How many tree species in French Guiana Tropical Moist Forest?

05 January 2022

The biodiversity of tropical rainforest is difficult to assess. Yet, its estimation is necessary for conservation purposes, to evaluate our level of knowledge and the risks faced by the forest in relation to global change. Our contribution is to estimate the regional richness of tree species from local but widely spread inventories. Guyadiv is a network of forest plots installed over the whole territory of French Guiana, where trees over 10 cm DBH are identified. We use its information (1180 species censused in 76 one-hectare plots) to estimate the exponent of the species-area relationship, assuming Arrhenius’s power law. We can then extrapolate the number of species from a local, wide inventory (62.5 ha in Paracou research station). We evaluate the number of tree species around 2000 over the territory.

# 1 Introduction

Biodiversity assessment in tropical moist forests is a practical challenge but a major goal considering they are the most diverse terrestrial ecosystems. Estimating the number of tree species is made possible by the long-term effort of sampling resulting in thousands of forest plots organized in various networks. In French Guiana, the Guyadiv network consists of xxx plots accross the whole forest. Based on similar datasets, the diversity of tree species has been estimated in Amazonia (ter Steege et al. 2013; ter Steege et al. 2020) and at the world scale (Slik et al. 2015). The methods used in these studies are not appropriate to estimate regional diversity, i.e. at a smaller scale where dispersal limitation is critical. The contibution of this paper is to estimate the number of tree species at the regional scale, in French Guiana (8 million hectares of tropical moist forest with no ecological boundary to distinguish them from the rest of Amazonia) and demonstrate which method is valid to do so. We build on Harte’s self similarity model (Harte, Kinzig, and Green 1999) that implies the power-law relatioship of Arrhenius (1921) and provides a technique to evaluate its parameters (Harte et al. 1999), previously applied by Krishnamani, Kumar, and Harte (2004) in the Western Ghats, India. We show that the log-series model underlying the work of ter Steege et al. (2013) does not apply at the regional scale.

# 2 Methods

Self-similarity (Harte, Kinzig, and Green 1999; Harte et al. 1999).

Estimation of the richness of an Indian forest (Krishnamani, Kumar, and Harte 2004).

## 2.1 Data

# Decrypt the vault  
library("secret") %>%  
 suppressMessages()  
name\_project <- "GF-Richness"  
vault <- "vault"  
Sys.setenv(USER\_KEY = usethis::proj\_path(paste0(name\_project,  
 ".rsa")))  
Plots <- get\_secret("Plots", vault = vault)  
Abundances <- get\_secret("Abundances", vault = vault)

# 3 Results

## 3.1 Plots

## 3.2 Random plot

One plot per location

# One plot per location...  
Plots %>%  
 mutate(Random = runif(n())) -> RandomizedPlots  
RandomizedPlots %>%  
 group\_by(Location) %>%  
 summarize(MaxRandom = max(Random)) %>%  
 rename(Random = MaxRandom) %>%  
 inner\_join(RandomizedPlots) %>%  
 select(Plot) -> SelectedPlots

## Joining, by = c("Location", "Random")

# ...or all plots SelectedPlots <- Plots['Plot']  
# Distances  
library("dbmss") %>%  
 suppressMessages()  
Plots %>%  
 inner\_join(SelectedPlots) %>%  
 rename(PointName = Plot, X = X\_UTM, Y = Y\_UTM,  
 PointType = Location) %>%  
 mutate(PointWeight = 1) %>%  
 wmppp(unitname = c("meter", "meters")) %>%  
 pairdist() %>%  
 as.dist -> Distances

## Joining, by = "Plot"

## 3.3 Sorensen

library("ade4") %>%  
 suppressMessages()  
Abundances %>%  
 inner\_join(SelectedPlots) %>%  
 select(-Plot) %>%  
 dist.binary(method = 5) -> Sorensen

## Joining, by = "Plot"

Sorensen <- 1 - Sorensen

## 3.4 Relation

tibble(Sorensen = as.numeric(log10(Sorensen)), Distance = as.numeric(log10(Distances))) %>%  
 filter(Distance > 3) -> dist\_plots  
lm(Sorensen ~ Distance, data = dist\_plots) -> regression  
regression %>%  
 summary()

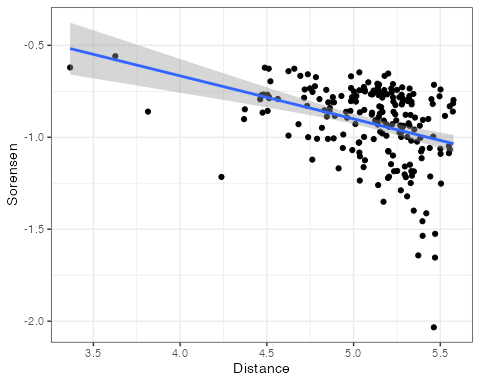
##   
## Call:  
## lm(formula = Sorensen ~ Distance, data = dist\_plots)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.02518 -0.09879 0.02880 0.15867 0.29373   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.27000 0.20849 1.295 0.197  
## Distance -0.23400 0.04088 -5.723 3.63e-08  
##   
## (Intercept)   
## Distance \*\*\*  
## ---  
## Signif. codes:   
## 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1985 on 207 degrees of freedom  
## Multiple R-squared: 0.1366, Adjusted R-squared: 0.1325   
## F-statistic: 32.76 on 1 and 207 DF, p-value: 3.63e-08

(-regression$coefficients[2]/2 -> z)

## Distance   
## 0.1170003

dist\_plots %>%  
 ggplot(aes(x = Distance, y = Sorensen)) + geom\_point() +  
 geom\_smooth(method = lm)

## `geom\_smooth()` using formula 'y ~ x'

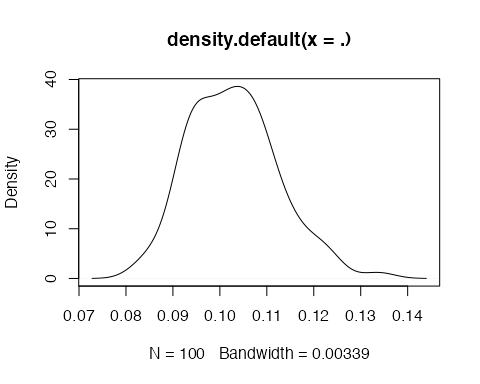


## Bootstrap

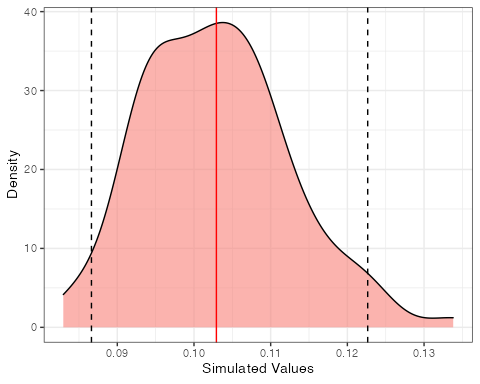
library("broom")  
estimate\_z <- function() {  
 # Select one plot per location...  
 Plots %>%  
 mutate(Random = runif(n())) -> RandomizedPlots  
 RandomizedPlots %>%  
 group\_by(Location) %>%  
 summarize(MaxRandom = max(Random)) %>%  
 rename(Random = MaxRandom) %>%  
 inner\_join(RandomizedPlots) %>%  
 suppressMessages %>%  
 select(Plot) -> SelectedPlots  
 # Calculate distances  
 Plots %>%  
 inner\_join(SelectedPlots) %>%  
 suppressMessages %>%  
 rename(PointName = Plot, X = X\_UTM, Y = Y\_UTM,  
 PointType = Location) %>%  
 mutate(PointWeight = 1) %>%  
 wmppp(unitname = c("meter", "meters")) %>%  
 suppressWarnings %>%  
 pairdist() %>%  
 as.dist -> Distances  
 # Calculate Sorensen divergence  
 Abundances %>%  
 inner\_join(SelectedPlots) %>%  
 suppressMessages %>%  
 select(-Plot) %>%  
 dist.binary(method = 5) -> Sorensen  
 Sorensen <- 1 - Sorensen  
 # Regress  
 tibble(Sorensen = as.numeric(log10(Sorensen)),  
 Distance = as.numeric(log10(Distances))) %>%  
 dplyr::filter(Distance > 3) %>%  
 lm(Sorensen ~ Distance, data = .) %>%  
 tidy %>%  
 dplyr::filter(term == "Distance") %>%  
 select(estimate) %>%  
 pull -> z  
 z <- -z/2  
 return(z)  
}  
n\_simulations <- 100  
pgb <- txtProgressBar(min = 0, max = n\_simulations)  
sim\_z <- rep(0, n\_simulations)  
for (i in 1:n\_simulations) {  
 sim\_z[i] <- estimate\_z()  
 setTxtProgressBar(pgb, i)  
}

## ==================================================

sim\_z %>%  
 density %>%  
 plot



entropart::as.SimTest(mean(sim\_z), sim\_z) %>%  
 autoplot



## 3.5 Extrapolation

# Extrapolation de Paracou : .625km², 604 sp >  
# c=631  
(c <- 600/0.625^z)

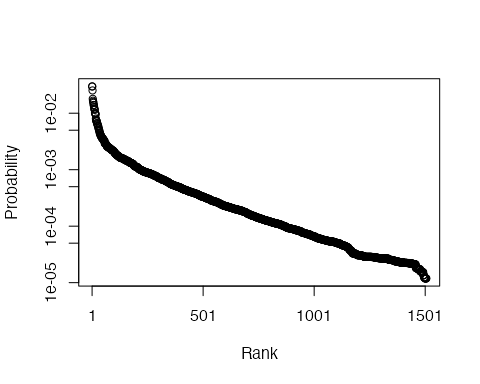
## Distance   
## 633.9184

# Nombre d'espèces total  
(spsim <- c \* 80000^z)

## Distance   
## 2375.18

## 3.6 Fisher

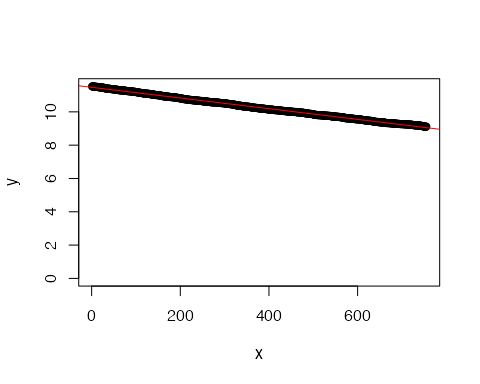
# Nombre d'arbres  
Ntrees <- 8 \* 10^6 \* mean(colSums(Abundances[, -1]))  
# Taux de couverture par placette  
library("entropart") %>%  
 suppressMessages()  
C <- apply(Abundances[, -1], 1, function(X) Coverage(X))  
# Distribution de probabilités, corrigée par les  
# taux de couverture  
ObsProba <- C %\*% as.matrix(Abundances[, -1]/rowSums(Abundances[,  
 -1]))/nrow(Abundances)  
plot(as.ProbaVector(ObsProba))



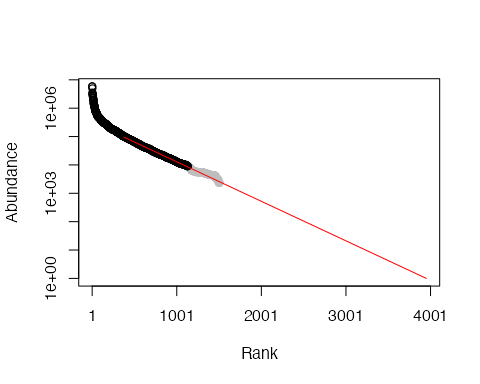
# Abondances  
EstimatedAbd <- as.AbdVector(round(ObsProba \* Ntrees,  
 0))  
# 50% centraux  
EA50 <- as.AbdVector(sort(EstimatedAbd, decreasing = T)[round(length(EstimatedAbd)/4):round(length(EstimatedAbd) \*  
 3/4)])  
x <- 1:length(EA50)  
y <- as.numeric(log(EA50))  
# Régression rang-abondance sur la partie  
# rectiligne de la courbe  
EA50lm <- lm(y ~ x)  
# Nombre d'espèces extrapolé  
NbSpecies <- round(-EA50lm$coefficients[1]/EA50lm$coefficients[2]) +  
 round(length(EstimatedAbd)/4) - 1  
# Régression de la partie rectiligne  
plot(y ~ x, ylim = c(0, max(y)))  
summary(EA50lm)

##   
## Call:  
## lm(formula = y ~ x)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.050607 -0.026011 -0.004327 0.017631 0.084264   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 1.147e+01 2.303e-03 4981.6  
## x -3.208e-03 5.278e-06 -607.8  
## Pr(>|t|)   
## (Intercept) <2e-16 \*\*\*  
## x <2e-16 \*\*\*  
## ---  
## Signif. codes:   
## 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.03161 on 753 degrees of freedom  
## Multiple R-squared: 0.998, Adjusted R-squared: 0.998   
## F-statistic: 3.694e+05 on 1 and 753 DF, p-value: < 2.2e-16

abline(a = EA50lm$coefficients[1], b = EA50lm$coefficients[2],  
 col = "red")



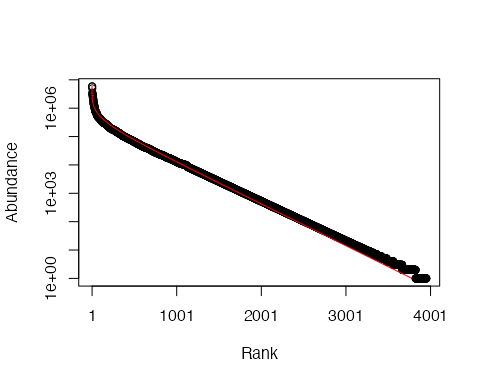
# Metacommunauté complète: partie observée  
EA75 <- sort(EstimatedAbd, decreasing = T)[1:round(length(EstimatedAbd) \*  
 3/4)]  
plot(EstimatedAbd, ylim = c(1, max(EA75)), xlim = c(0,  
 NbSpecies), col = "grey")  
points(EA75, ylim = c(1, max(EA75)), xlim = c(0, -EA50lm$coefficients[1]/EA50lm$coefficients[2]))  
Extra <- exp(EA50lm$coefficients[1]) \* cumprod(rep(exp(EA50lm$coefficients[2]),  
 NbSpecies - round(length(EstimatedAbd)/4) + 1))  
lines(x = round(length(EstimatedAbd)/4):NbSpecies,  
 y = Extra, col = "red")



# Vérification : effectif total  
(Ntotal <- sum(EA75) + sum(Extra[(length(EA75) - NbSpecies +  
 length(Extra) + 1):length(Extra)]))

## [1] 197186462

# Ajustement d'une log-série  
plot(as.AbdVector(round(c(EA75, Extra[(length(EA75) -  
 NbSpecies + length(Extra) + 1):length(Extra)]))),  
 Distribution = "lseries")



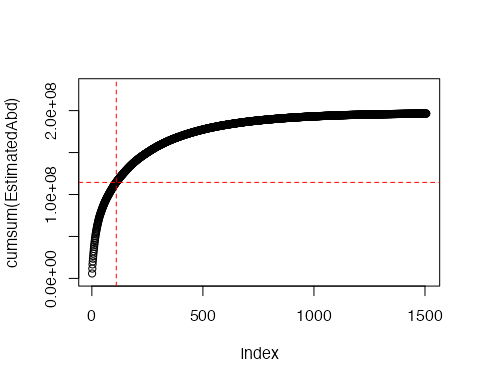
## $alpha  
## [1] 294.5374

# Nombre d'espèces estimées  
NbSpecies

## (Intercept)   
## 3951

## 3.7 Hyperdominance

# Hyperdominance  
EstimatedAbd <- sort(round(ObsProba \* Ntrees, 0), decreasing = TRUE)  
plot(cumsum(EstimatedAbd), ylim = c(0, Ntrees))  
abline(h = Ntrees/2, lty = 2, col = "red")  
hdSpecies <- min(which(cumsum(EstimatedAbd) > Ntrees/2))  
abline(v = hdSpecies, lty = 2, col = "red")



# References

Arrhenius, Olof. 1921. “Species and Area.” *Journal of Ecology* 9 (1): 95–99. <https://doi.org/10.2307/2255763>.

Harte, John, Ann Kinzig, and Jessica Green. 1999. “Self-Similarity in the Distribution and Abundance of Species.” *Science* 284 (5412): 334–36. <https://doi.org/10.1126/science.284.5412.334>.

Harte, John, Sarah Mccarthy, Kevin Taylor, Ann Kinzig, and Marc L. Fischer. 1999. “Estimating Species-Area Relationships from Scale Plot to Landscape Data Using Species Spatial-Turnover.” *Oikos* 86 (1): 45–54. <https://doi.org/10.2307/3546568>.

Krishnamani, R., A. Kumar, and John Harte. 2004. “Estimating Species Richness at Large Spatial Scales Using Data from Small Discrete Plots.” *Ecography* 27 (5): 637–42. <https://doi.org/10.1111/j.0906-7590.2004.03790.x>.

Slik, J. W. Ferry, Víctor Arroyo-Rodríguez, Shin-Ichiro Aiba, Patricia Alvarez-Loayza, Luciana F. Alves, Peter S. Ashton, Patricia Balvanera, et al. 2015. “An Estimate of the Number of Tropical Tree Species.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (24): 7472–77. <https://doi.org/10.1073/pnas.1423147112>.

Steege, Hans ter, Nigel C. A. Pitman, Daniel Sabatier, Christopher Baraloto, Rafael P. Salomão, Juan Ernesto Guevara, Oliver L. Phillips, et al. 2013. “Hyperdominance in the Amazonian Tree Flora.” *Science* 342 (6156): 1243092. <https://doi.org/10.1126/science.1243092>.

Steege, Hans ter, Paulo I. Prado, Renato A. F. de Lima, Edwin Pos, Luiz de Souza Coelho, Diogenes de Andrade Lima Filho, Rafael P. Salomão, et al. 2020. “Biased-Corrected Richness Estimates for the Amazonian Tree Flora.” *Scientific Reports* 10 (1): 10130. <https://doi.org/10.1038/s41598-020-66686-3>.