

Neither global nor consistent: a technical comment on the tree diversity-productivity analysis of Liang et al. (2016)

Carsten F. Dormann¹, Helge Schneider¹, and Jonas Gorges¹

¹Biometry & Environmental System Analysis, University of Freiburg, Germany
carsten.dormann@biom.uni-freiburg.de

January 18, 2019

Abstract

The publication of Liang et al. (2016, Science) seems to demonstrate very clearly that increasing tree species richness substantially increases forest productivity. To combine data from very different ecoregions, the authors constructed a relative measure of tree species richness. This relative richness however confounds plot-level tree species richness and the polar-tropical gradient of tree species richness. We re-analysed their original data, computing a regional measure of tree species richness and addressing several other issues in their analysis. We find that there is virtually no effect of relative tree species richness on productivity when computing species richness at the local scale. Also, different ecoregions have very different relationships between tree species richness and productivity. Thus, neither the “global” consistency nor the actual effect can be confirmed.

Observational studies of the species-richness-ecosystem functioning relationship are correlational and not causal. Observing a consistent correlative effect of tree species richness on forest productivity (TSP-P) across a wide range of biomes, however, would make this relationship a reliable principle on which to build recommendations for forestry and forest conservation. The publication of Liang *et al.* (2016) report on such a consistent, global TSP-P-relationship, based on forest inventory data from over 600,000 plots in 12 different ecoregions.

There are several logical and technical flaws in their analysis; correcting them also removes the reported global relationship and thus its role as principle for forest management. We identified the following problems with the original analysis: (1) The authors computed “relative tree species richness” as proportion of the maximum value. Thereby it represents a gradient from boreal to tropical plots, rather than in *local* species richness. When instead computing species richness relative to the maximum value in the region the effect of species richness on productivity is dramatically reduced. (2) Plots are overwhelmingly from temperate forest; indeed only some 2500 plots are from the tropics (equivalent to 0.4%), despite these forests representing around 30% of the world’s forest. Stratifying the plots accordingly weakens the TSR-P-relationship. (3) In the spatial regression model, distances between plots were computed without taking the spherical nature of earth into account. This had little effect on the slope estimate of the TSR-P-relationship. (4) The computational burden of the spatial model required subsampling the data to 500 data points. The authors did not correctly compute confidence intervals for this approach, wrongly interpreting subsampling as bootstrapping and additionally incorrectly computing bootstrap standard errors. A correct subsampling-based estimation led to approximate tripling of the reported confidence interval. (5) As noted earlier (Schulze *et al.*, 2018), some 4% of the plots had productivity values (far) beyond what is biologically plausible (Stape *et al.*, 2010). The likely reason is that small plots with large inventory errors in the productivity may lead to erratically high values. Not taking this into account in the analysis, e.g. by down-weighting plots with productivities above $30 \text{ m}^2\text{ha}^{-1}\text{y}^{-1}$ at least indicates an unreflected use of data. It also leads to an overestimation of the absolute productivity. Whether that is the main reason for productivity increases dramatically higher than any reported from experimental setups (Zhang *et al.*, 2012a; Vilà *et al.*, 2013; Huang *et al.*, 2018; Jactel *et al.*, 2018; Ammer, 2019) is unclear.

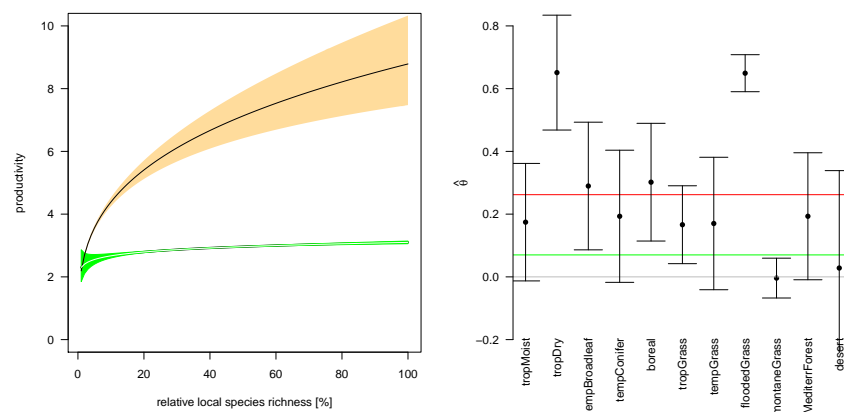


Figure 1: Comparison of reproduced analysis of Liang *et al.* (2016) (black) and corrected analysis of the same data (green), left, and re-analysis of the estimated species richness effect ($\hat{\theta}$) for each ecoregion separately, right. Red and green lines, right, are the original and the global re-analysis, respectively.

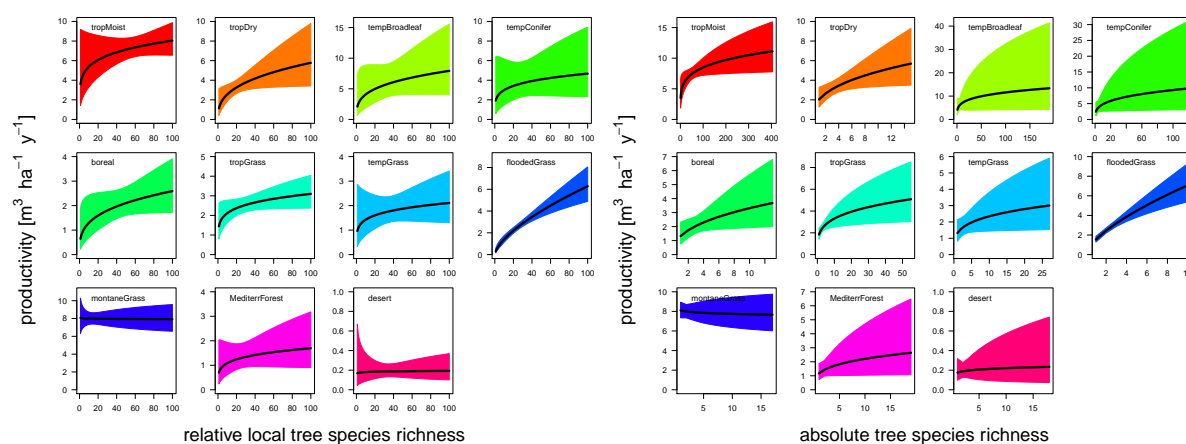


Figure 2: Biome-specific observed tree-species richness-productivity relationships (and 95% confidence interval). (Left) Species richness expressed as percentage of regional maximum. (Right) Productivity as changing with absolute species richness. Note that the shape of the curves can change when re-scaling species richness, as the values enter the analysis log-transformed.

Combining corrections for these points led to a new global analysis, with a negligible effect of tree species richness on a site's volumetric productivity (Fig. 1 left; see supplementary material for the full analysis including R-code). Furthermore, fitting the model to the 11 ecoregions with sufficient data separately shows that the global model is not consistent and provides a poor description of the ecoregions' specific TSR-P-relationships (Fig. 1 right). Also, there is no obvious benefit of scaling species richness, as the curves are more difficult to interpret when displayed as percentage of some maximal value (Fig. 2). Indeed, using a relative scale makes the species richness-effect (the θ -estimate) harder to compare across biomes and insignificant in several cases.

In summary, the analysis presented by Liang *et al.* (2016) is flawed in several respects, leading to a spuriously strong effect of tree species richness on forest productivity at the global level. While a re-analysis correlatively confirms a positive TSR-P-relationship for most ecoregion, these effects vary substantially in their strength and should be examined for each system separately. Overall, the original publication has thus added very little to our understanding, as both positive and biome-specific effects of species richness were already known before their publication (see, e.g., Whittaker & Heegaard, 2003; Vilà *et al.*, 2007; Paquette & Messier, 2010; Whittaker, 2010; Zhang *et al.*, 2012b), and the global consistency could not be upheld.

Appendix: Reproducible analysis with comments and conclusions for each step

Contents

A Introduction	3
B Reproducing the original tree richness-productivity relationship	4
B.1 Data preparation	4
B.2 Spatial model	8
C Corrections and improvements	11
C.1 Do implausible productivity values distort the analysis?	12
C.2 Aggregated subsets vs using the entire data set: do parameter estimates match? . .	13
C.2.1 Scaling standard errors from subsamples to full data	15
C.2.2 Increasing the size of the subsamples for the GLS	16
C.3 Stratified sampling of subsets	17
C.4 Compute distances between plots as Great Circle distances, rather than Euclidean (thereby accommodating the spherical nature of earth)	19
C.5 Define species richness relative to what is regionally possible	22
C.6 Use relative, rather than absolute, productivity	24
D A new analysis	26
E Comparison with ecoregion-specific fits	28
E.1 Analysis with actual species richness as predictor	31
F Conclusion	33
Bibliography	33

A Introduction

The positive effect of species richness on various measure of ecosystem functioning is generally undisputed (Cardinale *et al.*, 2012), even though this effect levels off fairly rapidly as the number of species beyond “few” (Cardinale *et al.*, 2006; Isbell *et al.*, 2011). Such fundamental understanding does not necessarily translate into applicable knowledge, and hence the publication of Liang *et al.* (2016) is a landmark step in converting academic biodiversity-ecosystem functioning into on-the-ground forest management. They claim to describe a globally applicable relationship between tree species richness and forest productivity, which, if correct, could guide forest management around the world without specific understanding of local or regional specifics.

On re-analysing the excellent data base compiled by Liang *et al.* (2016), we were originally curious to investigate an unanswered question, namely whether the global tree-richness-productivity-relationship is actually anywhere near as good as one built, on the same data, for a specific biome. Furthermore, we were also concerned with describing a partial dependence plot for the effect of tree species richness, rather than the conditional plot presented by Liang *et al.* (2016); the difference being that in their plot, the environmental predictors used in the model (such as basal area, temperature, precipitation and elevation) must be known, while in a partial dependency plot the *marginal* effect of tree species richness, independent of the actual local setting, would be depicted.

During the analysis, we realised that these initial questions were of secondary importance, as we discovered conceptual and statistical issues that cast substantial doubt on the results presented by Liang *et al.* (2016). Since the first author initially did not provide his code, we had to rely on the method description and guessing to reproduce the analysis.¹ From the reconstruction we modified the original model to correct for some mistakes, as detailed below, to derive our new version of the global relationship. In the next step, we predicted from the global model to the average conditions

¹This R-code of Liang is now available here: https://github.com/jingjingliang2018/GFB1/blob/master/R_script_Github10312008.R

of each biome, and compared this prediction with those of a biome-specific model from the same data base.

In the following, we present first the original analysis, then the modifications we introduce. We close with a new global and a biome-specific analysis.

B Reproducing the original tree richness-productivity relationship

B.1 Data preparation

The data are available under https://figshare.com/articles/GFB1_data_figshare_xlsx/4286552 for all countries except New Zealand, whose data are available from <https://datastore.landcareresearch.co.nz/dataset/new-zealand-forest-plot-data-in-global-forest-biodiversity-dataset>, both as Excel files. Upon downloading and merging the files, here is a summary of its content (in the following, code used for checking is outcommented):

```
library(readxl)
bpr_data_global <- read_excel("../Liang2016Science_GFB1_data_figshare.xlsx")
# dim(bpr_data_global) # check dimensions: 773100 x 21
# summary(bpr_data_global)
bpr_data_nz <- read_excel("../gfb1datanz.xlsx", sheet = 2)
# summary(bpr_data_nz) dim(bpr_data_nz) # check dimensions: 4026 x 21
# intersect(bpr_data_global$FID, bpr_data_nz$FID) # no overlap in FID
# intersect(bpr_data_global$PLOT, bpr_data_nz$PLOT) # no overlap in
# PLOT all.equal(colnames(bpr_data_global), colnames(bpr_data_nz)) #
# TRUE
bpr_data <- rbind(bpr_data_global, bpr_data_nz)
# dim(bpr_data) # check dimensions: 777126
summary(bpr_data)
```

FID	Country	dataset	PLOT
Min. : 0	Length:777126	Min. : 1.00	Min. : 0
1st Qu.:194281	Class :character	1st Qu.: 1.00	1st Qu.:193761
Median :388562	Mode :character	Median : 1.00	Median :388042
Mean :388562		Mean : 7.95	Mean :388043
3rd Qu.:582844		3rd Qu.: 8.00	3rd Qu.:582324
Max. :777125		Max. :45.00	Max. :776605
Lon	Lat	G	S
Min. : -170.80	Min. : -52.00	Min. : 0.00	Min. : 0.000
1st Qu.: -92.10	1st Qu.: 34.50	1st Qu.: 5.32	1st Qu.: 2.000
Median : -83.70	Median : 40.20	Median : 15.58	Median : 4.000
Mean : -64.31	Mean : 39.46	Mean : 19.00	Mean : 5.786
3rd Qu.: -70.30	3rd Qu.: 45.90	3rd Qu.: 27.46	3rd Qu.: 7.000
Max. : 178.46	Max. : 72.50	Max. :1239.63	Max. :405.000
P	I1	I2	T1
Min. : -358.360	Min. :0.0000	Min. :0.000	Min. : -165.0
1st Qu.: 1.830	1st Qu.:0.0000	1st Qu.:1.000	1st Qu.: 64.0
Median : 4.730	Median :0.0000	Median :1.000	Median : 106.0
Mean : 7.574	Mean :0.2315	Mean :1.426	Mean : 108.3
3rd Qu.: 8.780	3rd Qu.:0.0000	3rd Qu.:1.000	3rd Qu.: 156.0
Max. :1090.880	Max. :1.0000	Max. :4.000	Max. : 276.0
T2	T3	C1	C2
Min. : 0.00	Min. : 0	Min. : 0	Min. : 0.00
1st Qu.:30.00	1st Qu.: 6342	1st Qu.: 751	1st Qu.: 16.00
Median :36.00	Median : 7473	Median :1024	Median : 21.00
Mean :35.43	Mean : 7787	Mean :1020	Mean : 27.55
3rd Qu.:40.00	3rd Qu.: 9083	3rd Qu.:1251	3rd Qu.: 36.00
Max. :92.00	Max. :22147	Max. :5875	Max. :124.00
C3	E	PET	IAA
Min. : 0	Min. : -80.0	Min. : 0	Min. : 0
1st Qu.: 220	1st Qu.: 117.0	1st Qu.: 835	1st Qu.: 8388
Median : 291	Median : 288.0	Median :1016	Median : 9317
Mean : 282	Mean : 469.1	Mean :1063	Mean : 9915
3rd Qu.: 329	3rd Qu.: 515.0	3rd Qu.:1312	3rd Qu.:10661
Max. :1550	Max. :3876.0	Max. :1846	Max. :92589
Ecoregion			
Min. : 0.000			
1st Qu.: 4.000			
Median : 4.000			
Mean : 5.127			
3rd Qu.: 5.000			
Max. :98.000			

The meaning of the different column titles is provided in the meta-data to the figshare download. Most importantly, the response, productivity [$\text{m}^3\text{ha}^{-1}\text{y}^{-1}$] is given in column P, and the number of tree species observed in a plot is given in column S. As we can see from the summary, productivity ranges from -360 to $1090 \text{ m}^3\text{ha}^{-1}\text{y}^{-1}$, with most data between 1 and $9 \text{ m}^3\text{ha}^{-1}\text{y}^{-1}$. Species numbers range from 0 to 405 (we cannot provide a density here, as plot sizes are not provided and range from < 100 to $> 1100 \text{ m}^2$, only the 95% sampling interval was provided in the paper), with most plots between 2 and 7 species.

Liang *et al.* (2016) excluded plots with harvest volume $> 50\%$ of stocking volume (page 5), but do not provide the IDs of these plots. We thus excluded all plots with negative productivity and also those with productivity of 0 (as our model including them has a very different AIC to those reported by the authors).

Additionally, we remove plots with a species richness of 0, basal area of 0, those with a T3-value of 0 (temperature seasonality as standard deviation times 1000, making a value of 0 indicating no data, rather than no seasonality), IAA (indexed annual aridity times 10^{-4})-values of 0, and manually assigned four plots with an ecoregion value > 13 based on their coordinates in googleEarth. (Although there are 14 ecoregions in WWF's definition, the 14th in this data set was wrongly coded in the data and in fact plots were in rainforests.)

```
bpr_data <- bpr_data[bpr_data$P > 0, ] # Remove P values ≤ 0
bpr_data <- bpr_data[bpr_data$S != 0, ] # Delete species richness == 0
bpr_data <- bpr_data[bpr_data$G != 0, ] # remove plots with basal area == 0
bpr_data <- bpr_data[bpr_data$T3 != 0, ] # Delete rows where T3 == 0
bpr_data <- bpr_data[bpr_data$IAA != 0, ] # Delete rows where IAA and PET == 0
dim(bpr_data)
```

```
[1] 637710    21
```

```
# reassign ecoregions > 13 by their LonLat values LonLat = -48.7?,
# -26.2? Santa Catarina Rain forest, Brasil
bpr_data[bpr_data$Ecoregion == 14, ][1:3, ]$Ecoregion <- 1
# LonLat = -67.1?, 18.0? Llanos Lajas Dry forest, Puerto Rico
bpr_data[bpr_data$Ecoregion == 14, ]$Ecoregion <- 2
# LongLat = 30.5?, 1.3? Ituri forest, Rainforest
bpr_data[bpr_data$Ecoregion == 98, ]$Ecoregion <- 1
# key to ecoregions: see Liang et al. Fig. 1
rm(bpr_data_global, bpr_data_nz)
```

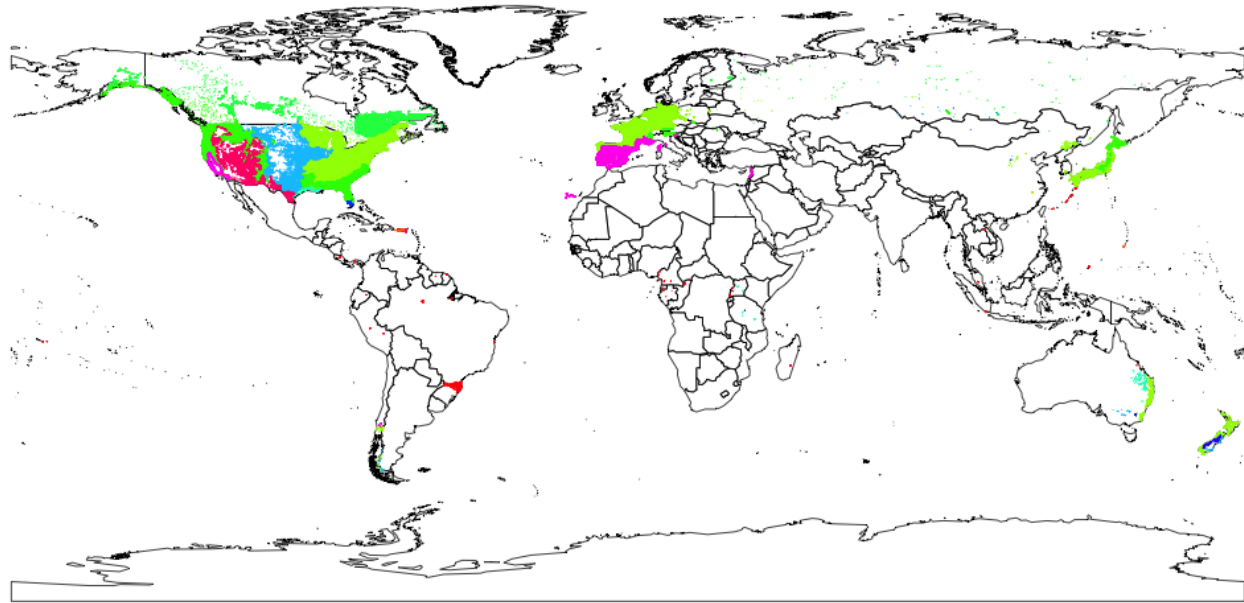
Some 1139 plots were recorded as ecoregion 0, which, when plotted, were distributed along the coasts across the world, i.e. belonging to very different ecoregions. We have no idea why these are classified as "0", but excluded them rather than manually and labouriously assigning them to appropriate ecoregions. Some plots were recorded repeatedly in the database (e.g. plot 0 occurs 370 times), but we kept them all in the data, as FIDs were unique, and so are their geographical coordinates.

```
library(maptools)
data(wrld_simpl)
# plot(wrld_simpl) # show where ecoregion 0 is
# with(bpr_data[bpr_data$Ecoregion==0,], points(Lon, Lat, col='green',
# pch=16, cex=0.5))
bpr_data <- bpr_data[bpr_data$Ecoregion != 0, ]
dim(bpr_data)
```

```
[1] 636616    21
```

```
# show all ecoregions:
library(maptools)
data(wrld_simpl)
par(mar = c(0, 0, 0, 0))
plot(wrld_simpl)
cols <- rainbow(13)
for (i in c(1, 2, 4:13)) {
  # there is no #3
  with(bpr_data[bpr_data$Ecoregion == i, ], points(Lon, Lat, col = cols[i],
    pch = 16, cex = 0.2))
}
table(bpr_data$Ecoregion)
```

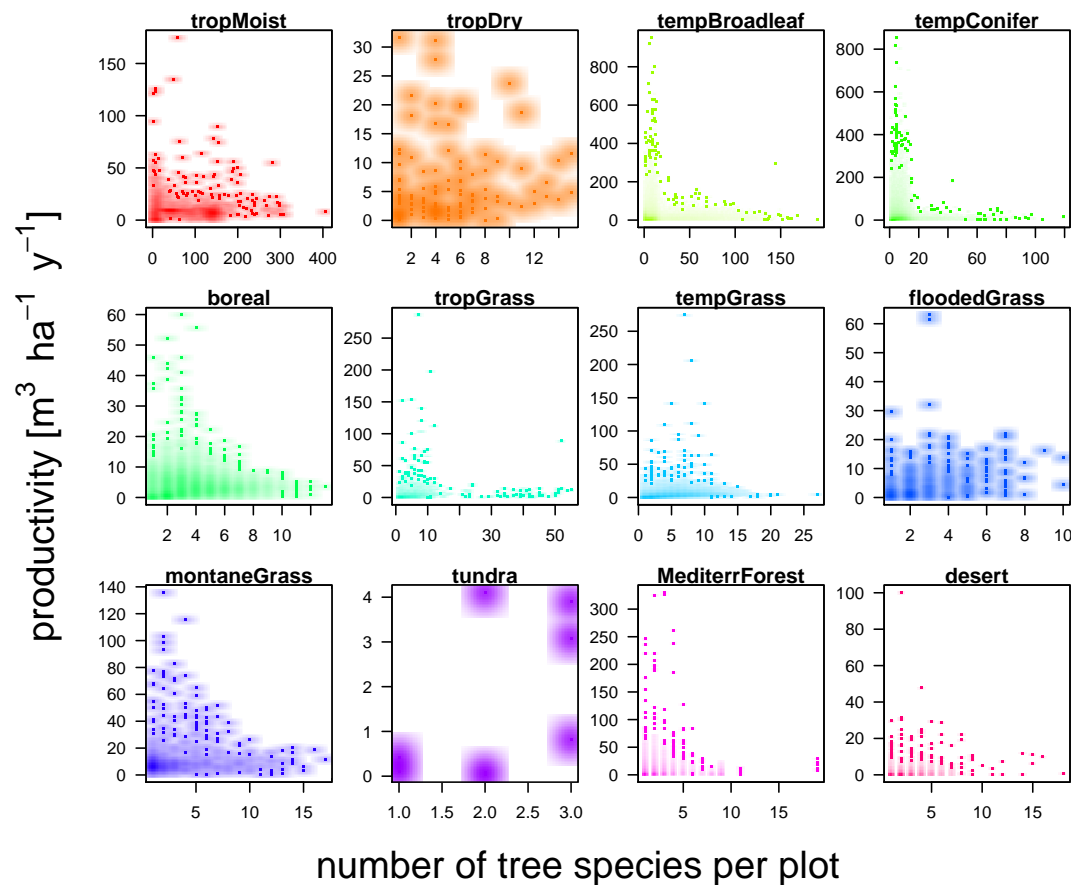
```
 1      2      4      5      6      7      8      9     10     11
2409   111 358551 177145 22651  2389 15146   547  1269      9
 12     13
39844 16545
```



At the end we are left with a data set of 636616 rows, heavily relying on the forest inventory data of temperate forests (ecoregions 4 and 5 comprise 615576 plots, i.e. 97% of all data points). Since for tundra, ecoregion 11, only 9 data points are available, we exclude it from the biome-specific analyses later on.

Before analysing these data, let's have a look at them, separately for each ecoregion:

```
# add information on tree proportions and biome names to data set:
biome_trees <- cbind.data.frame(c(1, 2, 4:13), c("tropMoist", "tropDry", "tempBroadleaf",
  "tempConifer", "boreal", "tropGrass", "tempGrass", "floodedGrass",
  "montaneGrass", "tundra", "MediterrForest", "desert"),
  c(799.4, 156.4, 362.6, 150.6, 749.3, 318, 148.3, 64.6, 60.3, 94.9,
    53.4, 53)) # in billions of trees, from Crowther et al. 2015
colnames(biome_trees) <- c("Ecoregion", "biomeName", "nOfTrees")
par(mfrow=c(3,4), mar=c(3,3,1,0), oma=c(4,4,0,0), las=1)
for (i in c(1,2,4:13)){
  with(bpr_data[bpr_data$Ecoregion == i, ], smoothScatter(P ~ S,
    colramp=colorRampPalette(c("white", cols[i])), pch=".", col=cols[i], xlab="",
    ylab="", main=biome_trees$biomeName[biome_trees$Ecoregion==i]))
}
mtext(side=1, text="number of tree species per plot", cex=1.5, line=1, outer=T)
mtext(side=2, text=expression(paste("productivity [", m^{3} ~ ha^{-1} ~ y^{-1}, "]")),
  cex=1.5, line=1, outer=T, las=0)
```

It is visible, at first glance, that these data are suspicious. The highest volume productivity ever measured in situ (rather than estimated from inventories), in a fertilised Eucalypt plantation in Brazil, was around $30 \text{ m}^3 \text{ ha}^{-1} \text{ y}^{-1}$ (Stape *et al.*, 2010). In this data set, the highest value is over $1000 \text{ m}^3 \text{ ha}^{-1} \text{ y}^{-1}$, and there are 23561 points with values larger than $30 \text{ m}^3 \text{ ha}^{-1} \text{ y}^{-1}$ (over 15000 of those are from the US, but Germany, France and Japan also contribute thousands of data points, respectively):

```
sum(bpr_data$P > 30)
```

```
[1] 23561
```

```
table(bpr_data$Country[which(bpr_data$P > 30)])
```

AU	BR	CA	CG	CL	CM	CN	CR	DE	EC	ES	FR
465	4	15	4	3	1	5	1	2803	1	678	2199
GA	GF	ID	JP	KR	LA	MG	NL	NZ	PE	PL	PT
1	1	1	1435	10	13	2	6	150	12	149	75
RO	SY	TZ	UG	US	US_AK	US_PR	US_VI				
2	2	1	4	15498	3	16	1				

```
table(bpr_data$Ecoregion[which(bpr_data$P > 10 & bpr_data$P < 30)])
```

1	2	4	5	6	7	8	9	10	12	13
667	19	91379	37742	555	174	992	76	329	2945	51

Moreover, as the last figure shows, some unlikely ecosystems contributed to high productivity sites (even after excluding obviously wrong data points): tundra, Mediterranean forest and deserts contributed dozens to thousands of plots with productivity exceeding $10 \text{ m}^3 \text{ ha}^{-1} \text{ y}^{-1}$. It is possible that the authors did remove some of these plots before the analysis without mentioning it explicitly in the methods (p. 5: “Intensively managed forests with harvests exceeding 50 percent of the stocking volume were excluded from this study”; there is no indication in the data, which plots are concerned).²

Nevertheless, as we are able to reproduce the analysis to a large extent (see below), we assume that many of these data points were still in the data set analysed.

²We have sent this document on 19 Sept 2018 to the first author and, in response, subsequently (on 31 Oct 2018) he made

B.2 Spatial model

Liang *et al.* (2016) employ a Generalised Least Squares (GLS) approach to accommodate spatial autocorrelation. Since the GLS is rather time-consuming, and it cannot handle data sets beyond a few thousand data points (depending on computer memory), they took random subsamples of 500 plots from the above data and ran the GLS on that subsample.³ For each such subsample, a new variable \hat{S} was computed as $\hat{S}_i = \frac{S_i}{\max(S_{\text{subsample}})}$. The actual model is at the log-log-scale (i.e. both P and \hat{S} were log-transformed). \hat{S} is interpreted as “relative species richness” and is claimed to “facilitate inter-biome comparison” (p. 7).⁴

So the actual steps of the analysis are:

1. Draw a random subsample of 500 data points (without replacement).
2. Compute \hat{S} for this sample.
3. Log-transform P and \hat{S} .
4. Run a GLS with log(P) as response and log(\hat{S}), G (basal area), T3 (standard deviation of temperature), C1 (annual precipitation), C3 (precipitation of warmest quarter), PET (potential evapotranspiration), IAA (indexed annual aridity) and E (elevation) as predictors.⁵ The spatial distances are implicitly computed from geographic coordinates using the corSpher correlation structure.⁶ Since coordinates of the released data were rounded, we jitter them slightly to avoid distances of zero.
5. Store model estimates and repeat.

While Liang *et al.* (2016) run 10000 bootstraps, we only do 50, as the results do not change substantially anymore after that. (Note that our code is not made for speed but for intellegibility.) Also, as the authors kept the nugget and range fixed in their final run, we use these values here (0.8 and 50, respectively).

As illustration, here is a single model run:

```
library(nlme) # for gls
bpr_data$lat <- jitter(bpr_data$Lat)
bpr_data$lon <- jitter(bpr_data$Lon)
set.seed(12)
index <- sample(nrow(bpr_data), 500)
subset <- bpr_data[index, ]
subset$Sbreve <- subset$S/max(subset$S)
fgls <- gls(log(P) ~ log(Sbreve) + G + T3 + C1 + C3 + PET + IAA + E, correlation = corSpher(c(50,
  0.8), form = ~lon + lat, nugget = T, fixed = T), data = subset)
summary(fgls)
```

```
Generalized least squares fit by REML
Model: log(P) ~ log(Sbreve) + G + T3 + C1 + C3 + PET + IAA + E
Data: subset
      AIC      BIC    logLik
1639.374 1681.338 -809.6869

Correlation Structure: Spherical spatial correlation
Formula: ~lon + lat
Parameter estimate(s):
range nugget
50.0   0.8
```

R-code of (the main part of) their analysis available under https://github.com/jingjingliang2018/GFB1/blob/master/R_script_Github10312008.R. In that code it is clear that plots with productivity larger than 533 m³ha⁻¹yr⁻¹ and with species richness larger than 270 species were removed. The reasoning behind these values is not communicated in that document. Going through their code, we are pleased to notice that we have correctly interpreted their methods section and that our points of content are not mere misunderstandings.

³Liang *et al.* (2016) invent the name “spatial random forest” for this procedure, which is rather misleading. First, the randomForest algorithm (Breiman, 2001) uses classification and regression trees as building blocks, hence the name “forest” as a large group of “trees”. Second, in randomForest the actual model is altered in each tree, by randomly selecting the variables trialed at each node. Third, while randomForest samples data points, it does so *with replacement*, i.e. as a bootstrap. As a result, the number of data points are the same as the full data set in each tree. What the subsampling approach of Liang *et al.* (2016) and the randomForest have in common is the use of aggregating many models. Since these are not actual bootstraps (because Liang *et al.* subsample), this is not proper “bagging” either. Thus, the term “spatial random forest” is misleading and suggests a sampling theory which actually does not apply to this approach. Accordingly, we shun this term, and also will not use the term “bootstrapping” unless sampling is carried out *with replacement*.

⁴We will return to the computation of relative species richness as a critical flaw later.

⁵Sources for each are given in the metadata to the New Zealand data.

⁶We will return to the misinterpretation of this correlation as being computed on a sphere later.


```

Coefficients:
              Value Std.Error   t-value p-value
(Intercept)  4.578043  1.1995697   3.816405  0.0002
log(Sbreve)  0.322772  0.0827239   3.901805  0.0001
G            0.023619  0.0036675   6.439911  0.0000
T3          -0.000119  0.0000646  -1.836109  0.0669
C1           0.000918  0.0004701   1.953651  0.0513
C3           0.001763  0.0008737   2.017603  0.0442
PET          -0.002111  0.0006586  -3.205601  0.0014
IAA          -0.000094  0.0000333  -2.832398  0.0048
E            -0.000617  0.0001474  -4.188606  0.0000

Correlation:
              (Intr) lg(Sb) G      T3      C1      C3      PET      IAA
log(Sbreve)  0.357
G            -0.175 -0.211
T3           -0.790 -0.130  0.106
C1           -0.079 -0.284 -0.039  0.264
C3           0.089  0.093 -0.054 -0.188 -0.169
PET          -0.751  0.001  0.083  0.425 -0.419 -0.123
IAA          -0.358  0.137  0.011  0.128 -0.793 -0.160  0.688
E            -0.378  0.001 -0.067  0.229  0.101 -0.058  0.332  0.080

Standardized residuals:
              Min      Q1      Med      Q3      Max
-4.814311 -0.181500  0.354186  0.902762  3.180726

Residual standard error: 1.19801
Degrees of freedom: 500 total; 491 residual

```

```
cor(predict(fgls), log(subset$P))^2 # R2
```

```
[1] 0.3696743
```

The part after setting the seed will be repeated below, and from the summary we only keep the coefficient estimates.

```

repeatedGLS <- function(N=500){ # function running one random subset GLS
  index <- sample(nrow(bpr_data), N)
  subset <- bpr_data[index, ]
  subset$Sbreve <- subset$S / max(subset$S)
  fgls <- gls(log(P) ~ log(Sbreve) + G + T3 + C1 + C3 + PET + IAA + E,
    correlation=corSpher(c(50, 0.8), form=~lon+lat, nugget=T, fixed=T),
    data=subset)
  r2 <- cor(predict(fgls), log(subset$P))^2 # R2
  return( c(ell=logLik(fgls), AIC=AIC(fgls), BIC=BIC(fgls), R2=r2, coef(fgls)) )
}
set.seed(123)
coefMat <- t(replicate(50, repeatedGLS() ))
library(matrixStats) # for colSds
signif(rbind(mean= colMeans(coefMat), se=colSds(coefMat) ) , 3)

```

```

      ell    AIC    BIC    R2 (Intercept) log(Sbreve)      G      T3
mean -817.0 1650.0 1700.0 0.3470         4.760        0.3020 0.01740 -6.70e-05
se    20.3   40.5   40.5 0.0418         0.981        0.0921 0.00426  4.73e-05
      C1      C3      PET      IAA      E
mean 0.001650 0.001100 -0.002710 -1.27e-04 -0.000618
se   0.000663 0.000812  0.000636  5.21e-05  0.000144

```

```

# for comparison, there the results from 5000 replicates:
if ("coefMat5000.Rdata" %in% dir()){
  load("coefMat5000.Rdata")
} else {
  coefMat5000 <- t(replicate(5000, repeatedGLS() ))
  save(coefMat5000, file="coefMat5000.Rdata")
}
signif(rbind(mean= colMeans(coefMat5000), se=colSds(coefMat5000) ) , 3)

```

```

      ell    AIC    BIC    R2 (Intercept) log(Sbreve)      G      T3
mean -817.0 1650.0 1700.0 0.3450         4.78        0.3010 0.01710 -7.50e-05
se    21.7   43.3   43.3 0.0482         1.07        0.0907 0.00426  5.23e-05
      C1      C3      PET      IAA      E
mean 0.001670 0.001120 -0.002710 -1.30e-04 -0.000602
se   0.000652 0.000761  0.000703  4.91e-05  0.000139

```

results are similar enough to allow us to work with 50 repetitions of the subsampling

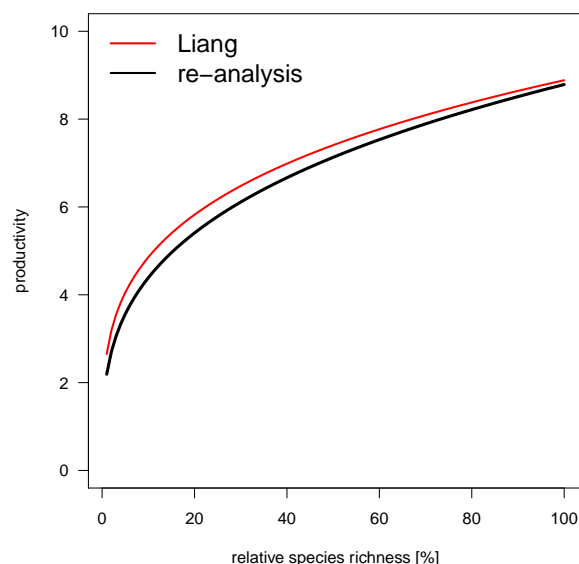
This approximates the results of Liang *et al.* (2016)'s Table 2, with the exception of the standard errors for the model coefficients, which Liang computed as incorrectly as standard deviation divided by the square root of the number of replicate models: correctly computed, the standard deviation of a bootstrap is the standard error of the statistic (here the goodness of fit measures and the coefficients). Hence, the row "SE" in the authors' table 2 must be multiplied by 100 to get to the correct values.⁷

Table 2. Parameters of the global geospatial random forest model in 10,000 iterations of 500 randomly selected (with replacement) GFB plots. Mean and SE of all the parameters were estimated by using bootstrapping. Effect sizes were represented by the Akaike information criterion (AIC), Bayesian information criterion (BIC), and generalized R^2 ($G-R^2$). Const. constant.

	Loglik	AIC	BIC	$G-R^2$	const	θ	G	T3	C1	C3	PET	IAA	E
Mean	-761.41	1546.71	1597.08	0.354	3.816	0.2625243	0.014607	-0.000106	0.001604	0.001739	-0.002566	-0.000134	-0.000809
SE	0.54	1.10	1.13	0.001	0.011	0.0009512	0.000039	0.000001	0.000008	0.000008	0.000009	0.000001	0.000002

Finally, we plot this relationship with the "correct" error bars. As parameter estimates are correlated, we cannot simply use the above standard errors to compute an error interval around the line. Instead, we predict with each of the 50 parameter sets the curve and compute 95% quantiles from them as confidence intervals.⁸

```
beta <- coefMat[, 5:13]
# set all values but log(Sbreve) to their mean:
X <- as.matrix(data.frame(Intercept = 1, Sbreve = log(seq(0.01, 1, by = 0.01)),
  t(colMeans(bpr_data[, c("G", "T3", "C1", "C3", "PET", "IAA", "E"))))))
preds <- X %*% t(beta)
par(mar = c(5, 4, 1, 1))
plot(1:100, exp(rowMeans(preds)), type = "l", las = 1, lwd = 3, ylim = c(0,
  10), xlab = "relative species richness [%]", ylab = "productivity")
betaLiang <- c(3.816, 0.2625, 0.0146, -0.00011, 0.0016, 0.00174, -0.002566,
  -0.000134, -0.000809)
Xliang <- X
Xliang[, 2] <- log(seq(0.01, 1, by = 0.01) * 100) # uses %, not fraction
liangLine <- Xliang %*% betaLiang
lines(1:100, exp(liangLine), col = "red", lwd = 2)
legend("topleft", legend = c("Liang", "re-analysis"), col = c("red", "black"),
  lwd = 2, lty = 1, bty = "n", cex = 1.5)
```



We were thus able to reproduce the analysis of Liang *et al.* relatively faithfully, at least with respect to the expectation (we shall return to the error envelop in due course).

⁷We shall return to the computation of error bars in a few sections. For now, please note that the classical bootstrap's idea is to use the *entire* data set when bootstrapping; only then will the bootstrapped standard errors be correct. Using the bootstrap on subsamples, in the way done by Liang *et al.* (2016), does not constitute a valid statistical approach for computing standard errors for the full data set, neither in their way, nor in the way presented here (with the multiplication by 100). What this subsampling-bootstrapped standard errors deliver is an estimate of where the line would be with 500 data points, not for the full data set.

There are subsampling procedures (see Politis *et al.*, 1999), but they require additional corrections on top of the subsampling itself.

⁸We do not know how Liang computed these, as he did not reply to our emails on this topic.

Since their log-likelihood is 50 units higher, we suspect that in their analysis further plots were excluded.

C Corrections and improvements

In summary, we have been able to reproduce the analysis of Liang *et al.* (2016).

During this first step, we may already have noticed some issues with the analysis. Here is our list of things we improve on in the next sections, starting with the most important ones first:

1. Evaluate the effect of unreasonably large values of productivity on the result.
2. Aggregating analyses of subsets, which comprise only a permille of the full data set, may introduce a bias relative to the full dataset. The mechanism behind this bias is the same that we criticised as first point, i.e. the computation of \hat{S} on the basis of highest tree species richness in the subset, thereby ranging \hat{S} differently for each subset.
3. Minor point: The GLS can easily accommodate more than 500 points, so we can run sub-samples of 1000, 2000 and 5000 points. This will slightly increase the chance of plots being relatively close to each other, and hence allows for a better estimation of the short-range spatial autocorrelation, possibly improving the estimates for nugget and range of the variogram.
4. Stratified draws of plots in proportion to the area this ecoregion's forests cover on the world. Currently only some 2500 points (out of 636000) are representing the tropical forests (labelled 1 and 2 below), despite their proportion of the world's forests being over 30%.

```
tapply(bpr_data$FID, bpr_data$Ecoregion, length)
```

1	2	4	5	6	7	8	9	10	11
2409	111	358551	177145	22651	2389	15146	547	1269	9
12	13								
39844	16545								

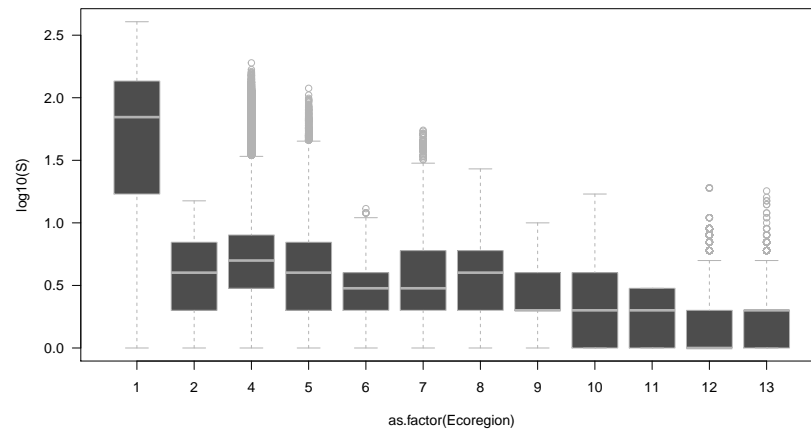
```
t(biome_trees[, 1:2]) # labels
```

	[,1]	[,2]	[,3]	[,4]	[,5]
Ecoregion	" 1"	" 2"	" 4"	" 5"	" 6"
biomeName	"tropMoist"	"tropDry"	"tempBroadleaf"	"tempConifer"	"boreal"
	[,6]	[,7]	[,8]	[,9]	[,10]
Ecoregion	" 7"	" 8"	" 9"	"10"	"11"
biomeName	"tropGrass"	"tempGrass"	"floodedGrass"	"montaneGrass"	"tundra"
	[,11]	[,12]			
Ecoregion	"12"	"13"			
biomeName	"MediterrForest"	"desert"			

5. The corSpher-structure used by Liang *et al.* (2016) suggests, in its name, that it would be appropriate for data on a sphere. That is not the case. The spherical model is only the name for a different shape of how spatial autocorrelation decreases with distance (Pinheiro & Bates, 2000), but it does not correct for the fact that on a sphere geographic distances cannot be computed from coordinates as Euclidean distance. To do so, we have to compute the so-called great-circle (or orthodromic) distance.
6. Calculate relative species richness \hat{S} relative to what is the maximal species richness *in that region*. Currently, the definition of relative species richness is not intuitive. Since only the tropics have plots with dozens to hundreds of species, \hat{S} is effectively representing a gradient from cold to tropical plots.⁹

```
plot(log10(S) ~ as.factor(Ecoregion), data = bpr_data, border = "grey70",
     col = "grey30", las = 1)
```

⁹This is crucial to understand, thus we try to repeat with other words. If we draw 500 plots randomly, with 97% of them coming from temperate forests, then the highest species richness will be set by the very few plots from the tropics, more precisely ecoregion 1: tropical moist broadleaf forest. Thus, we divide the species richness of our temperate plots, say something between 1 and 10, by 80-400. Thus, low \hat{S} -values will inevitably be from the temperate or boreal forests, while high \hat{S} -values will be from the tropics.



Thus, the so-called species richness gradient is in fact a latitudinal gradient. This is, in our opinion, a serious problem and leads to fundamental misinterpretation of the results. We propose to compute \hat{S} differently, namely relative to what is possible for a given plot. We could choose the maximal value for an ecoregion to scale tree species richness of each plot, or even more locally, the maximal species richness observed in, say, 100 km radius around a plot. This would lead to plots from any ecoregion being able to occupy the upper end of the relative species richness gradient, if they are species rich *relative to their wider neighbourhood*.¹⁰

In the following, we demonstrate the effect of each of these improvements separately, always comparing it to our analysis of the data presented above. Arguably, as the data set changed, we may have to re-estimate the coefficients of the spatial autocorrelation, which we did not.

C.1 Do implausible productivity values distort the analysis?

As shown above, some 23,000 data points, around 4% of the entire data set, have productivity values above the largest measured in the field. The probably reason is that forest inventories measure with a certain error, and if the error is large relative to the increase in biomass, it is easy to get unrealistically high productivity estimates, particularly in small plots with harvesting.

There are, essentially, two ways to accommodate values that are clearly implausible: (i) omit them, or (ii) give the lower weight in the model. We only investigate the effect of omitting all values where $P > 30$ as the strongest, not necessarily most appropriate, way.

```
bpr_data_less30 <- bpr_data[bpr_data$P <= 30, ]
repeatedGLSless30 <- function(N = 500) {
  # function running one random subset GLS
  index <- sample(nrow(bpr_data_less30), N)
  subset <- bpr_data_less30[index, ]
  subset$Sbreve <- subset$S/max(subset$S)
  fgls <- gls(log(P) ~ log(Sbreve) + G + T3 + C1 + C3 + PET + IAA + E,
    correlation = corSpher(c(50, 0.8), form = ~lon + lat, nugget = T,
    fixed = T), data = subset)
  r2 <- cor(predict(fgls), log(subset$P))^2 # R2
  return(c(ell = logLik(fgls), AIC = AIC(fgls), BIC = BIC(fgls), R2 = r2,
    coef(fgls)))
}
set.seed(123)
coefMatless30 <- t(replicate(50, repeatedGLSless30()))
library(matrixStats) # for colSds
signif(rbind(mean = colMeans(coefMatless30), se = colSds(coefMatless30)),
  3)
```

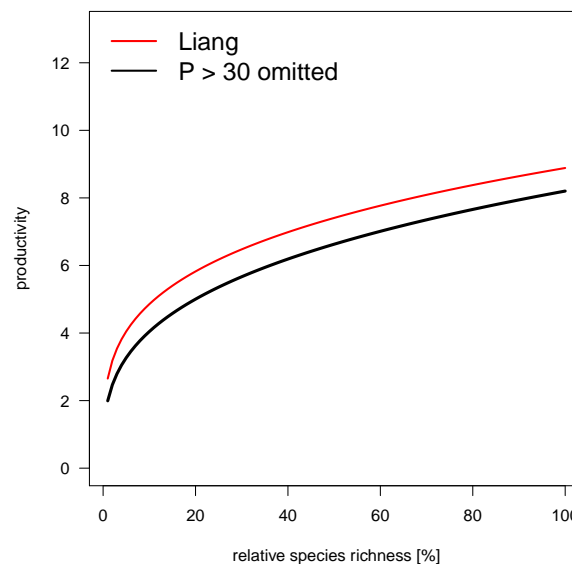
	ell	AIC	BIC	R2	(Intercept)	log(Sbreve)	G	T3
mean	-790.0	1600	1640	0.375	4.76	0.307	0.01830	-6.97e-05
se	22.5	45	45	0.047	1.07	0.088	0.00349	4.75e-05
	C1	C3	PET	IAA	E			
mean	0.001690	0.001160	-0.002770	-1.34e-04	-0.000619			
se	0.000581	0.000628	0.000694	4.41e-05	0.000135			

¹⁰Clearly, this is still not an ideal relative species richness, as we can imagine local conditions to vary at relatively small scale. However, we think that this step is already a major improvement in the actual meaning and interpretability of “relative species richness” and will come closer to what we think most of the authors actually understood \hat{S} to be.

```

betaless30 <- coefMatless30[, 5:13]
predsless30 <- X %>% t(betaless30)
par(mar = c(5, 4, 1, 1))
plot(1:100, exp(rowMeans(predsless30)), type = "l", las = 1, lwd = 3, ylim = c(0,
  13), xlab = "relative species richness [%]", ylab = "productivity")
betaLiang <- c(3.816, 0.2625, 0.0146, -0.00011, 0.0016, 0.00174, -0.002566,
  -0.000134, -0.000809)
Xliang <- X
Xliang[, 2] <- log(seq(0.01, 1, by = 0.01) * 100) # uses %, not fraction
liangLine <- Xliang %>% betaLiang
lines(1:100, exp(liangLine), col = "red", lwd = 2)
legend("topleft", legend = c("Liang", "P > 30 omitted"), col = c("red",
  "black"), lwd = 2, lty = 1, bty = "n", cex = 1.5)

```



Excluding the high-value plots leads to the expected drop in the absolute values of the TSR-P-relationship, but it does not substantially affect the slope estimate.

C.2 Aggregated subsets vs using the entire data set: do parameter estimates match?

As sample size of a regression problem increases, standard errors of model parameters decrease, typically as a function of \sqrt{n} . Thus, the 500-data point subset is likely to have much wider error bars than an analysis of the full 636616 data points. How much wider is somewhat difficult to estimate, as data are spatially autocorrelated, making the \sqrt{n} an optimistic estimate. Following the logic of data cloning (Lele *et al.*, 2010), who show that repeating their data set 100 times yields $\sqrt{100} = 10$ times to narrow standard errors, one can argue that the standard error computed from the subset-models will be $\sqrt{636616}/\sqrt{500} = 34.6$ times too wide.

We cannot test this idea on the GLS, as we have no computer at our disposal that can invert a 600000×600000 matrix as required by the GLS. Thus, we cannot fit a GLS on the full data set (otherwise Liang *et al.* would have done that, too). Instead, we use the non-spatial linear model, which is biased due to spatial autocorrelation, as we shall see. Still, if our 35-fold correction worked for the non-spatial regression (the ordinary least square: OLS), we would be happy to use it for the GLS.

```

repeatedOLS <- function(N = 500) {
  # function running one random subset GLS
  index <- sample(nrow(bpr_data), N)
  subset <- bpr_data[index, ]
  subset$Sbreve <- subset$S/max(subset$S)
  flm <- lm(log(P) ~ log(Sbreve) + G + T3 + C1 + C3 + PET + IAA + E,
    data = subset)
  r2 <- cor(predict(flm), log(subset$P))^2 # R2
  return(c(ell = logLik(flm), AIC = AIC(flm), BIC = BIC(flm), R2 = r2,
    coef(flm)))
}
set.seed(123)
coefMatOLS <- t(replicate(50, repeatedOLS()))
signif(rbind(mean = colMeans(coefMatOLS), se = colSds(coefMatOLS)), 3)

```

	ell	AIC	BIC	R2	(Intercept)	log(Sbreve)	G	T3
mean	-774.0	1570.0	1610.0	0.3810	6.03	0.2710	0.0168	-1.47e-04
se	21.1	42.1	42.1	0.0365	0.75	0.0861	0.0044	2.71e-05
	C1	C3	PET	IAA	E			
mean	0.001780	0.002160	-0.003110	-1.66e-04	-0.000921			
se	0.000681	0.000771	0.000632	5.41e-05	0.000137			

```
# full data set:
Sbreve <- bpr_data$S/max(bpr_data$S)
summary(fullfm <- lm(log(P) ~ log(Sbreve) + G + T3 + C1 + C3 + PET + IAA +
  E, data = bpr_data))
```

```
Call:
lm(formula = log(P) ~ log(Sbreve) + G + T3 + C1 + C3 + PET +
  IAA + E, data = bpr_data)

Residuals:
    Min       1Q   Median       3Q      Max
-17.6776  -0.6125   0.0694   0.7169   5.7970

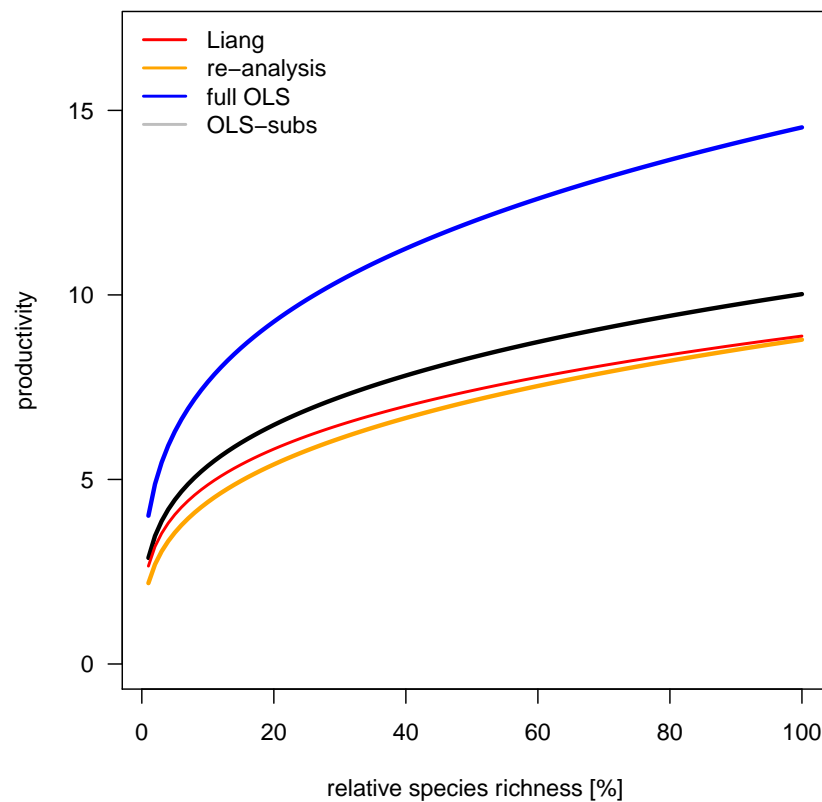
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.368e+00  2.196e-02  289.9  <2e-16 ***
log(Sbreve)  2.793e-01  2.295e-03  121.7  <2e-16 ***
G            1.514e-02  8.205e-05  184.5  <2e-16 ***
T3          -1.510e-04  9.050e-07 -166.9  <2e-16 ***
C1           1.748e-03  1.393e-05  125.5  <2e-16 ***
C3           2.111e-03  1.781e-05  118.6  <2e-16 ***
PET          -3.023e-03  1.297e-05 -233.1  <2e-16 ***
IAA          -1.609e-04  1.032e-06 -155.9  <2e-16 ***
E           -9.123e-04  2.868e-06 -318.1  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.159 on 636607 degrees of freedom
Multiple R2: 0.3654, Adjusted R2: 0.3654
F-statistic: 4.582e+04 on 8 and 636607 DF, p-value: < 2.2e-16
```

Resampling OLS and full data set analysis yield similar parameter estimates (apart from the intercept). As result, the **estimated effect of increasing “relative species richness” is substantially lower** in the subsampling approach than in the full model. We are confident that this effect will be similarly present in the GLS analysis and return to this issue in the section on increasing sample size of the subsampling.

Let's plot the results:

```
betaOLS <- coefMatOLS[, 5:13]
predsOLS <- X %*% t(betaOLS)
# CIsOLS <- apply(predsOLS, 1, quantile, c(0.025, 0.975))
par(mar = c(5, 4, 1, 1))
# subsample CIs:
plot(1:100, exp(rowMeans(predsOLS)), type = "l", las = 1, lwd = 3, ylim = c(0,
  17), xlab = "relative species richness [%]", ylab = "productivity")
# GLS
lines(1:100, exp(rowMeans(preds)), lwd = 3, col = "orange")
# full data
Xforfull <- data.frame(Intercept = 1, Sbreve = seq(0.01, 1, by = 0.01),
  t(colMeans(bpr_data[, c("G", "T3", "C1", "C3", "PET", "IAA", "E")]))))
predsLM <- predict(fullfm, newdata = Xforfull, se.fit = T)
lines(1:100, exp(predsLM$fit), col = "blue", lwd = 3)
# Liang line
lines(1:100, exp(liangLine), col = "red", lwd = 2)
legend("topleft", bty = "n", col = c("red", "orange", "blue", "grey"),
  legend = c("Liang", "re-analysis", "full OLS", "OLS-subs"), lwd = 2,
  lty = 1)
```

The take-home message of this is: subsampling causes bias, at least for the specific way \hat{S} is computed by Liang *et al.* (2016).

C.2.1 Scaling standard errors from subsamples to full data

To take the subsampling correction a step further, we now run subsamples of different size and try to identify a relationship between their (subsampling-derived) standard error for $\hat{\theta}$ and size of the subsample, B . As it happens, this relationship is best presented as a straight line on a log-log-plot.

```
SEQ <- c(100, 500, 1000, 5000, 10000, 50000, 1e+05, 5e+05)
set.seed(123)
count <- 0
coefThetaOLS500tofull <- 0
for (i in SEQ) {
  count <- count + 1
  fms <- replicate(50, repeatedOLS(N = i)[6]) # value 6 is theta
  coefThetaOLS500tofull[count] <- sd(fms)
  cat(i, "\n")
  rm(fms)
}
```

```
100 500 1000 5000 10000 50000 1e+05 5e+05
```

```
sefull <- summary(fullfm)$coefficients[2, 2] # se of theta
plot(c(SEQ, 636616), c(coefThetaOLS500tofull, sefull), log = "xy", xlab = "B",
     ylab = "sd_theta", pch = 16, las = 1)
summary(lm(log(c(coefThetaOLS500tofull, sefull)) ~ log(c(SEQ, 636616))))
```

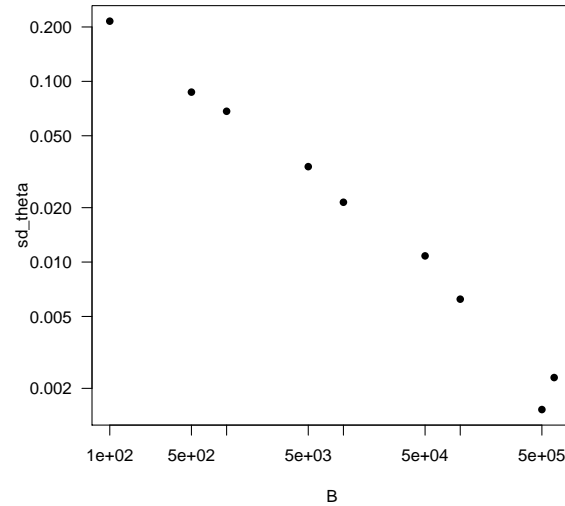
```
Call:
lm(formula = log(c(coefThetaOLS500tofull, sefull)) ~ log(c(SEQ,
636616)))
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.44880 -0.09864  0.08049  0.08931  0.26604
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.04955     0.25037   4.192  0.00408 **
log(c(SEQ, 636616)) -0.53995     0.02553 -21.146 1.33e-07 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2229 on 7 degrees of freedom
Multiple R2: 0.9846, Adjusted R2: 0.9824
F-statistic: 447.2 on 1 and 7 DF, p-value: 1.331e-07
```



The standard error of the estimate of θ clearly decreases nicely as a power-law function of sample size, and fits well to the observed values:

```
exp(1.05 - 0.54 * log(636616)) # coefs from regression
```

```
[1] 0.002098531
```

```
sfull
```

```
[1] 0.002295168
```

To apply this to the GLS, we have to make the assumption that the same proportionality holds for the GLS as for the OLS. If so, we can compute the standard error of the GLS, based on the subsampling of size $B = 500$, as follows (as a simple rule of three):

$$\hat{s}_{\text{GLS}} = e^{\frac{\ln(s_{\text{OLSfull}})}{\ln(s_{\text{OLS500}})} \ln(s_{\text{GLS500}})} = e^{\frac{\ln(0.002295)}{\ln(0.07239)} \ln(s_{\text{GLS500}})} = e^{2.314 \ln(s_{\text{GLS500}})}$$

Thus, for the observed standard error of the GLS for subsampling size $B = 500$ we have $s_{\text{GLS500}} = 0.0921$ (see section B.2), which leads to an estimated standard error of the full model of $\hat{s}_{\text{GLSfull}} = 0.00263$:

```
exp(log(sfull)/log(coefThetaOLS500tofull[2]) * log(colSds(coefMat)[6]))
```

```
[1] 0.002632888
```

This is (obviously) dramatically better than the “bootstrap” estimate presented above, but still thrice the 0.0009 of Liang *et al.* (2016).

C.2.2 Increasing the size of the subsamples for the GLS

As another step, we can briefly check whether increasing the size of the subsample from 500 to 5000 for the GLS makes an appreciable difference.¹¹

```
if ("coefMatGLS500to5000.Rdata" %in% dir()) {
  load("coefMatGLS500to5000.Rdata")
} else {
  set.seed(123)
  system.time(coefMatGLS500 <- t(replicate(10, repeatedGLS(N = 500)))) #5s
  system.time(coefMatGLS1000 <- t(replicate(10, repeatedGLS(N = 1000)))) # 1000: 72s
```

¹¹Beyond 5000 the GLS becomes extremely time consuming, running for many hours per model run.

```
system.time(coefMatGLS2000 <- t(replicate(10, repeatedGLS(N = 2000)))) # 2000: 437s
system.time(coefMatGLS5000 <- t(replicate(10, repeatedGLS(N = 5000)))) # 5000: 6238s
save(coefMatGLS500, coefMatGLS1000, coefMatGLS2000, coefMatGLS5000,
      file = "coefMatGLS500to5000.Rdata")
}
signif(rbind(mean = colMeans(coefMatGLS500), se = colSds(coefMatGLS500)),
       3)
```

	ell	AIC	BIC	R2 (Intercept)	log(Sbrev)	G	T3
mean	-819.0	1660.0	1700.0	0.3370	5.12	0.312	0.01880
se	32.8	65.6	65.6	0.0325	1.13	0.101	0.00415
	C1	C3	PET	IAA	E		
mean	0.001850	0.000933	-0.002900	-1.57e-04	-0.000561		
se	0.000829	0.001170	0.000756	5.55e-05	0.000149		

```
signif(rbind(mean = colMeans(coefMatGLS5000), se = colSds(coefMatGLS5000)),
       3)
```

	ell	AIC	BIC	R2 (Intercept)	log(Sbrev)	G	T3
mean	-7520	15100	15100	0.2410	4.140	0.3100	0.01550
se	111	223	223	0.0316	0.498	0.0424	0.00289
	C1	C3	PET	IAA	E		
mean	0.001510	0.000209	-0.002300	-1.03e-04	-1.75e-04		
se	0.000231	0.000284	0.000232	1.53e-05	6.41e-05		

```
# R2 goes down, as do SEs plot(log10(c(5, 72, 437, 6238)) ~
# log10(c(500, 1000, 2000, 5000))) runtime increases as power law ...
# plot(c(500, 1000, 2000, 5000), c(0.105, 0.062, .045, 0.0194), log='')
# SE goes down as power law ... lm(log10(c(0.105, 0.062, .045, 0.0194))
# ~ log10(c(500, 1000, 2000, 5000))) # logY=0.955-0.714logX ->
# logY(X=630000) = 0.955-0.714*56.799 = -39 -> ~0
```

The estimate of θ and most other model parameters is similar for sample size 500 and 5000, while the intercept and the model's R^2 goes down.

What does that mean?

1. Using small subsamples of a large data set leads to dramatic increases in parameter uncertainty. Thus, the (corrected) “bootstrap” approach of Liang *et al.* (2016) is likely to be very pessimistic. Since they did an error in the computation of their standard errors, the uncertainty limits provided in the original paper cannot be trusted.
2. Using the full data set led to very different estimates for the model (comparing OLS for the resampled and the full data set, i.e. the grey and the blue curves above). By analogy, also the resampling-based estimates of the GLS are likely to be substantially biased.¹²
3. Correction for spatial autocorrelation leads to lower estimates for productivity, roughly 80% of the expected value of P along the tree species richness gradient (orange line).

Although we have introduced a way to estimate the uncertainty around the predicted relationship from the data by approximation in the previous section, we have not used it until the very end.

C.3 Stratified sampling of subsets

Each ecoregion has a certain share of the terrestrial surface, but within each ecoregion, not all area is forested. Crowther *et al.* (2015) provide estimates of the total land area as well as the number of trees for each ecoregion. We use the latter as means of stratification, i.e. draw plots proportional to the number of trees in that ecoregion.

```
library(nlme)
library(matrixStats)
biome_trees$stratProb <- biome_trees$nOfTrees/sum(biome_trees$nOfTrees)
bpr_data <- merge(bpr_data, biome_trees)
# head(bpr_data)

repeatedGLSstrat <- function(N = 500) {
  # function running one stratified random subset GLS sample the
  # Ecoregion-ID according to its probability:
  ecoSample <- sample(biome_trees$Ecoregion, N, prob = biome_trees$stratProb,
```

¹²The origin of this bias has probably to do with the re-scaling of species richness in each subset. If no high-diversity plot is present, then the left part of the relationship is stretched out to the right hand side without higher P-values, thereby pulling the entire curve down.

```

      replace = T) # how many samples from all ecoregions
# table(ecoSample) Now, for each ecoregion, take a random subset from
# the full data set and rbind them:
subset <- NA
for (i in biome_trees$Ecoregion) {
  subs <- bpr_data[bpr_data$Ecoregion == i, ] # only this biome
  subset2nd <- subs[subs$FID %in% sample(subs$FID, sum(ecoSample ==
    i), replace = {
      if (sum(ecoSample == i) > length(subs$FID))
        TRUE else FALSE
    }), ]
  # sorry for this convoluted expression ...
  subset <- rbind(subset, subset2nd)
}
subset <- subset[-1, ] # remove the NA-row
# re-jitter geographical coordinates, if data points we resampled!
if (sum(ecoSample == i) > length(subs$FID)) {
  subset$lat <- jitter(subset$Lat)
  subset$lon <- jitter(subset$Lon)
}
# Compute Sbreve as before and run GLS:
subset$Sbreve <- subset$S/max(subset$S)
fgls <- gls(log(P) ~ log(Sbreve) + G + T3 + C1 + C3 + PET + IAA + E,
  correlation = corSpher(c(50, 0.8), form = ~lon + lat, nugget = T,
    fixed = F), data = subset)
r2 <- cor(predict(fgls), log(subset$P))^2 # R2
return(c(ell = logLik(fgls), AIC = AIC(fgls), BIC = BIC(fgls), R2 = r2,
  coef(fgls)))
}

if ("coefMatStrat.Rdata" %in% dir()) {
  load("coefMatStrat.Rdata")
} else {
  set.seed(1234)
  coefMatStrat <- t(replicate(50, repeatedGLSstrat()))
  save(coefMatStrat, file = "coefMatStrat.Rdata")
}
signif(rbind(mean = colMeans(coefMatStrat), se = colSds(coefMatStrat)),
  3)

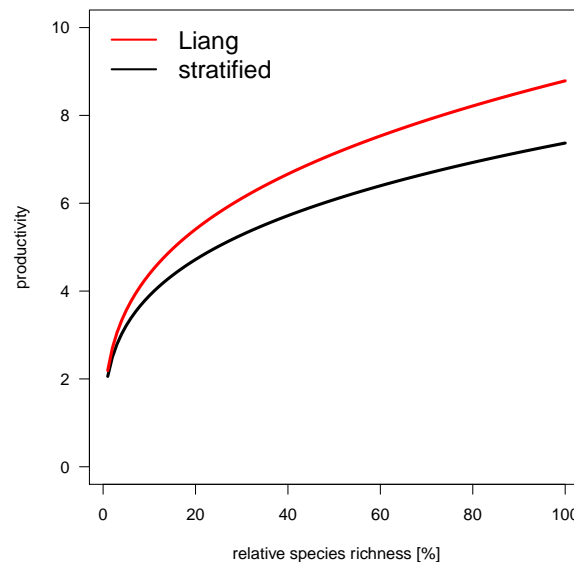
```

	ell	AIC	BIC	R2	(Intercept)	log(Sbreve)	G	T3
mean	-770.0	1560.0	1610.0	0.2980	3.05	0.2770	0.02060	-1.00e-04
se	26.9	53.7	53.8	0.0463	0.99	0.0797	0.00596	4.01e-05
	C1	C3	PET	IAA	E			
mean	5.14e-05	0.001100	-0.000628	-3.36e-05	-0.000238			
se	3.27e-04	0.000772	0.000502	3.36e-05	0.000226			

```

load("coefMatStrat.Rdata")
betaStrat <- coefMatStrat[, 5:13]
# set all values but log(Sbreve) to their mean: X <-
# as.matrix(data.frame('Intercept'=1, Sbreve=log(seq(0.01, 1,
# by=0.01)), t(colMeans(bpr_data[, c('G', 'T3', 'C1', 'C3', 'PET',
# 'IAA', 'E')]))))
predsStrat <- X %*% t(betaStrat)
par(mar = c(5, 4, 1, 1))
plot(1:100, exp(rowMeans(predsStrat)), type = "l", las = 1, lwd = 3, ylim = c(0,
  10), xlab = "relative species richness [%]", ylab = "productivity",
  log = "")
# add standard model:
lines(1:100, exp(rowMeans(preds)), lwd = 3, col = "red")
# lines(1:100, exp(rowMeans(predsStrat)), lwd=3, col='white')
legend("topleft", legend = c("Liang", "stratified"), col = c("red", "black"),
  lwd = 2, lty = 1, bty = "n", cex = 1.5)

```



The effect is moderate, with slightly lower values than the original non-stratified approach. This result suggests that also with non-stratified sampling always some tropical plots with high species richness are drawn, making the original \hat{S} robust to unrepresentative sampling.

C.4 Compute distances between plots as Great Circle distances, rather than Euclidean (thereby accommodating the spherical nature of earth)

In the spatial model, the GLS, the distances between plots are computed from x-y-coordinates as Euclidean distance.¹³ On earth, the distance between two points cannot be computed as Euclidean distance, and instead requires the computation of the so-called Great Circle distance.¹⁴

We employ the “haversine” correlation structure provided on stackexchange to implement a correct distance measure within GLS.¹⁵ We define the model and run it repeatedly.

```
# Calculates the geodesic distance between two points specified by
# radian latitude/longitude using Haversine formula. output in km
haversine <- function(x0, x1, y0, y1) {
  a <- sin((y1 - y0)/2)^2 + cos(y0) * cos(y1) * sin((x1 - x0)/2)^2
  v <- 2 * asin(min(1, sqrt(a)))
  6371 * v
}

# function to compute geodesic haversine distance given two-column
# matrix of longitude/latitude input is assumed in form decimal degrees
# if radians = F note fields::rdist.earth is more efficient
haversineDist <- function(xy, radians = F) {
  if (ncol(xy) > 2)
    stop("Input must have two columns (longitude and latitude)")
  if (radians == F)
    xy <- xy * pi/180
  hMat <- matrix(NA, ncol = nrow(xy), nrow = nrow(xy))
  for (i in 1:nrow(xy)) {
    for (j in i:nrow(xy)) {
      hMat[j, i] <- haversine(xy[i, 1], xy[j, 1], xy[i, 2], xy[j,
        2])
    }
  }
  as.dist(hMat)
}

## for most methods, machinery from corSpatial will work without
## modification
Initialize.corHaversine <- nlme::Initialize.corSpatial
```

¹³This is implicit in the `corSpher`-argument (Pinheiro & Bates, 2000). The different correlation structures (exponential, Gaussian, linear and spherical) describe how the covariance between two locations decreases with their distance. Spherical, for example, is similar to exponential, but with a more linear shape.

¹⁴https://en.wikipedia.org/wiki/Great-circle_distance

¹⁵Taken from <https://stackoverflow.com/questions/18857443/specifying-a-correlation-structure-for-a-linear-mixed-model-using-the-ramps-pack>.

```

recalc.corHaversine <- nlme:::recalc.corSpatial
Variogram.corHaversine <- nlme:::Variogram.corSpatial
corFactor.corHaversine <- nlme:::corFactor.corSpatial
corMatrix.corHaversine <- nlme:::corMatrix.corSpatial
coef.corHaversine <- nlme:::coef.corSpatial
"coef<-corHaversine" <- nlme:::"coef<-corSpatial"

## Constructor for the corHaversine class
corHaversine <- function(value = numeric(0), form = ~1, mimic = "corSpher",
  nugget = FALSE, fixed = FALSE) {
  spClass <- "corHaversine"
  attr(value, "formula") <- form
  attr(value, "nugget") <- nugget
  attr(value, "fixed") <- fixed
  attr(value, "function") <- mimic
  class(value) <- c(spClass, "corStruct")
  value
} # end corHaversine class
environment(corHaversine) <- asNamespace("nlme")

Dim.corHaversine <- function(object, groups, ...) {
  if (missing(groups))
    return(attr(object, "Dim"))
  val <- Dim.corStruct(object, groups)
  val[["start"]] <- c(0, cumsum(val[["len"]] * (val[["len"]] - 1)/2)[-val[["M"]]])
  ## will use third component of Dim list for spClass
  names(val)[3] <- "spClass"
  val[[3]] <- match(attr(object, "function"), c("corSpher", "corExp",
    "corGaus", "corLin", "corRatio"), 0)
  val
}
environment(Dim.corHaversine) <- asNamespace("nlme")

## getCovariate method for corHaversine class
getCovariate.corHaversine <- function(object, form = formula(object), data) {
  if (is.null(covar <- attr(object, "covariate"))) {
    # if object lacks covariate attribute if object lacks data
    if (missing(data)) {
      stop("need data to calculate covariate")
    }
    covForm <- getCovariateFormula(form)
    if (length(all.vars(covForm)) > 0) {
      # if covariate present if formula includes intercept
      if (attr(terms(covForm), "intercept") == 1) {
        covForm <- eval(parse(text = paste("~", deparse(covForm[[2]]),
          "-1", sep = ""))) # remove intercept
      }
      # can only take covariates with correct names
      if (length(all.vars(covForm)) > 2)
        stop("corHaversine can only take two covariates, 'lon' and 'lat'")
      if (!all(all.vars(covForm) %in% c("lon", "lat")))
        stop("covariates must be named 'lon' and 'lat'")
      covar <- as.data.frame(unclass(model.matrix(covForm, model.frame(covForm,
        data, drop.unused.levels = TRUE))))
      covar <- covar[, order(colnames(covar), decreasing = T)] # order as lon ... lat
    } else {
      covar <- NULL
    }
  }

  if (!is.null(getGroupsFormula(form))) {
    # if groups in formula extract covar by groups
    grps <- getGroups(object, data = data)
    if (is.null(covar)) {
      covar <- lapply(split(grps, grps), function(x) as.vector(dist(1:length(x))))
    } else {
      giveDist <- function(el) {
        el <- as.matrix(el)
        if (nrow(el) > 1)
          as.vector(haversineDist(el)) else numeric(0)
      }
      covar <- lapply(split(covar, grps), giveDist)
    }
    covar <- covar[sapply(covar, length) > 0] # no 1-obs groups
  } else {
    # if no groups in formula extract distance

```



```

      if (is.null(covar)) {
        covar <- as.vector(dist(1:nrow(data)))
      } else {
        covar <- as.vector(haversineDist(as.matrix(covar)))
      }
    }
    if (any(unlist(covar) == 0)) {
      # check that no distances are zero
      stop("cannot have zero distances in \"corHaversine\"")
    }
  }
  covar
} # end method getCovariate
environment(getCovariate.corHaversine) <- asNamespace("nlme")

```

```

set.seed(12)
index <- sample(nrow(bpr_data), 500)
subset <- bpr_data[index, ]
subset$Sbreve <- subset$S/max(subset$S)
fgls <- gls(log(P) ~ log(Sbreve) + G + T3 + C1 + C3 + PET + IAA + E, correlation = corHaversine(form =
  lat, mimic = "corSpher"), data = subset)
summary(fgls)

```

```

Generalized least squares fit by REML
Model: log(P) ~ log(Sbreve) + G + T3 + C1 + C3 + PET + IAA + E
Data: subset
      AIC      BIC    logLik
1697.479 1743.64 -837.7396

Correlation Structure: corHaversine
Formula: ~lon + lat
Parameter estimate(s):
      range
0.001371426

Coefficients:
              Value Std.Error   t-value p-value
(Intercept)  6.264020  0.7383701   8.483578  0.0000
log(Sbreve)   0.341606  0.0833646   4.097728  0.0000
G              0.017556  0.0030472   5.761253  0.0000
T3            -0.000197  0.0000334  -5.876495  0.0000
C1              0.000379  0.0005550   0.682931  0.4950
C3              0.002481  0.0006231   3.981967  0.0001
PET            -0.002139  0.0005147  -4.156098  0.0000
IAA            -0.000088  0.0000421  -2.089247  0.0372
E             -0.001024  0.0001069  -9.578685  0.0000

Correlation:
      (Intr) lg(Sb) G      T3      C1      C3      PET      IAA
log(Sbreve)  0.626
G            -0.185 -0.254
T3           -0.680 -0.453  0.140
C1            0.194 -0.332  0.044  0.258
C3            0.046  0.004  0.041 -0.273 -0.250
PET           -0.576  0.098 -0.034  0.067 -0.831 -0.045
IAA           -0.384  0.235 -0.108 -0.084 -0.931  0.045  0.898
E            -0.111  0.058 -0.186  0.057  0.149  0.168 -0.031 -0.119

Standardized residuals:
      Min      Q1      Med      Q3      Max
-3.9392597 -0.5070892  0.0228970  0.5715560  3.5128986

Residual standard error: 1.162089
Degrees of freedom: 500 total; 491 residual

```

```

repeatedGLShaversine <- function(N = 500) {
  # function running one random subset GLS
  index <- sample(nrow(bpr_data), N)
  subset <- bpr_data[index, ]
  subset$Sbreve <- subset$S/max(subset$S)
  fgls <- gls(log(P) ~ log(Sbreve) + G + T3 + C1 + C3 + PET + IAA + E,
    correlation = corHaversine(form = ~lon + lat, mimic = "corSpher"),
    data = subset)
  r2 <- cor(predict(fgls), log(subset$P))^2 # R2
  return(c(ell = logLik(fgls), AIC = AIC(fgls), BIC = BIC(fgls), R2 = r2,
    coef(fgls)))
}

```

```

}
set.seed(123)
coefMathaversine <- t(replicate(50, repeatedGLShaversine()))
library(matrixStats) # for colSds
signif(rbind(mean = colMeans(coefMathaversine), se = colSds(coefMathaversine)),
3)

```

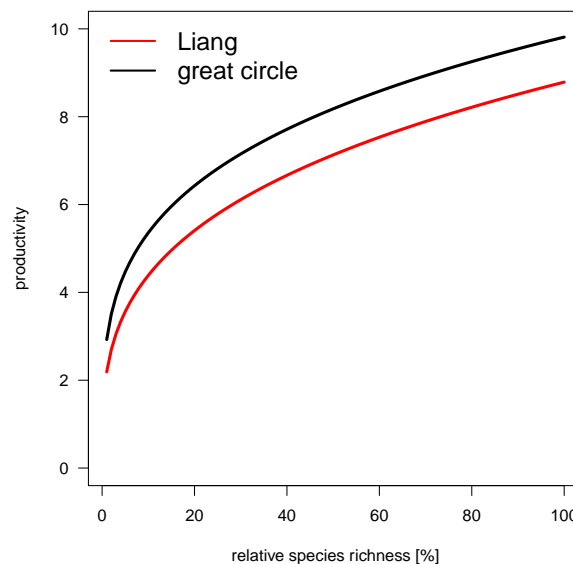
	ell	AIC	BIC	R2	(Intercept)	log(Sbrev)	G	T3
mean	-830.0	1680.0	1730.0	0.3770	6.050	0.2630	0.01580	-0.000155
se	20.6	41.1	41.1	0.0387	0.779	0.0896	0.00441	0.000031

	C1	C3	PET	IAA	E
mean	0.001800	0.002010	-0.003050	-1.65e-04	-0.000942
se	0.000636	0.000808	0.000547	4.78e-05	0.000105

```

betahaversine <- coefMathaversine[, 5:13]
predshaversine <- X %*% t(betahaversine)
par(mar = c(5, 4, 1, 1))
plot(1:100, exp(rowMeans(predshaversine)), type = "l", las = 1, lwd = 3,
      ylim = c(0, 10), xlab = "relative species richness [%]", ylab = "productivity",
      log = "")
# add standard model:
lines(1:100, exp(rowMeans(preds)), lwd = 3, col = "red")
legend("topleft", legend = c("Liang", "great circle"), col = c("red", "black"),
      lwd = 2, lty = 1, bty = "n", cex = 1.5)

```



In this case, the correct representation of spatial distances increases overall productivity estimates a bit, but leaves the TSP-P-relationship unchanged.

C.5 Define species richness relative to what is regionally possible

The definition of \tilde{S} in the subsets is relative to the largest value in the subset. When analysing the full data set, by the same logic, all tree species richness values are a fraction of the most diverse plot with 405 species. Thus, the reported relationship must be read as: “If we increase whatever number of species we currently have in a plot to 400, then we are moving on this curve.” Ecologically, this makes no sense. Whatever the reason why there are only a handful species in a plot of 1/10th of a hectare in temperate and boreal forests, it also prevents a “tropic richness” in these sites. Thus, one cannot meaningfully increase species richness outside the tropics to a tropical level. As a consequence, the depicted figure lacks ecological interpretability.

What the reader, and we suspect also some of the authors, probably interpret into the x-axis is “species richness relative to what would be possible at this site”. There are different ways to define “what is possible at this site”: (a) relative to what the maximum recorded value for this biome is, (b) relative to the richness of plots in the vicinity, (c) relative to the number of trees in the local/regional species pool, and possibly others. In the following, we use approach (b), relative to other plots in the region. The biome is, in our opinion, too large, while the third approach would require an analysis far beyond what current data allow us to do.

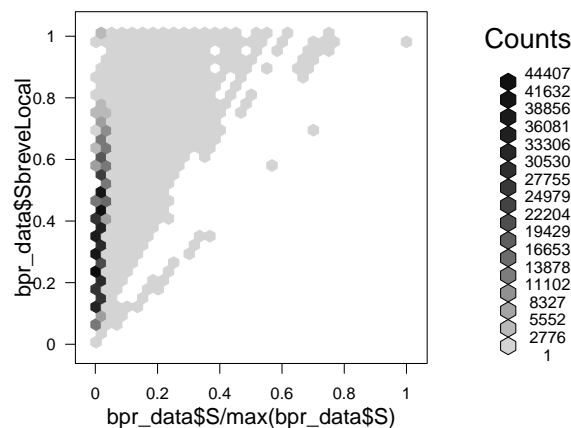
We define the “region” around a plot as being within a raster cell of $100 \times 100 \text{ km}^2$.

```
library(raster)
library(sp)
# do everything in equal-area Mollweide projection:
prj <- "+proj=moll +lon_0=0 +x_0=0 +y_0=0 +ellps=WGS84"
# set up a raster grid for the world:
wrld_moll <- spTransform(wrld_simpl, CRS = prj)
wrld_moll_raster <- raster(wrld_moll, res = 1e+05)
# transform plot coordinates into Mollweide:
plots <- SpatialPoints(coords = bpr_data[, c("Lon", "Lat")], proj4string = crs(wrld_simpl))
plots_moll <- spTransform(plots, CRS = prj)
# rasterise the data into it:
bpr_moll_raster <- rasterize(x = plots_moll, y = wrld_moll_raster, field = bpr_data$,
  fun = max)
bpr_moll_raster
```

```
class      : RasterLayer
dimensions : 178, 355, 63190 (nrow, ncol, ncell)
resolution : 1e+05, 1e+05 (x, y)
extent     : -17702743, 17797257, -9052898, 8747102 (xmin, xmax, ymin, ymax)
coord. ref.: +proj=moll +lon_0=0 +x_0=0 +y_0=0 +ellps=WGS84
data source: in memory
names      : layer
values     : 1, 405 (min, max)
```

```
# plot(bpr_moll_raster) # check extract Smax of the cell into which a
# plot falls
SmaxPerPlot <- extract(bpr_moll_raster, plots_moll)
SbreveLocal <- bpr_data$S/SmaxPerPlot
bpr_data$SbreveLocal <- SbreveLocal

# plot(bpr_data$S/max(bpr_data$S), bpr_data$SbreveLocal, pch='.',
# cex=2) #slow!
library(hexbin)
plot(hexbin(bpr_data$S/max(bpr_data$S), bpr_data$SbreveLocal))
```



The local relative species richness (\hat{S}_{local}) shows substantially more variability for low-richness plots.

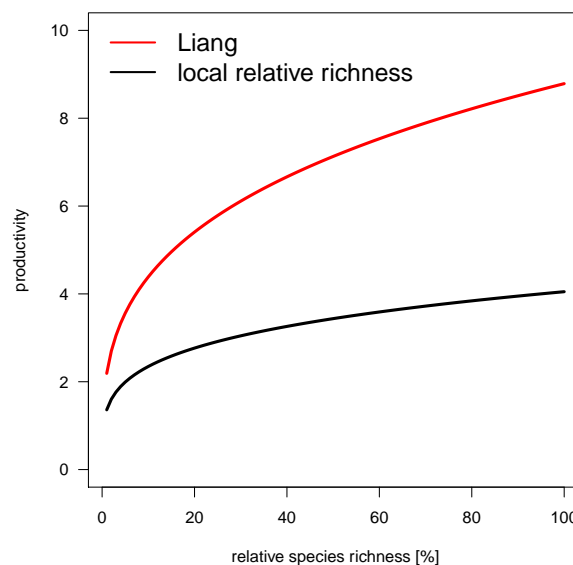
Now we can repeat the above analysis with a different relative tree species richness as response:

```
repeatedGLSlocalSbreve <- function(N = 500) {
  # function running one random subset GLS
  index <- sample(nrow(bpr_data), N)
  subset <- bpr_data[index, ]
  fgls <- gls(log(P) ~ log(SbreveLocal) + G + T3 + C1 + C3 + PET + IAA +
    E, correlation = corSpher(c(50, 0.8), form = ~lon + lat, nugget = T,
    fixed = T), data = subset)
  r2 <- cor(predict(fgls), log(subset$P))^2 # R2
  return(c(ell = logLik(fgls), AIC = AIC(fgls), BIC = BIC(fgls), R2 = r2,
    coef(fgls)))
}
set.seed(123)
coefMatlocalSbreve <- t(replicate(50, repeatedGLSlocalSbreve()))
```

```
library(matrixStats) # for colSds
signif(rbind(mean = colMeans(coefMatlocalSbreve), se = colSds(coefMatlocalSbreve)),
3)
```

	e11	AIC	BIC	R2	(Intercept)	log(SbreveLocal)	G
mean	-818.0	1660.0	1700.0	0.3360	3.870	0.2370	0.01710
se	19.8	39.7	39.7	0.0406	0.979	0.0995	0.00433
	T3	C1	C3	PET	IAA	E	
mean	-7.03e-05	0.002090	0.001160	-0.002790	-1.51e-04	-0.000641	
se	5.49e-05	0.000583	0.000695	0.000643	4.04e-05	0.000122	

```
betalocalSbreve <- coefMatlocalSbreve[, 5:13]
predslocalSbreve <- X %>% t(betalocalSbreve)
par(mar = c(5, 4, 1, 1))
plot(1:100, exp(rowMeans(predslocalSbreve)), type = "l", las = 1, lwd = 3,
      ylim = c(0, 10), xlab = "relative species richness [%]", ylab = "productivity",
      log = "")
# add standard model:
lines(1:100, exp(rowMeans(preds)), lwd = 3, col = "red")
legend("topleft", legend = c("Liang", "local relative richness"), col = c("red",
"black"), lwd = 2, lty = 1, bty = "n", cex = 1.5)
```



Representing species richness relative to what is possible in that region substantially reduces the effect of species richness on productivity, as well as the error margin of this relationship.

To repeat the message of this last plot: When moving from a plot that only comprises, say, 10% of the potential local tree species to one that contains all 100%, productivity increases from around 2.2 to 4.5 $\text{m}^3\text{ha}^{-1}\text{y}^{-1}$ (much less than the increase from 2.5 to 8 $\text{m}^3\text{ha}^{-1}\text{y}^{-1}$ reported by Liang *et al.* 2016).

C.6 Use relative, rather than absolute, productivity

It appears a bit odd to scale species richness relative to what is locally possible, but not also productivity. Tropical forests can be expected to be far more productive than boreal ones, but the effect of relative species richness on *relative* productivity may be similar.

To explore this idea, we follow the same approach as with local relative species richness and compute local relative productivity, P_{local} . This we then analyse with the local relative species richness (\hat{S}_{local}).

Again we define the “region” around a plot as being within a raster cell of $100 \times 100 \text{ km}^2$.

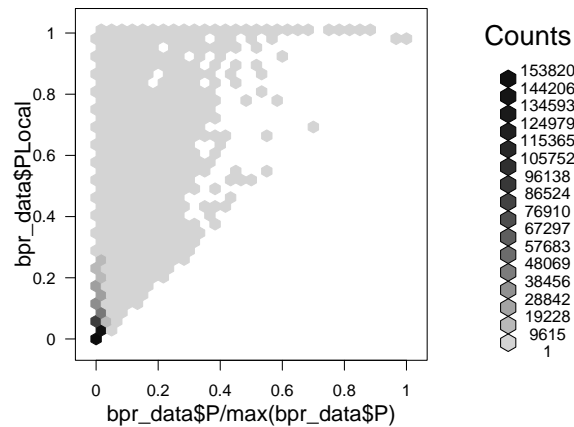
```
library(raster)
library(maptools)
data(wrld_simpl)
# do everything in equal-area Mollweide projection:
prj <- "+proj=moll +lon_0=0 +x_0=0 +y_0=0 +ellps=WGS84"
# set up a raster grid for the world:
wrld_moll <- spTransform(wrld_simpl, CRS = prj)
wrld_moll_raster <- raster(wrld_moll, res = 1e+05)
# transform plot coordinates into Mollweide:
plots <- SpatialPoints(coords = bpr_data[, c("Lon", "Lat")], proj4string = crs(wrld_simpl))
```

```
plots_moll <- spTransform(plots, CRS = prj)
# rasterise the data into it:
bpr_moll_raster <- rasterize(x = plots_moll, y = wrld_moll_raster, field = bpr_data$P,
  fun = max)
bpr_moll_raster
```

```
class      : RasterLayer
dimensions : 178, 355, 63190 (nrow, ncol, ncell)
resolution : 1e+05, 1e+05 (x, y)
extent     : -17702743, 17797257, -9052898, 8747102 (xmin, xmax, ymin, ymax)
coord. ref.: +proj=moll +lon_0=0 +x_0=0 +y_0=0 +ellps=WGS84
data source: in memory
names      : layer
values     : 0.01, 952.02 (min, max)
```

```
# plot(bpr_moll_raster) # check extract Smax of the cell into which a
# plot falls
PmaxPerPlot <- extract(bpr_moll_raster, plots_moll)
PLocal <- bpr_data$P/PmaxPerPlot
bpr_data$PLocal <- PLocal

library(hexbin)
plot(hexbin(bpr_data$P/max(bpr_data$P), bpr_data$PLocal))
```



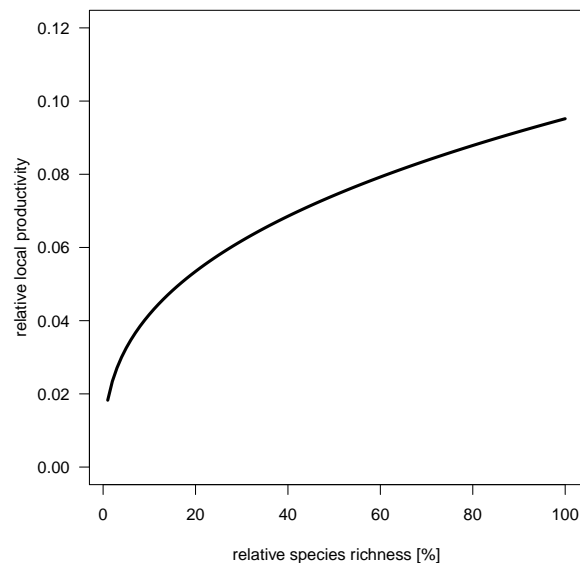
Now we can repeat the above analysis with a different relative (local) productivity as response:

```
repeatedGLSlocalP <- function(N = 500) {
  # function running one random subset GLS
  index <- sample(nrow(bpr_data), N)
  subset <- bpr_data[index, ]
  fgls <- gls(log(PLocal) ~ log(SbreveLocal) + G + T3 + C1 + C3 + PET +
    IAA + E, correlation = corSpher(c(50, 0.8), form = ~lon + lat,
    nugget = T, fixed = T), data = subset)
  r2 <- cor(predict(fgls), log(subset$P))^2 # R2
  return(c(ell = logLik(fgls), AIC = AIC(fgls), BIC = BIC(fgls), R2 = r2,
    coef(fgls)))
}
set.seed(123)
coefMatlocalP <- t(replicate(50, repeatedGLSlocalP()))
library(matrixStats) # for colSds
signif(rbind(mean = colMeans(coefMatlocalP), se = colSds(coefMatlocalP)),
  3)
```

	ell	AIC	BIC	R2 (Intercept)	log(SbreveLocal)	G	T3
mean	-859	1740	1780	0.1350	-0.308	0.358	0.0127
se	19	38	38	0.0449	1.170	0.100	0.0036
		C1	C3	PET	IAA	E	
mean	-0.000752	0.003310	-0.001900	-1.18e-06	-0.000598		
se	0.000809	0.000905	0.000876	5.86e-05	0.000147		

```
# plot(hexbin(log10(bpr_data$S), log10(bpr_data$PLocal)))
# plot(hexbin(bpr_data$S/405, bpr_data$PLocal))
```

```
betalocalP <- coefMatlocalP[, 5:13]
predslocalP <- X %*% t(betalocalP)
par(mar = c(5, 4, 1, 1))
plot(1:100, exp(rowMeans(predslocalP)), type = "l", las = 1, lwd = 3, ylim = c(0,
  0.12), xlab = "relative species richness [%]", ylab = "relative local productivity",
  log = "")
```



This does not look like an improvement, as the model fit is poorer (R^2 of 0.135 compared to 0.34). Interestingly, however, (a) the effect of species richness is stronger (larger value of θ); and (b) even at maximal species richness only 10% of maximal productivity is reached.

Note that we used the *local* relative species richness (\tilde{S}_{local}); the model with the original \tilde{S} is similarly bad, and with a lower value for θ (around 0.213). While this may actually be interesting in itself, we do not pursue this line of thought further here.

D A new analysis

In summary, our explorations above have shown substantial effects of (in that order)

1. computing \tilde{S} relative to maximal *local* species richness;
2. stratifying the data proportional to an ecoregion's share of global forest cover;
3. computing spatial distances on a sphere;
4. stratification of samples by ecoregion forest cover;
5. correcting for spatial autocorrelation;
6. correcting standard errors for subsampling.

The relative scaling of productivity did not improve the global model and is not further considered here.

It is quite possible that these alterations of the original model interact, i.e. that the substantial bias we have seen introduced by stratification may be caused by the way \tilde{S} is computed, and would be removed with a locally computed \tilde{S}_{local} . Exploring *all* these option combinations is beyond the scope of this re-analysis.

So, in a nutshell, the *new global model* uses a locally-scaled species richness as predictor, computes spatial distances on a sphere when correcting for spatial autocorrelation in the data and stratifies sampling by each ecoregion's forest cover.

```
repeatedGLSnew <- function(N = 500) {
  # function running one stratified random subset GLS sample the
  # Ecoregion-ID according to its probability:
  ecoSample <- sample(biome_trees$Ecoregion, N, prob = biome_trees$stratProb,
    replace = T) # how many samples from all ecoregions
  # table(ecoSample) Now, for each ecoregion, take a random subset from
  # the full data set and rbind them:
  subset <- NA
  for (i in biome_trees$Ecoregion) {
    subs <- bpr_data[bpr_data$Ecoregion == i, ] # only this biome
```



```

subset2nd <- subs[subs$FID %in% sample(subs$FID, sum(ecoSample ==
  i), replace = {
    if (sum(ecoSample == i) > length(subs$FID))
      TRUE else FALSE
  }), ]
# sorry for this convoluted expression ...
subset <- rbind(subset, subset2nd)
}
subset <- subset[-1, ] # remove the NA-row
# re-jitter geographical coordinates, if data points were resampled!
if (sum(ecoSample == i) > length(subs$FID)) {
  subset$lat <- jitter(subset$lat)
  subset$lon <- jitter(subset$lon)
}
fgls <- gls(log(P) ~ log(SbreveLocal) + G + T3 + C1 + C3 + PET + IAA +
  E, correlation = corHaversine(form = ~lon + lat, mimic = "corSpher"),
  data = subset)
r2 <- cor(predict(fgls), log(subset$P))^2 # R2
return(c(ell = logLik(fgls), AIC = AIC(fgls), BIC = BIC(fgls), R2 = r2,
  coef(fgls)))
}
set.seed(123)
coefMatNew <- t(replicate(50, repeatedGLSnew()))
signif(rbind(mean = colMeans(coefMatNew), se = colSds(coefMatNew)), 3)

```

	ell	AIC	BIC	R2	(Intercept)	log(SbreveLocal)	G
mean	-819.0	1660.0	1710.0	0.326	2.970	0.0646	0.03020
se	23.9	47.8	47.8	0.032	0.826	0.1050	0.00579
	T3	C1	C3	PET	IAA	E	
mean	-1.16e-04	0.000649	0.001610	-0.001300	-1.05e-04	-0.000661	
se	3.53e-05	0.000250	0.000437	0.000407	2.76e-05	0.000189	

Here we add the correction for subsampling, specifically computing SDs of the predictions from the subsamples and correcting them to the full data. We first do that for the fits provided by the re-analysis in the original form.

```

predsMeansGLS <- rowMeans(preds)
CIsdGLS <- apply(preds, 1, sd)
CIsdGLScorr <- exp(2.341 * log(CIsdGLS))
CIsGLS <- rbind(predsMeansGLS + 2 * CIsdGLScorr, predsMeansGLS - 2 * CIsdGLScorr)

```

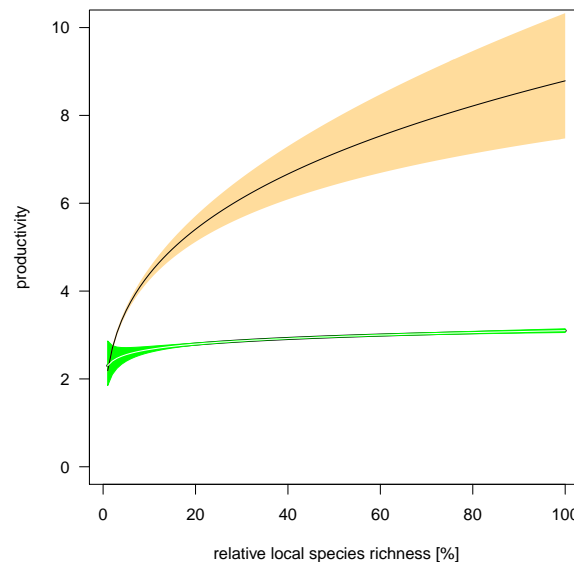
Now we plot the new model, and as comparison the original.

```

betaNew <- coefMatNew[, 5:13]
predsNew <- X %*% t(betaNew)
predsMeans <- rowMeans(predsNew)
CIsd <- apply(predsNew, 1, sd)
CIsdcorr <- exp(2.341 * log(CIsd))
CIsNew <- rbind(predsMeans + 2 * CIsdcorr, predsMeans - 2 * CIsdcorr)

par(mar = c(5, 4, 1, 1))
plot(1:100, exp(rowMeans(predsNew)), type = "l", las = 1, lwd = 3, ylim = c(0,
  10), xlab = "relative local species richness [%]", ylab = "productivity",
  log = "")
polygon(c(1:100, 100:1), c(exp(CIsNew[1, ]), rev(exp(CIsNew[2, ]))), border = "green",
  col = "green")
# add standard model:
polygon(c(1:100, 100:1), c(exp(CIsGLS[1, ]), rev(exp(CIsGLS[2, ]))), border = NA,
  col = rgb(255, 165, 0, 100, maxColorValue = 255))
lines(1:100, exp(predsMeansGLS), lwd = 1)
lines(1:100, exp(predsMeans), lwd = 1, col = "white")

```



As we can clearly see, the pattern observed by Liang *et al.* (2016) does not hold up to scrutiny when correcting the mistakes of their analysis. The overall effect of increasing species richness to the maximum possible at that site is negligible and levels off at less than 10%.¹⁶

The absolute levels of productivity presented in the above figure are conditional on the values chosen for environmental covariates and are thus not interpretable “globally”. To do so would require a marginal, rather than a conditional approach. As already the conditional effect at the mean value of the covariates is practically absent, so would probably be the marginal.¹⁷ Although that was one of our initial questions, it is not worth pursuing under these conditions.

E Comparison with ecoregion-specific fits

The lack of species richness-effect begs the question whether *within* ecoregions such a species richness-effect is present. It could well be that the variation in species-richness effects across ecoregions leads to a dilution of this signal at the global scale and the overall lack of effect in the full analysis.

We thus recompute with the same approach the fits to each ecoregion separately.

```
repeatedGLSbiome <- function(N = 500, data = NULL) {
  # function running random subset GLS take subset
  subset <- data[sample(nrow(data), min(N, nrow(data))), ]
  # fit model
  fgls <- try(gls(log(P) ~ log(SbreveLocal) + G + T3 + C1 + C3 + PET +
    IAA + E, correlation = corHaversine(form = ~lon + lat, mimic = "corSpher"),
    data = subset))
  if (inherits(fgls, "try-error")) {
    return(rep(NA, 13))
  } else {
    r2 <- cor(predict(fgls), log(subset$P))^2 # R2
    return(c(ell = logLik(fgls), AIC = AIC(fgls), BIC = BIC(fgls),
      R2 = r2, coef(fgls)))
  }
}

regionalFits <- list()
set.seed(123)
for (i in c(1, 2, 4:10, 12, 13)) {
  # 3: no data; 11: only 9 data points select region:
  thisEcoregion <- bpr_data[bpr_data$Ecoregion == i, ]
```

¹⁶As written above, it is probably in the interaction of the changes that the effect disappears. Changing, for example, the Haversine back to corSpher has hardly any effect in itself. One could now start to remove the changes one at a time to identify the step(s) that caused the original signal to disappear.

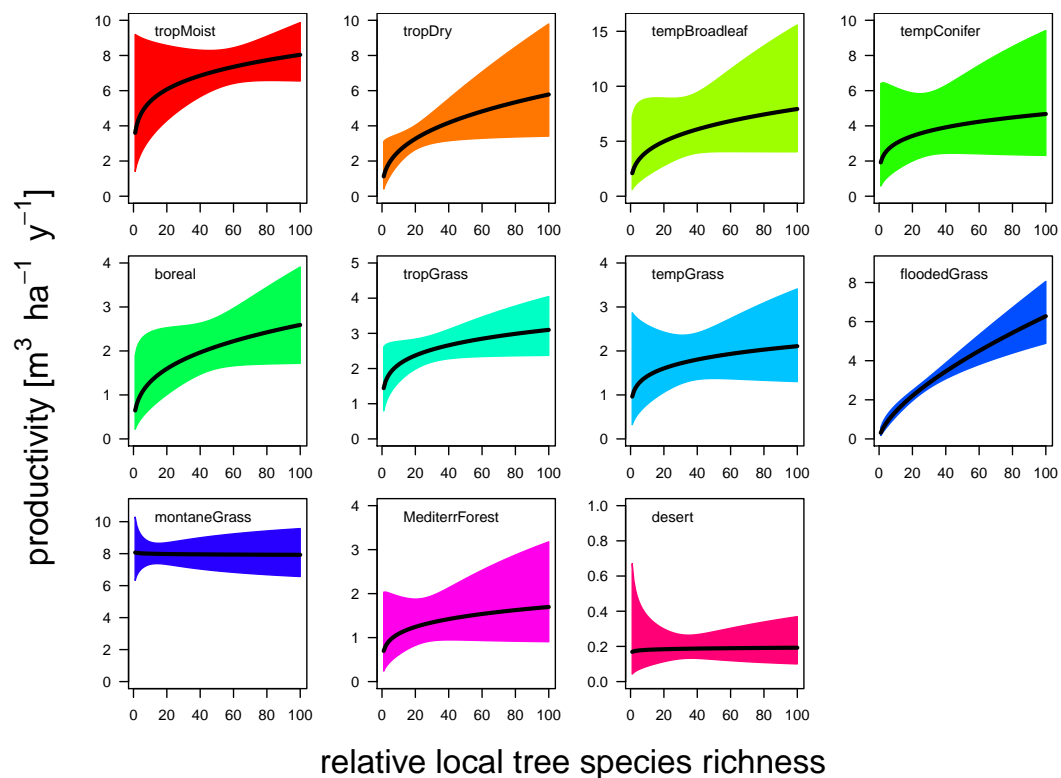
¹⁷The marginal effect asks the question: “I’m in a forest somewhere on this planet. I know that this plot has 50% of its potential tree species richness. How productive is this plot?” In other words, the marginal effect is ignorant of the values of the environmental covariates, and as a consequence is far less certain. The conditional effect (“I’m standing in a forest patch in north Sumatra, with that much rainfall, tree density and so forth, and tree species richness is at 50% of its maximum potential: How productive is this site?”) obviously requires more information, but is thereby, by definition, not global anymore.

```
# fit new model:
coefMatNew <- t(replicate(500, repeatedGLSbiome(N = 500, data = thisEcoregion)))
regionalFits[[i]] <- coefMatNew
cat(i, " ")
rm(coefMatNew)
}
```

```
1 2 4 5 6 7 8 9 10 12 13
```

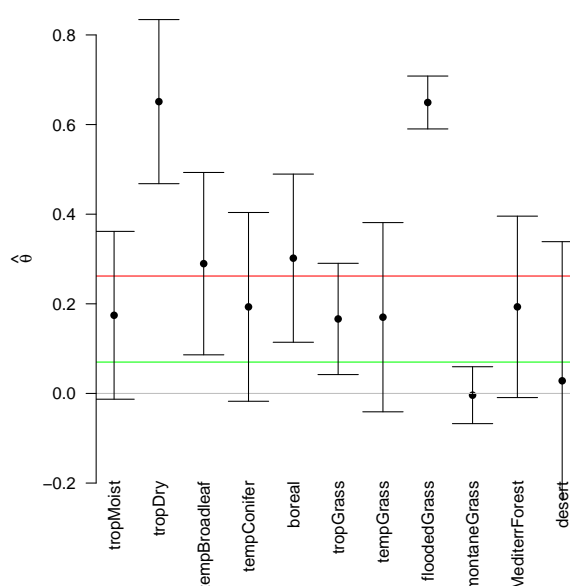
```
thetamat <- sapply(regionalFits, function(x) c(mean(x[, 6], na.rm = T),
sd(x[, 6], na.rm = T)))
```

```
sampleSize <- table(bpr_data$Ecoregion)
par(mar=c(3,3,0,1), mfrow=c(3,4), oma=c(4,4,0,0))
for (i in c(1,2,4,10,12,13)){
  XReg <- as.matrix(data.frame("Intercept"=1, "SbrevLocal"=log(seq(0.01, 1, by=0.01)),
    t(colMeans(bpr_data[bpr_data$Ecoregion == i, c("G", "T3", "C1", "C3", "PET",
      "IAA", "E")]))))
  if (i %in% c(2, 9)){
    # Ecoregion 2 has only 111 plots, 9: 547; here we use the full model!
    subSet <- bpr_data[bpr_data$Ecoregion == i, -c(2,6)]
    # apparently predictSE cannot handle the log() in a formula:
    subSet$logSbrevLocal <- log(subSet$SbrevLocal)
    fglsReg <- nlme::glsl(log(P) ~ logSbrevLocal + G + T3 + C1 + C3 + PET + IAA + E,
      correlation=corHaversine(form=~lon+lat, mimic="corSpher"),
      data=subSet)
    thetamat[,2] <- c(coef(fglsReg)["logSbrevLocal"],
      sqrt(diag(summary(fglsReg)$varBeta))["logSbrevLocal"] )
    # somehow knitr didn't like this line:
    #XReg <- as.matrix(data.frame("logSbrevLocal"=log(seq(0.01, 1, by=0.01)),
    #  t(colMeans(subSet)))
    XReg <- as.matrix(data.frame("logSbrevLocal"=log(seq(0.01, 1, by=0.01)),
      "G"=mean(subSet$G), "T3"=mean(subSet$T3), "C1"=mean(subSet$C1),
      "C3"=mean(subSet$C3), "PET"=mean(subSet$PET),
      "IAA"=mean(subSet$IAA), "E"=mean(subSet$E)))
    predsReg <- AICcmmodavg::predictSE(fglsReg, newdata=XReg, se.fit=T)
    predsRegMeans <- predsReg$fit
    CIsdcorrReg <- predsReg$se.fit # no need to correct
  } else {
    betaReg <- regionalFits[[i]][, 5:13]
    predsReg <- XReg %*% t(betaReg)
    predsRegMeans <- rowMeans(predsReg, na.rm=T)
    CIsdReg <- apply(predsReg, 1, sd, na.rm=T)
    CIsdcorrReg <- exp(log(500) / log(sampleSize[names(sampleSize) == i]) * log(CIsdReg))
  }
  CIsNewReg <- rbind(predsRegMeans + 2 * CIsdcorrReg, predsRegMeans - 2 * CIsdcorrReg)
  ylimUpper <- ceiling(exp(max(CIsNewReg)))
  plot(1:100, exp(predsRegMeans), type="n", las=1, ylim=c(0, ylimUpper),
    #xlab="relative local species richness [%]", ylab="productivity", log="",
    main="", xlab="", ylab="")
  polygon(c(1:100, 100:1), c(exp(CIsNewReg[1,]), rev(exp(CIsNewReg[2,]))),
    border=cols[i], col=cols[i])
  lines(1:100, exp(predsRegMeans), lwd=3)
  legend("topleft", legend=biome_trees$biomeName[biome_trees$Ecoregion == i],
    bty="n", cex=1)
}
mtext(side=1, text="relative local tree species richness", cex=1.5, line=1, outer=T)
mtext(side=2, text=expression(paste("productivity [", m^{"3"} ~ ha^{"-1"} ~ y^{"-1"}, "]")),
  cex=1.5, line=1, outer=T, las=0)
```



As a summary, we can plot the estimates for each ecoregion.

```
thetamat <- as.data.frame(thetamat[, -c(3, 11)])
colnames(thetamat) <- as.character(biome_trees$biomeName[-c(10)]) # 'tundra' excluded
rownames(thetamat) <- c("theta", "se")
par(mar = c(5, 5, 1, 1), las = 1)
plot(1:11, thetamat[1, ], pch = 16, ylim = c(-0.2, 0.8), ylab = expression(hat(theta)),
     xlab = "", axes = F)
abline(h = 0, col = "grey")
abline(h = 0.07, col = "green")
abline(h = 0.262, col = "red")
arrows(1:11, unlist(thetamat[1, ] - 2 * thetamat[2, ]), 1:11, unlist(thetamat[1,
     ] + 2 * thetamat[2, ]), angle = 90, code = 3, length = 0.2)
axis(2)
axis(1, tick = F, at = 1:11, labels = colnames(thetamat), las = 3, line = -2)
```



In six out of 11 cases (tropical moist rainforest, temperate coniferous forest, temperate grassland, mountain grassland, Mediterranean forest and desert) the estimates are not significant. All others point at positive effects of relative local species richness of rather different strength (very strong in tropical dry rainforests and flooded grasslands). The original global model of Liang *et al.* (2016) (in

red) shows a substantially stronger effect than our re-analysis (green).

E.1 Analysis with actual species richness as predictor

One of the interpretational problems of the above analyses, and indeed the key neat idea of the original paper, is that a change from, say, 10 to 30% on the x-axis means completely different things in different ecoregions. In the tropics, this may be a change in species richness from 30 to 90 species, while in the boreal system from “less than one” to “just under two”.

As a final step, we re-analyse the data with the same model as in the previous section, just with $\log(S)$ as predictor, rather than a somehow corrected species richness.

```
repeatedGLSbiomeS <- function(N = 500, data = NULL) {
  # function running random subset GLS take subset
  subset <- data[sample(nrow(data), min(N, nrow(data))), ]
  # fit model
  fgls <- try(gls(log(P) ~ log(S) + G + T3 + C1 + C3 + PET + IAA + E,
    correlation = corHaversine(form = ~lon + lat, mimic = "corSpher"),
    data = subset))
  if (inherits(fgls, "try-error")) {
    return(rep(NA, 13))
  } else {
    r2 <- cor(predict(fgls), log(subset$P))^2 # R2
    return(c(ell = logLik(fgls), AIC = AIC(fgls), BIC = BIC(fgls),
      R2 = r2, coef(fgls)))
  }
}

regionalFitsS <- list()
set.seed(123)
for (i in c(1, 2, 4:10, 12, 13)) {
  # 3: no data; 11: only 9 data points select region:
  thisEcoregion <- bpr_data[bpr_data$Ecoregion == i, ]
  # fit new model:
  coefMatNewS <- t(replicate(50, repeatedGLSbiomeS(N = 500, data = thisEcoregion)))
  regionalFits[[i]] <- coefMatNewS
  cat(i, " ")
  rm(coefMatNewS)
}
```

```
1 2 4 5 6 7 8 9 10 12 13
```

```
thetamatS <- sapply(regionalFits, function(x) c(mean(x[, 6], na.rm = T),
  sd(x[, 6], na.rm = T)))
thetamatS
```

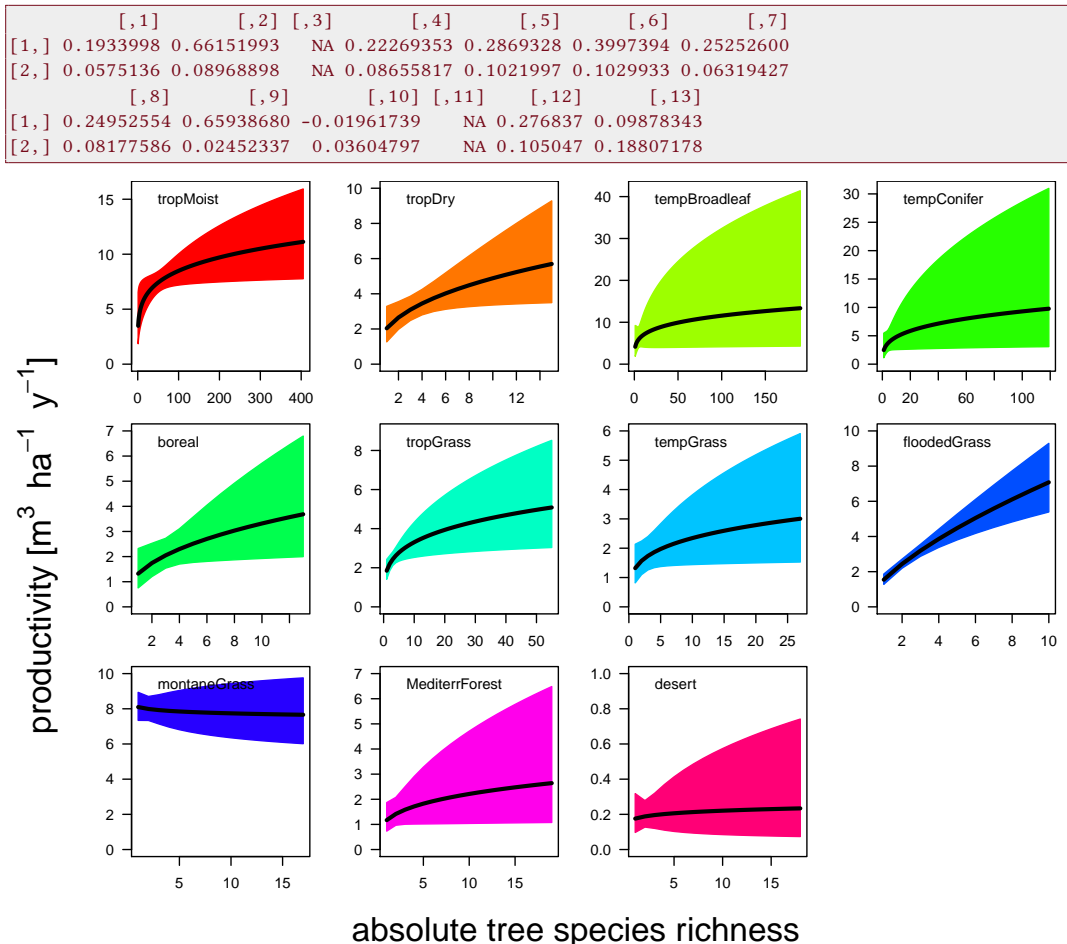
	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	0.1933998	3.795883e-01	NA	0.22269353	0.2869328	0.3997394	0.25252600
[2,]	0.0575136	3.520413e-12	NA	0.08655817	0.1021997	0.1029933	0.06319427
	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	
[1,]	0.24952554	0.65938680	-0.01961739	NA	0.276837	0.09878343	
[2,]	0.08177586	0.02452337	0.03604797	NA	0.105047	0.18807178	

```
sampleSize <- table(bpr_data$Ecoregion)
par(mar=c(3,3,0,1), mfrow=c(3,4), oma=c(4,4,0,0))
for (i in c(1,2,4:10,12,13)){
  XReg <- as.matrix(data.frame("Intercept"=1,
    "S"=log(seq(1, max(bpr_data$S[bpr_data$Ecoregion == i]), by=1)),
    t(colMeans(bpr_data[bpr_data$Ecoregion == i, c("G", "T3", "C1",
      "C3", "PET", "IAA", "E")]))))
  if (i %in% c(2, 9)){
    # Ecoregion 2 has only 111 plots, 9: 547; here we use the full model!
    subSet <- bpr_data[bpr_data$Ecoregion == i, -c(2,6)]
    # apparently predictSE cannot handle the log() in a formula:
    subSet$logS <- log(subSet$S)
    fglsReg <- nlme::glms(log(P) ~ logS + G + T3 + C1 + C3 + PET + IAA + E,
      correlation=corHaversine(form=~lon+lat, mimic="corSpher"),
      data=subSet)
    thetamatS[,2] <- c(coef(fglsReg)["logS"], sqrt(diag(summary(fglsReg)$varBeta))["logS"])
    XReg <- as.matrix(data.frame("logS"=log(seq(1, max(subSet$S), by=1)),
      "G"=mean(subSet$G), "T3"=mean(subSet$T3), "C1"=mean(subSet$C1),
      "C3"=mean(subSet$C3), "PET"=mean(subSet$PET),
      "IAA"=mean(subSet$IAA), "E"=mean(subSet$E)))
    predsRegS <- AICcmmodavg::predictSE(fglsReg, newdata=XReg, se.fit=T)
```

```

predsRegMeansS <- predsRegS$fit
CIsdcorrRegS <- predsRegS$se.fit # no need to correct
Smax <- max(subSet$S)
} else {
  Smax <- max(bpr_data$S[bpr_data$Ecoregion == i])
  betaReg <- regionalFits[[i]][, 5:13]
  predsRegS <- XReg %>% t(betaReg)
  predsRegMeansS <- rowMeans(predsRegS, na.rm=T)
  CIsdRegS <- apply(predsRegS, 1, sd, na.rm=T)
  CIsdcorrRegS <- exp(log(500) / log(sampleSize[names(sampleSize) == i]) * log(CIsdRegS))
}
CIsNewRegS <- rbind(predsRegMeansS + 2 * CIsdcorrRegS, predsRegMeansS - 2 * CIsdcorrRegS)
ylimUpper <- ceiling(exp(max(CIsNewRegS)))
plot(1:Smax, exp(predsRegMeansS), type="n", las=1, ylim=c(0, ylimUpper),
     main="", xlab="", ylab="")
polygon(c(1:Smax, Smax:1), c(exp(CIsNewRegS[1,]), rev(exp(CIsNewRegS[2,]))),
        border=cols[i], col=cols[i])
lines(1:Smax, exp(predsRegMeansS), lwd=3)
legend("topleft", legend=biome_trees$biomeName[biome_trees$Ecoregion == i], bty="n", cex=1)
}
mtext(side=1, text="absolute tree species richness", cex=1.5, line=1, outer=T)
mtext(side=2, text=expression(paste("productivity [", m^3, " ha^-1 ~ ha^-1 ~ y^-1", " ]")),
      cex=1.5, line=1, outer=T, las=0)
thetamatS

```

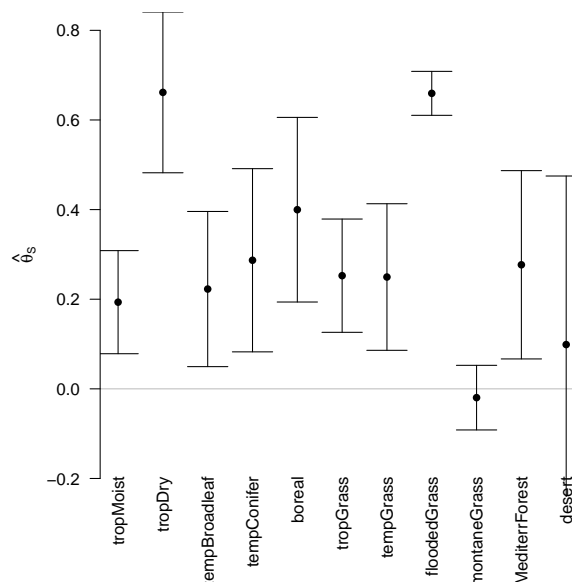


As a summary, we can plot the estimates and their 95%-confidence interval for each ecoregion.

```

thetamatS <- as.data.frame(thetamatS[, -c(3, 11)])
colnames(thetamatS) <- as.character(biome_trees$biomeName[-c(10)]) # 'tundra' excluded
rownames(thetamatS) <- c("theta", "se")
par(mar = c(5, 5, 1, 1), las = 1)
plot(1:11, thetamatS[1, ], pch = 16, ylim = c(-0.2, 0.8), ylab = expression(hat(theta)[S]),
     xlab = "", axes = F)
abline(h = 0, col = "grey")
arrows(1:11, unlist(thetamatS[1, ] - 2 * thetamatS[2, ]), 1:11, unlist(thetamatS[1, ]
+ 2 * thetamatS[2, ]), angle = 90, code = 3, length = 0.2)
axis(2)
axis(1, tick = F, at = 1:11, labels = colnames(thetamatS), las = 3, line = -2)

```

We see that in almost all ecoregions higher tree species richness is correlated with higher productivity. We also see the common pattern of a quick decrease of a distinguishable diversity effect beyond a few species (in tropical moist forests, “few” in this case means a few dozen, while in all other systems “few” means less than 10).¹⁸

F Conclusion

The “global” relationship between “relative tree species richness” and productivity reported by Liang *et al.* (2016) hinges on an analysis with several logical and technical flaws. Removing them yields essentially a flat line and no such global relationship. It is entirely possible that there are mistakes in the above analysis, and that is why we provide this document.¹⁹

This relationship is *not* globally consistent; rather, ecosystems vary substantially, but most show a positive tree-species richness-productivity relationship. Analysing the data with a relative species richness axis provides no benefit or insight when compared to absolute species richness analyses.

Bibliography

- Ammer, C. (2019) Diversity and forest productivity in a changing climate. *New Phytologist*, **221**, 50–66.
- Breiman, L. (2001) Random forests. *Machine Learning Journal*, **45**, 5–32.
- Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S. & Naeem, S. (2012) Biodiversity loss and its impact on humanity. *Nature*, **486**, 59–67.
- Cardinale, B.J., Srivastava, D., Duffy, J.E., Wright, J.P., Downing, A.L. & Sankaran, M. (2006) Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature*, **443**, 989–992.
- Crowther, T.W., Glick, H.B., Covey, K.R., Bettigole, C., Maynard, D.S., Thomas, S.M., Smith, J.R., Hintler, G., Duguid, M.C., Amatulli, G., Tuanmu, M.N., Jetz, W., Salas, C., Stam, C., Piotto, D., Tavani, R., Green, S., Bruce, G., Williams, S.J., Wiser, S.K., Huber, M.O., Hengeveld, G.M., Nabuurs, G.J., Tikhonova, E., Borchardt, P., Li, C.F., Powrie, L.W., Fischer, M., Hemp, A., Homeier, J., Cho, P., Vibrans, A.C., Umunay, P.M., Piao, S.L., Rowe, C.W., Ashton, M.S., Crane, P.R. & Bradford, M.A. (2015) Mapping tree density at a global scale. *Nature*, **525**, 201–205.
- Huang, Y., Chen, Y., Castro-Izaguirre, N., Baruffol, M., Brezzi, M., Lang, A., Li, Y., Härdtle, W., Oheimb, G.v., Yang, X. & et al. (2018) Impacts of species richness on productivity in a large-scale subtropical forest experiment. *Science*, **362**, 80–83.

¹⁸These curves are not identical to those from the previous analysis with relative local species richness, as the re-scaling of the x-axis is non-linear: species richness enters the model after log-transformation, thereby also affecting estimation of θ .

¹⁹Note that some steps require substantial computing time (all those running a model often).

- Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W.S., Reich, P.B., Scherer-Lorenzen, M., Schmid, B., Tilman, D., van Ruijven, J., Weigelt, A., Wilsey, B.J., Zavaleta, E.S. & Loreau, M. (2011) High plant diversity is needed to maintain ecosystem services. *Nature*, **477**, 199–202.
- Jactel, H., Gritti, E.S., Drössler, L., Forrester, D.I., Mason, W.L., Morin, X., Pretzsch, H. & Castagneyrol, B. (2018) Positive biodiversity–productivity relationships in forests: climate matters. *Biology Letters*, **14**, 20170747.
- Lele, S.R., Nadeem, K. & Schmuland, B. (2010) Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, **105**, 1617–1625.
- Liang, J., Crowther, T.W., Picard, N., Wiser, S., Zhou, M., Alberti, G., Schulze, E.D., McGuire, D., Bozzato, F., Pretzsch, H., De-Miguel, S., Paquette, A., Hérault, B., Scherer-Lorenzen, M., Barrett, C.B., Glick, H.B., Hengeveld, G.M., Nabuurs, G.J., Pfautsch, S., Viana, H., Vibrans, A.C., Ammer, C., Schall, P., Verbyla, D.L., Tchebakova, N., Fischer, M., Watson, J.V., Chen, H.Y., Lei, X., Schelhaas, M.J., Lu, H., Gianelle, D., Parfenova, E.I., Salas, C., Lee, E., Lee, B., Kim, H.S., Bruehlheide, H., Coomes, D.A., Piotta, D., Sunderland, T., Schmid, B., Gourlet-Fleury, S., Sonké, B., Tavani, R., Zhu, J., Brandl, S., Vayreda, J., Kitahara, F., Searle, E.B., Neldner, V.J., Ngugi, M.R., Baraloto, C., Frizzera, L., Bałazy, R., Oleksyn, J., Zawila-Niedzwiecki, T., Bouriaud, O., Bussotti, F., Finér, L., Jaroszewicz, B., Jucker, T., Valladares, F., Jagodzinski, A.M., Peri, P.L., Gonmadje, C., Marthy, W., O'Brien, T., Martin, E.H., Marshall, A., Rovero, F., Bitariho, R., Niklaus, P.A., Alvarez-Loayza, P., Chamuya, N., Valencia, R., Mortier, F., Wortel, V., Engone-Obiang, N.L., Ferreira, L.V., Odeke, D.E., Vasquez, R.M. & Reich, P.B. (2016) Positive biodiversity–productivity relationship predominant in global forests. *Science*, **354**, 196.
- Paquette, A. & Messier, C. (2010) The effect of biodiversity on tree productivity: from temperate to boreal forests. *Global Ecology and Biogeography*, **20**, 170–180.
- Pinheiro, J.C. & Bates, D.M. (2000) *Mixed-Effects Models in S and S-Plus*. Springer, Berlin. ISBN 0-387-98957-0.
- Politis, D.N., Romano, J.P. & Wolf, M. (1999) *Subsampling*. Springer, Berlin.
- Schulze, E.D., Bouriaud, O., Weber, U., Roscher, C., Hessenmoeller, D., Kroiher, F. & Schall, P. (2018) Management breaks the natural productivity-biodiversity relationship in forests and grassland: an opinion. *Forest Ecosystems*, **5**, 3.
- Stape, J.L., Binkley, D., Ryan, M.G., Fonseca, S., Loos, R.A., Takahashi, E.N., Silva, C.R., Silva, S.R., Hakamada, R.E., Ferreira, J.M.d.A., Lima, A.M.N., Gava, J.L., Leite, F.P., Andrade, H.B., Alves, J.M., Silva, G.G.C. & Azevedo, M.R. (2010) The Brazil Eucalyptus Potential Productivity Project: Influence of water, nutrients and stand uniformity on wood production. *Forest Ecology and Management*, **259**, 1684–1694.
- Vilà, M., Vayreda, J., Comas, L., Ibáñez, J.J., Mata, T. & Obón, B. (2007) Species richness and wood production: A positive association in Mediterranean forests. *Ecology Letters*, **10**, 241–250.
- Vilà, M., Carrillo-Gavilán, A., Vayreda, J., Bugmann, H., Fridman, J., Grodzki, W., Haase, J., Kunstler, G., Schelhaas, M. & Trasobares, A. (2013) Disentangling biodiversity and climatic determinants of wood production. *PLoS ONE*, **8**, e53530.
- Whittaker, R.J. (2010) Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness-productivity relationship. *Ecology*, **91**, 2522–2533.
- Whittaker, R.J. & Heegaard, E. (2003) What is the observed relationship between species richness and productivity? (Comment). *Ecology*, **84**, 3384–3390.
- Zhang, Y., Chen, H.Y.H. & Reich, P.B. (2012a) Forest productivity increases with evenness, species richness and trait variation: a global meta-analysis. *Journal of Ecology*, **100**, 742–749.
- Zhang, Y., Chen, H.Y.H. & Reich, P.B. (2012b) Forest productivity increases with evenness, species richness and trait variation: A global meta-analysis. *Journal of Ecology*, **100**, 742–749.