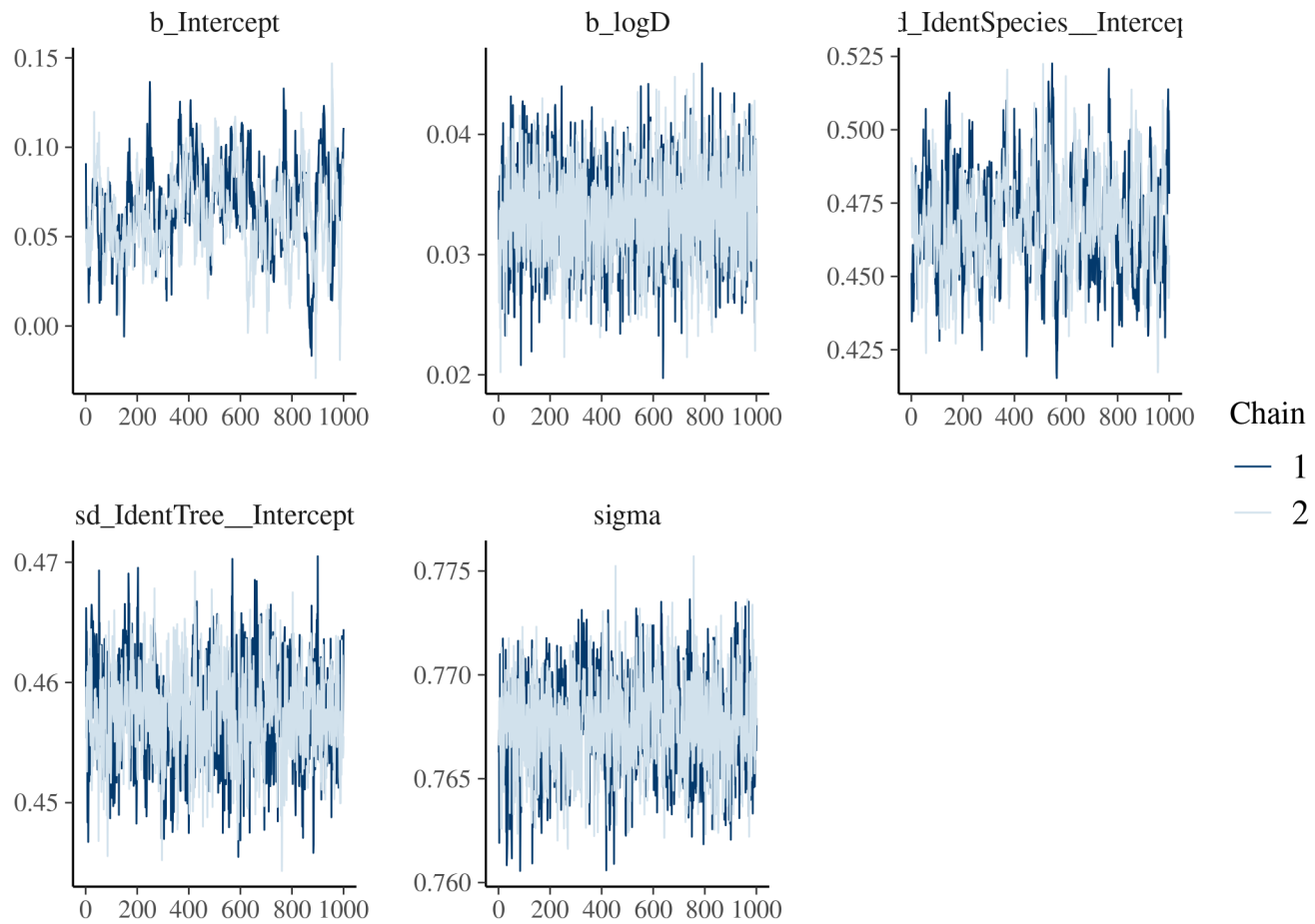
4.1 Paracou

```
## No divergences to plot.
```

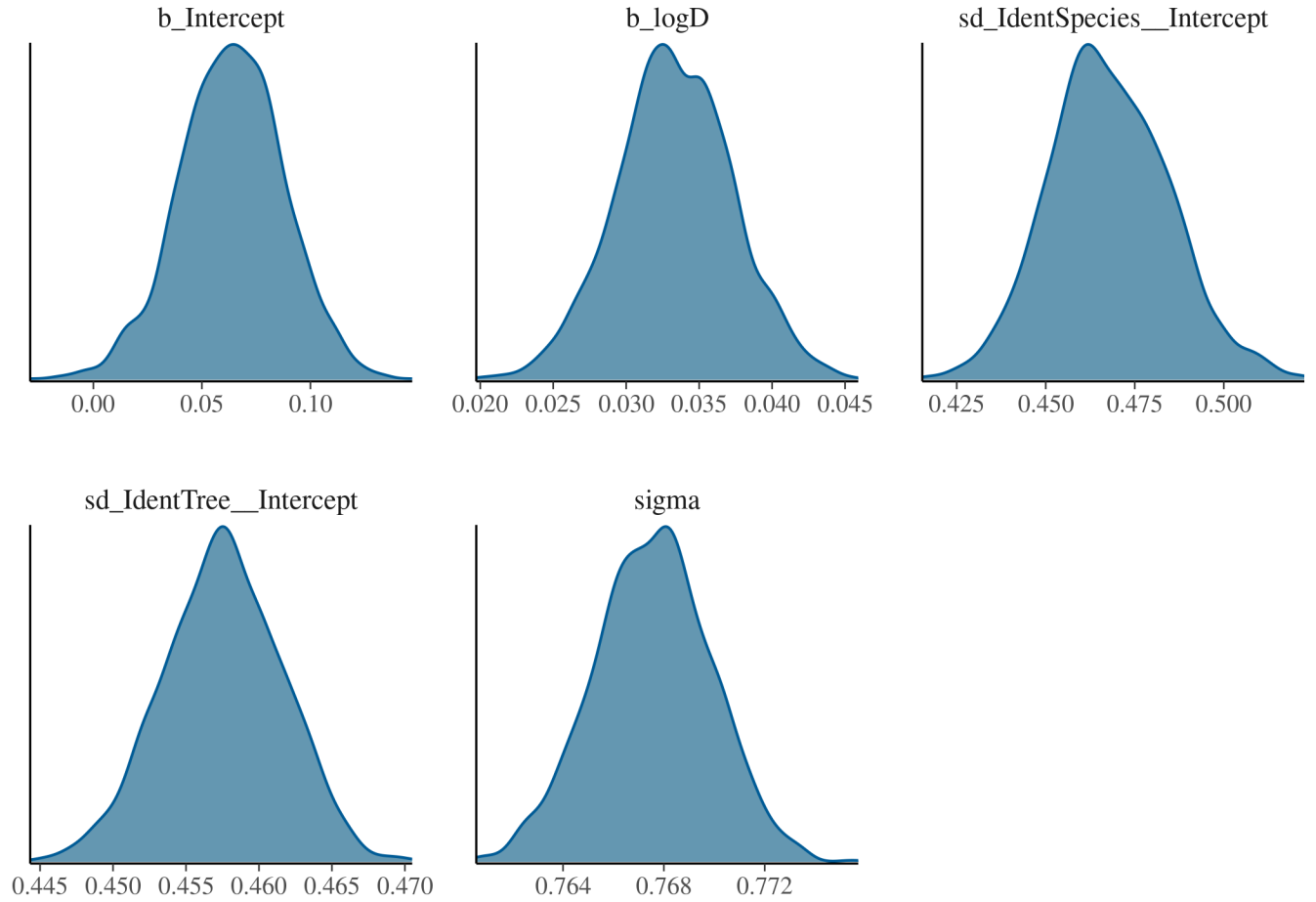Figure 4.1: Trace of the posterior estimates for Paracou

Figure 4.2: Density of the posterior estimates for Paracou

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: logG ~ 1 + logD + (1 | IdentSpecies) + (1 | IdentTree)
##    Data: Data_stan_Paracou (Number of observations: 100000)
## Samples: 2 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup samples = 2000
##
## Group-Level Effects:
## ~IdentSpecies (Number of levels: 575)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.47      0.02     0.44     0.50 1.01      255      471
##
## ~IdentTree (Number of levels: 48636)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.46      0.00     0.45     0.47 1.00      484      928
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept     0.06      0.02     0.01     0.11 1.02      116      199
## logD          0.03      0.00     0.03     0.04 1.00     2094     1634
```

```
## 
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.77      0.00     0.76     0.77 1.00      961     1340
## 
## 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).
```

Here the intraspecific variability is about the same as interspecific variability.

4.2 Uppangala

We apply the same model on the Uppangala dataset.

```
## No divergences to plot.
```
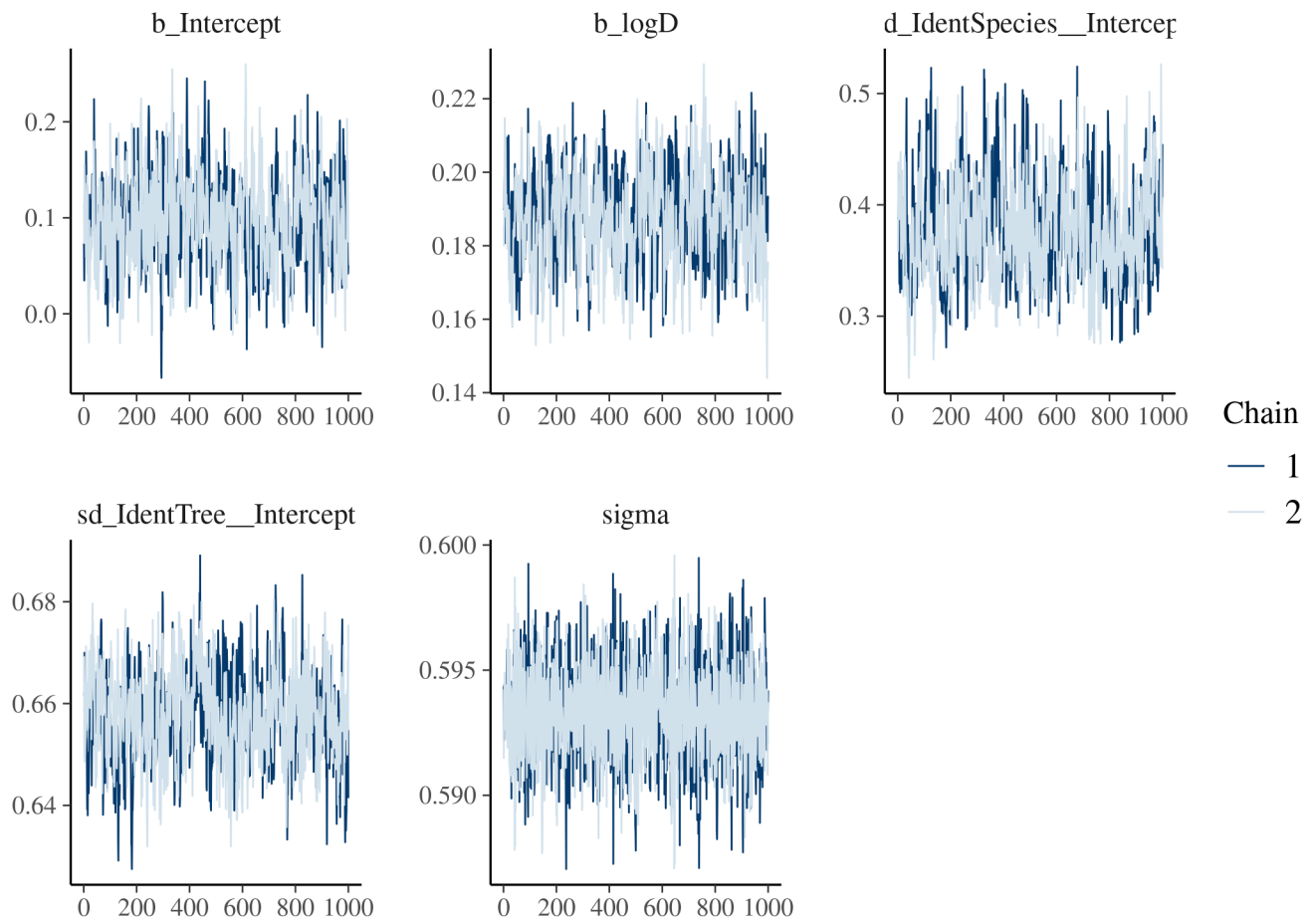

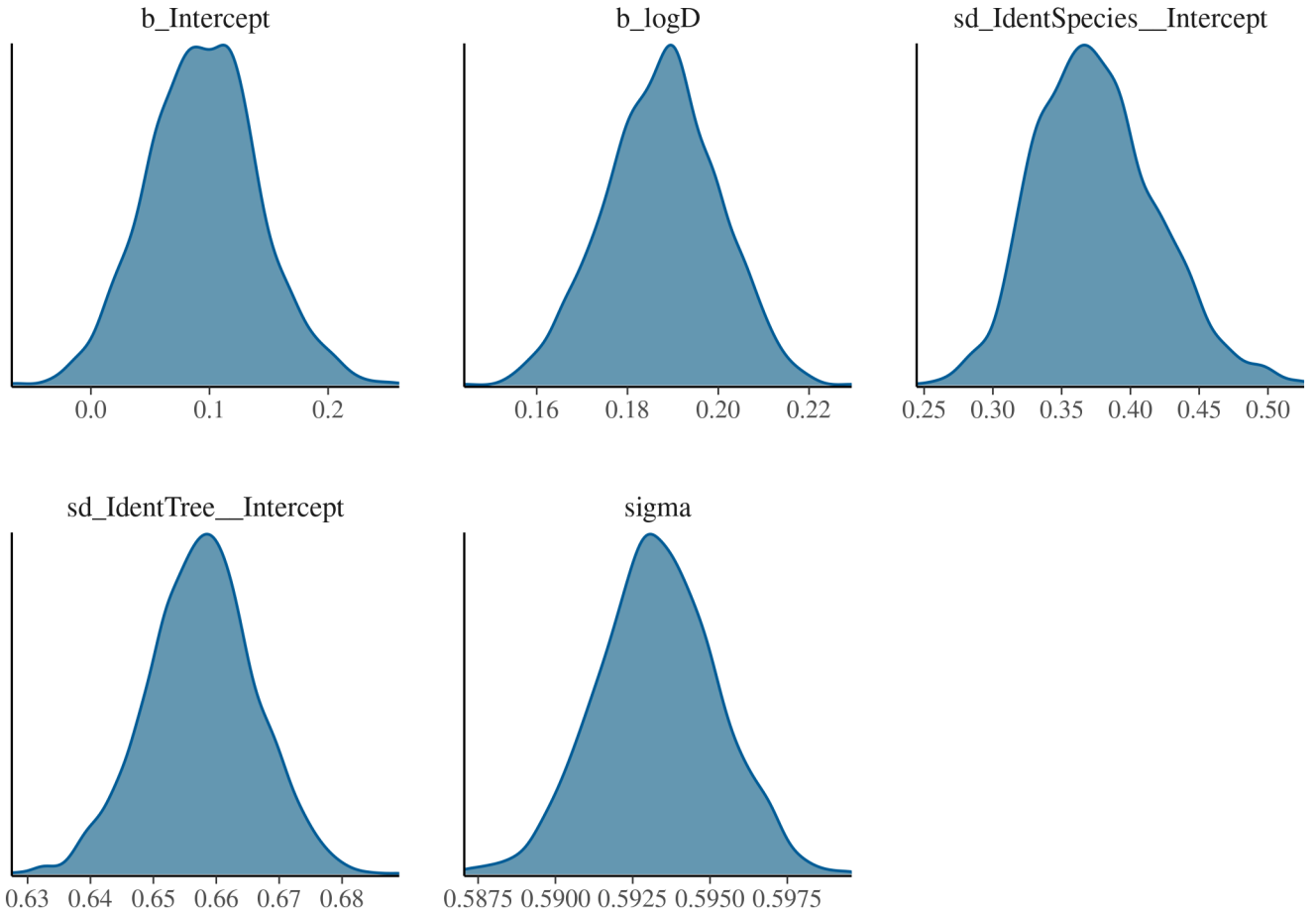Figure 4.3: Trace of the posterior estimates for Paracou

Figure 4.4: Density of the posterior estimates for Uppangala

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: logG ~ 1 + logD + (1 | IdentSpecies) + (1 | IdentTree)
##    Data: Data_stan_Uppangala (Number of observations: 52524)
## Samples: 2 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup samples = 2000
##
## Group-Level Effects:
## ~IdentSpecies (Number of levels: 99)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.38      0.04     0.30     0.47 1.00      294      487
##
## ~IdentTree (Number of levels: 3645)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.66      0.01     0.64     0.67 1.01      233      744
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept     0.09      0.05     0.01     0.19 1.00      474      825
## logD          0.19      0.01     0.16     0.21 1.01      341      687
```

```
## 
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.59      0.00     0.59     0.60 1.00     2620     1069
## 
## 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).
```

We can also conclude that intraspecific variability is high, but even that it is higher than interspecific variability.

4.3 BCI

We apply the same model to the BCI dataset.
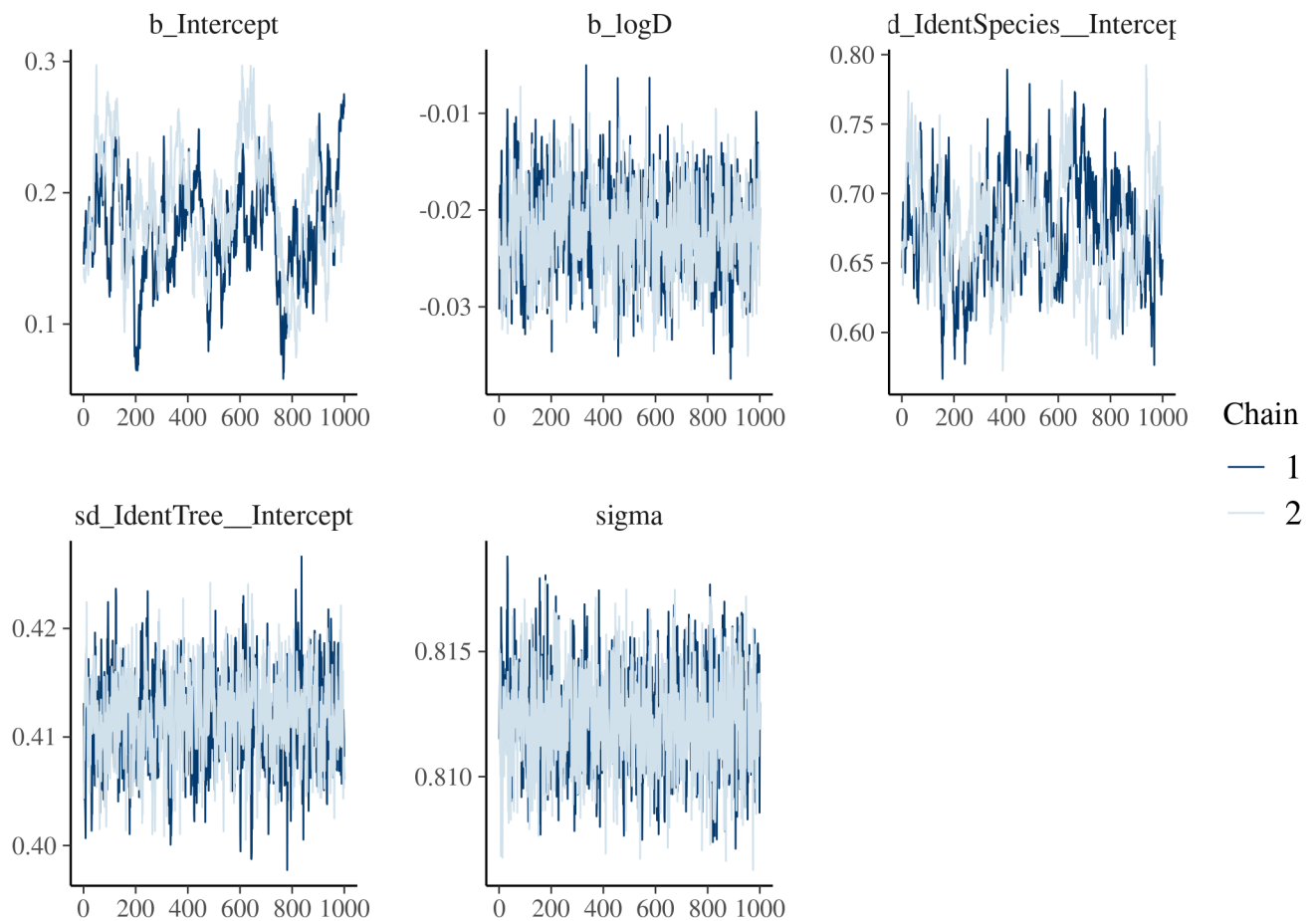
```
## No divergences to plot.
```


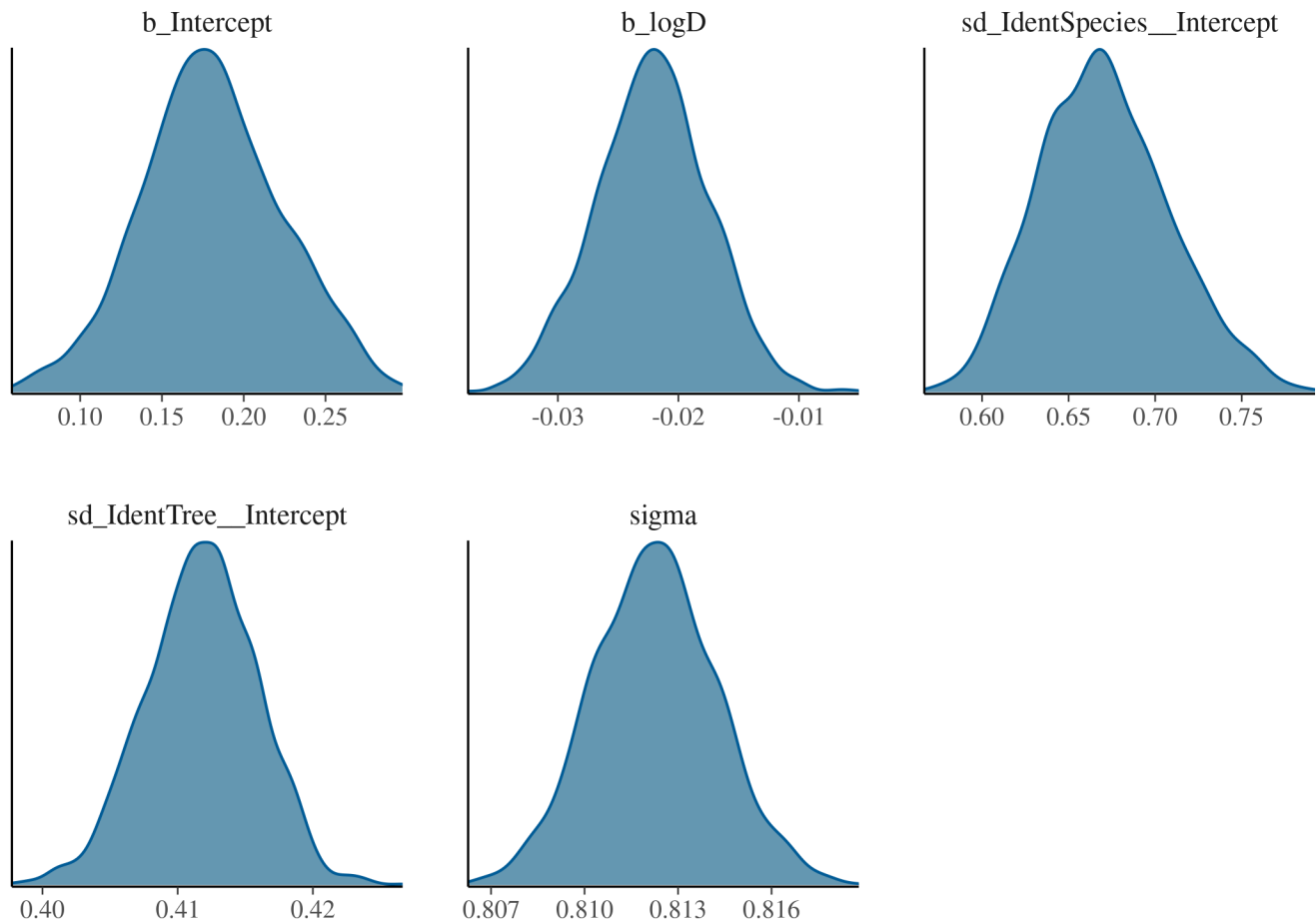Figure 4.5: Trace of the posterior estimates for BCI

14

Figure 4.6: Density of the posterior estimates for BCI

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: logG ~ 1 + logD + (1 | IdentSpecies) + (1 | IdentTree)
##    Data: Data_stan_BCI (Number of observations: 110710)
## Samples: 2 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup samples = 2000
##
## Group-Level Effects:
## ~IdentSpecies (Number of levels: 231)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.67      0.04     0.60     0.75 1.01       62      208
##
## ~IdentTree (Number of levels: 24320)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.41      0.00     0.40     0.42 1.00      565     1031
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept     0.18      0.04     0.10     0.27 1.03       44       80
## logD         -0.02      0.00    -0.03    -0.01 1.00      653     1276
```

```
## 
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.81      0.00     0.81     0.82 1.00     1446     1686
## 
## 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).
```

Here, intraspecific variability is slower than interspecific variability, although it is still about half the order of magnitude, which is significant. Note that here the mean estimated fixed effect of diameter on growth is slightly negative, which is counter-intuitive.

The following table enables to compare the results of the models for the three datasets.

Table 1: Mean estimates and their estimation errors for the three models

	Intercept (β_0)	Diameter (β_1)	Species variance (V_b)	Individual variance (V_d)	Residual variance (V)
Paracou					
Estimate	6.4e-02	3.3e-02	4.7e-01	4.6e-01	7.7e-01
Estimation error	2.4e-02	4e-03	1.7e-02	4.1e-03	2.3e-03
Uppangala					
Estimate	9.5e-02	1.9e-01	3.8e-01	6.6e-01	5.9e-01
Estimation error	4.7e-02	1.2e-02	4.4e-02	8.6e-03	1.9e-03
BCI					
Estimate	1.8e-01	-2.2e-02	6.7e-01	4.1e-01	8.1e-01
Estimation error	4.3e-02	4.6e-03	3.7e-02	4.1e-03	2e-03

We compute the proportion of unexplained variance of growth that is contained in the individual effects :
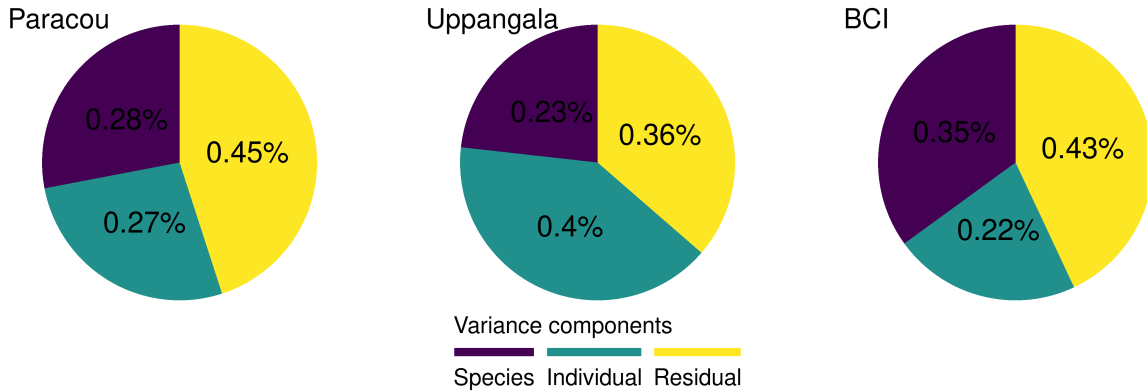

Figure 4.7: Proportions of each variance components for the three datasets. The variance of individual effects is in blue, the variance of species effect is in purple and the variance of residuals is yellow

Overall, a large part of variability is imputable to individual effects, showing a high intraspecific variability in growth in tropical forests, independently from the diameter.

5 Moran's I analysis

In order to have a numeric evaluation of spatial autocorrelation, we perform Moran's test on individual mean growth for each species. Moran's test function function was re-written from the ape package's Moran.I function [9] in order to be able to select the pairs of neighbours which are less than 100 m apart. The p-values are obtained with a "greater" alternative hypothesis.

As this analysis takes time, we only present the graphical result from a previously obtained result, but feel free to re-run it from the .Rmd file.

We compute the proportion of species with significant positive spatial autocorrelation with a 0.05 alpha risk and the proportion of species with non significant spatial autocorrelation. The proportions are computed relative to the species on which the test was computed.

Table 2: Proportions of species and individuals with a significant Moran's I test result

	Significant (i)	Not significant (ii)
Paracou		
Proportion of species	28.7	71.3
Proportion of individuals	77.7	22.3
Uppangala		
Proportion of species	18.5	81.5
Proportion of individuals	45.3	54.7
BCI		
Proportion of species	16.9	83.1
Proportion of individuals	51.4	48.6

We can observe that the proportion of species with significant spatial autocorrelation is low (29% in Paracou, 18% for Uppangala, 17% for BCI). However, the mean number of individuals per category shows that there is a much higher number of individuals in species with significant spatial autocorrelation. This motivates to analyse the proportion of individuals represented in each category.

We test if species abundance significantly explains the significance of Moran's test :

```
##
## Call:
## glm(formula = factor(signif) ~ nb_ind, family = binomial(link = "logit"),
##     data = na.omit(Moran_I_Paracou))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.0432  -0.5836  -0.5083   0.0470   2.0907
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.140573   0.196665 -10.884  < 2e-16 ***
## nb_ind       0.010620   0.001518   6.995 2.65e-12 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 455.47  on 379  degrees of freedom
## Residual deviance: 317.97  on 378  degrees of freedom
## AIC: 321.97
##
## Number of Fisher Scoring iterations: 7


##
## Call:
## glm(formula = factor(signif) ~ nb_ind, family = binomial(link = "logit"),
##     data = na.omit(Moran_I_Uppangala))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9313  -0.5575  -0.5338  -0.5228   2.0219
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.957224   0.439264  -4.456 8.36e-06 ***
## nb_ind       0.005754   0.002750   2.093   0.0364 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 51.75  on 53  degrees of freedom
## Residual deviance: 46.27  on 52  degrees of freedom
## AIC: 50.27
##
## Number of Fisher Scoring iterations: 4


##
## Call:
## glm(formula = factor(signif) ~ nb_ind, family = binomial(link = "logit"),
##     data = na.omit(Moran_I_BCI))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0333  -0.4755  -0.4223  -0.4016   2.2617
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.527817   0.317787  -7.954 1.80e-15 ***
## nb_ind       0.005645   0.001334   4.233 2.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 156.06  on 171  degrees of freedom
```

```
## Residual deviance: 122.57  on 170  degrees of freedom
## AIC: 126.57
##
## Number of Fisher Scoring iterations: 5
```

For all three datasets, the effect of the number of individuals on test significance is significant. However, it is less significant in Uppangala.

For Paracou, the proportion of individuals belonging to as species which had a significant spatial autocorrelation is high (> 77%). For Uppangala and BCI, it is lower (~45% and 51% respectively).

Important parameters contrasting between those forests are the topography, which is hillier in Uppangala, the number of species and of individuals per species, which are higher in Paracou, and the date of the last known disturbance, which is much more recent in Paracou than in BCI and Uppangala. Moreover, the same analysis of the undisturbed plots of Paracou give about the same proportion of individuals (exposed lower).

Finally, the magnitude of spatial autocorrelation varied across the three sites, with Paracou presenting the biggest proportion and Uppangala the lowest, which can be related to each site specificities. The plots of the Paracou site were subjected to disturbance treatments that created numerous large gaps [10]. This resulted in a high heterogeneity and spatial structuration of the environment. When restricting our analysis to undisturbed control plots, the proportion of detected spatial autocorrelation in conspecific growth drops from 77.1% to about 50% individuals, supporting the hypothesis that the more important spatial autocorrelation in Paracou is due to the stronger spatial structuration of the micro-environment associated with disturbance treatments. Conversely, the disturbance regime in BCI is known to produce frequent small gaps [11], while the mature forest in Uppangala has been reported as barely disturbed [3]. We acknowledge that the hilly topography of Uppangala could make our estimates of distances among individuals based on pairs of tree coordinates provided by field inventory less accurate than in the other two flatter sites.

6   Is variance within conspecifics inferior to variance within heterospecifics ?

In order to compare intra- and interspecific variance, we compute the semivariance of growth of individuals that are less than 100 m apart. We compare the semivariances obtained with conspecifics and heterospecifics with a Mann-Whitney test. Semivariance is estimated as the mean of the squared difference of the mean growth of all pairs individuals of the focal species which are less than 100 m apart. Interspecific variance is estimated with the semivariance computed as the mean of the squared difference of the mean growth of all pairs individuals of the containing the focal species and another species which are less than 100 m apart.

Here again, the analysis is long to run and we display results from a pre-generated result.

Computing interspecific variance with a sampling of maximum ten indidividuals per species, we obtain a theoretical interspecific variance, not taking the actual proportions of species into account.

We can see that in all forests, in about half of the species, local intraspecific growth semivariance is significantly smaller than local semivariance of heterospecifics growth.

In all forests, a large majority of individuals belong to species in which intraspecific growth semivariance is significantly smaller than interspecifc growth semivariance.

Sampling individuals within species to estimate interspecific variance did not change the main results.

7   Other tests and verifications

Table 3: Proportions of species and individuals with a significant semivariance comparison test result

	Intraspecific variability < interspecific variability (i)	Intraspecific variability ~ interspecific variability (ii)	Intraspecific variability > interspecific variability (iii)
Paracou			
Proportion of species	61.7	39.7	0.333
Proportion of individuals	89.1	10.9	0.0699
Uppangala			
Proportion of species	42.2	62.2	4.44
Proportion of individuals	57.7	23.6	18.8
BCI			
Proportion of species	46.6	44.3	9.09
Proportion of individuals	75.2	17.2	7.6

Table 4: Proportions of species and individuals with a significant semivariance comparison test result, using samples to estimate interspecific variance.

	Intraspecific variability < interspecific variability (i)	Intraspecific variability ~ interspecific variability (ii)	Intraspecific variability > interspecific variability (iii)
Paracou			
Proportion of species	61.7	37	1.33
Proportion of individuals	86.3	12.1	1.56
Uppangala			
Proportion of species	42.2	55.6	2.22
Proportion of individuals	61.2	20.4	18.4
BCI			
Proportion of species	50	47.7	2.27
Proportion of individuals	71.2	27.8	1.02

7.1 Undisturbed plots in Paracou

One of our main hypotheses is that many environmental variables are spatially structured. Consequently, as growth is largely influenced by environmental variables like light, it should also be spatially structured. Our analysis using Moran's I test showed that growth was indeed spatially structured. The Paracou dataset offers the opportunity to test this hypothesis further. Indeed, some plots were disturbed in the early eighties, creating artificial gaps, whereas others were not disturbed. As the creation of gaps results in a strong spatial structure of the available light under the canopy, growth should be less structured in plots that were not disturbed. That is what we test here.

This leads to a dataset containing 14266 trees of 451 species.

The proportion of individuals represented in the species which have a significant Moran's test are about 50%. With the disturbed plots, it was more than 70%. This is due to the openness of the canopy triggering important spatial structure in disturbed plots due to light gap. In these gaps, tree tend to grow faster, and

Table 5: Proportions of species and individuals of the Paracou site which had a significant or insignificant Moran's test results

	Significant (i)	Not significant (ii)
Proportion of species	18	82
Proportion of individuals	54.7	45.3

therefore the spatial structure of growth is stronger.

7.2   Smaller stems in BCI

In order to be able to compare all three datasets together, we removed all stems that had a DBH inferior to 10 cm in BCI, where all stems with DBH $\geq$ 1 cm are measured. Using the complete dataset, we concluded that the spatial structure in individual growth was even more significant.

7.3   Other discussion of the results

We here used data on tree growth, which is one component of individual fitness and an imperfect proxy of tree competitive ability, and our analyses could be further tested with other traits in the future.

8   Code implementation

The whole analysis is done using the R language [12] in the Rstudio environment [13]. The tables are made with the kableExtra package [14], the figures with the package ggplot2 [15], and the code uses other packages of the Tidyverse [16] (tidyr [17], dplyr [18], readr [19], lubridate [20], glue [21], magrittr [22]) and other R packages (here [23], sp [25]). Some functions were implemented in C++ thanks to the R packages Rcpp [26,27,28] and RcppArmadillo [29]. The pdf and html documents are produced thanks to the R packages rmarkdown [32], knitr [35] and bookdown [36].

References

1   Gourlet-Fleury, S. *et al.*, eds. (2004) *Ecology and management of a neotropical rainforest: Lessons drawn from Paracou, a long-term experimental research site in French Guiana*, Elsevier.

2   Hérault, B. and Piponiot, C. (2018) Key drivers of ecosystem recovery after disturbance in a neotropical forest: Long-term lessons from the Paracou experiment, French Guiana. *Forest Ecosystems* 5, 2

3   Pélissier, R. *et al.* (2011) Tree demography in an undisturbed Dipterocarp permanent sample plot at Uppangala, Western Ghats of India: *Ecological Archives* E092-115. *Ecology* 92, 1376–1376

4   Bec, J.L. *et al.* (2015) Characterizing tropical tree species growth strategies: Learning from interindividual variability and scale invariance. *PLoS ONE* 10,

5   Pascal, J.-P. and Pelissier, R. (1996) Structure and Floristic Composition of a Tropical Evergreen Forest in South-West India. *Journal of Tropical Ecology* 12, 191–214

6 Condit, R. *et al.* Complete data from the Barro Colorado 50-ha plot: 423617 trees, 35 years. (2019), Dryad

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

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

9 Paradis, E. and Schliep, K. (2019) Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528

10 Molino, J.-F. and Sabatier, D. (2001) Tree Diversity in Tropical Rain Forests: A Validation of the Intermediate Disturbance Hypothesis. *Science* 294, 1702–1704

11 Hubbell, S.P. (1999) Light-Gap Disturbances, Recruitment Limitation, and Tree Diversity in a Neotropical Forest. *Science* 283, 554–557

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

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

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

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

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

17 Wickham, H. (2021) *Tidyr: Tidy messy data*,

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

19 Wickham, H. and Hester, J. (2020) *Readr: Read rectangular text data*,

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

21 Hester, J. (2020) *Glue: Interpreted string literals*,

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

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

24 Pebesma, E.J. and Bivand, R.S. (2005) Classes and methods for spatial data in R. *R News* 5, 9–13

25 Bivand, R. *et al.* (2013) *Applied spatial data analysis with R*, Second edition.Springer.

26 Eddelbuettel, D. and François, R. (2011) Rcpp : Seamless R and C++ Integration. *Journal of Statistical Software* 40,

27   Eddelbuettel, D. (2013) *Seamless R and C++ integration with Rcpp*, Springer.

28   Eddelbuettel, D. and Balamuta, J.J. (2018) Extending *r* with C++: A Brief Introduction to Rcpp. *The American Statistician* 72, 28–36

29   Eddelbuettel, D. and Sanderson, C. (2014) RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics & Data Analysis* 71, 1054–1063

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

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

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

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

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

35   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

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