

Understanding Forest Structure

Philipp Burckhardt

August 12, 2014

1 Introduction

For 53 forest sites in Colombia and Ecuador, data was collected about the structure of the forest (Phillips, 2002). Besides the total number of plants, a breakdown of the composition of the plants was collected: It was recorded how many of them were ordinary *Trees*, and how many *Lianas* and *Hemiepiphytes* had been found. *Hemiepiphytes* are a type of epiphytes which grow upon other plants but develop roots over the course of their life, whereas *Lianas* grow from the ground but climb up on trees in order to reach the parts of the forest canopy which are well-lit (Britannica, 2006). There is an intense tree-liana competition, as studied e.g. by Ingwell et al. (2010) in a long-time study spanning 10 years. Research questions about the factors influencing the structure of forests are abundant in ecology, with one question centering on the proportion of lianas in tropical vs. temperate forests (Phillips, 2002). For example, it can be observed that dry forests generally have a higher proportion of lianas than forests with higher rainfall, see Swaine and Grace (2007) for a detailed study.

In analyzing our data set, we try to investigate whether the forest sites in Colombia and Ecuador conform to these laws or diverge from them. Specifically, we want to investigate how forest structure varies with the two variables *Rainfall* (in mm/year) and *Elevation* (in meters above sea level), which have been collected for each forest site. We start our analysis with an exploratory data analysis in Section 2. A multinomial logit model is proposed in Section 3 and the results from fitting this model are presented in Section 4. In Section 5, a Bayesian analysis is presented as an addition to the frequentist approach. The results are then compared with the previously obtained ones and the mixing of the chain is assessed. In Section 6, the results are summarized and potential for future research is discussed.

2 Exploratory Analysis

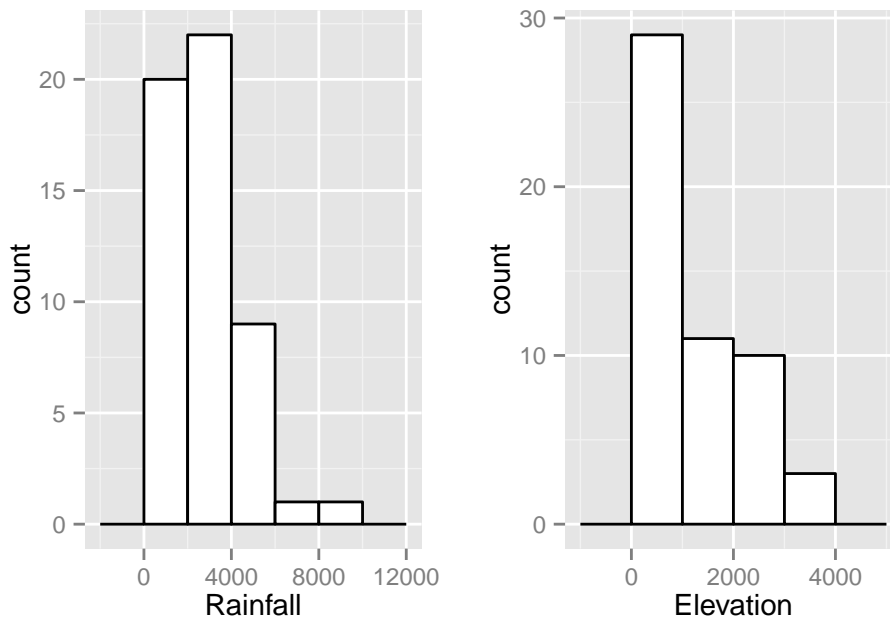


Figure 1: Histograms for explanatory variables *Rainfall* and *Elevation*

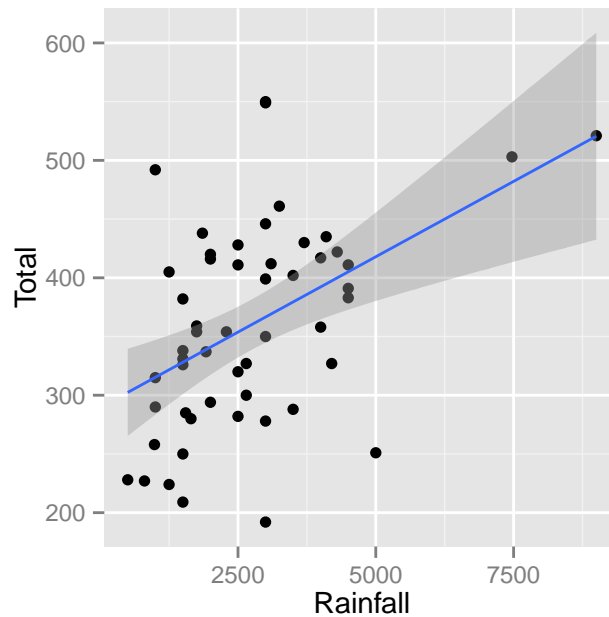


Figure 2: Scatterplot of the annual rainfall against the total number of plants

First, we examined whether the data set contained missing values or exhibited other properties one would need to deal with before conducting any statistical analysis. In general, the data set is well formatted, requiring only minor pre-processing from our part. One observation (row index 41) has missing values for both *Lianas* and *Hemiepiphytes*. It is obvious, however, that they must correspond to zero values, as the total number of plants and the amount of trees coincide for this observation. It was recoded accordingly.

Recall that for each forest site, the values for two possible factors which shape forest structure were recorded. In Figure 1, their empirical distributions are displayed as a histogram. As one can see, the respective forest sites vary considerably with respect to these characteristics. Most of them are fairly dry, but there are a few places with annual *Rainfall* higher than 6000mm, which could even be called rain-forests. *Elevation* is distributed in the range of 0 – 4000, without any extreme observations. In return, the distribution of *Elevation* is highly right-skewed, with most observations having values on the lower end of the spectrum.

Indeed, the considered forest sites differ both in terms of weather conditions and position from each other. But how do these different characteristics affect the flora of the forests? One quite obvious relationship is that more rainfall is generally associated with a larger amount of plants, as can be seen in the scatterplot of Figure 2: The superimposed regression line shows a clear linear trend. It must be noted that it would be even steeper if the two observations with very high rainfall did not exist.

Visualizing the relationships between the composition of the forest and the features of the forest site is a much more complicated task, though. In fact, one of the main reasons why one should favour a statistical modelling approach in this context is that it facilitates comprehension. Before turning to that task, a look at Figure 3 might notwithstanding clarify some patterns. First notice that the previously detected relationship between rainfall and the total number of trees is also visible here: The bars further down in the plot are for the most part shorter than the ones at the top. Also, we can see that the trees are the most common plant group. In fact, 80% of all counted plants

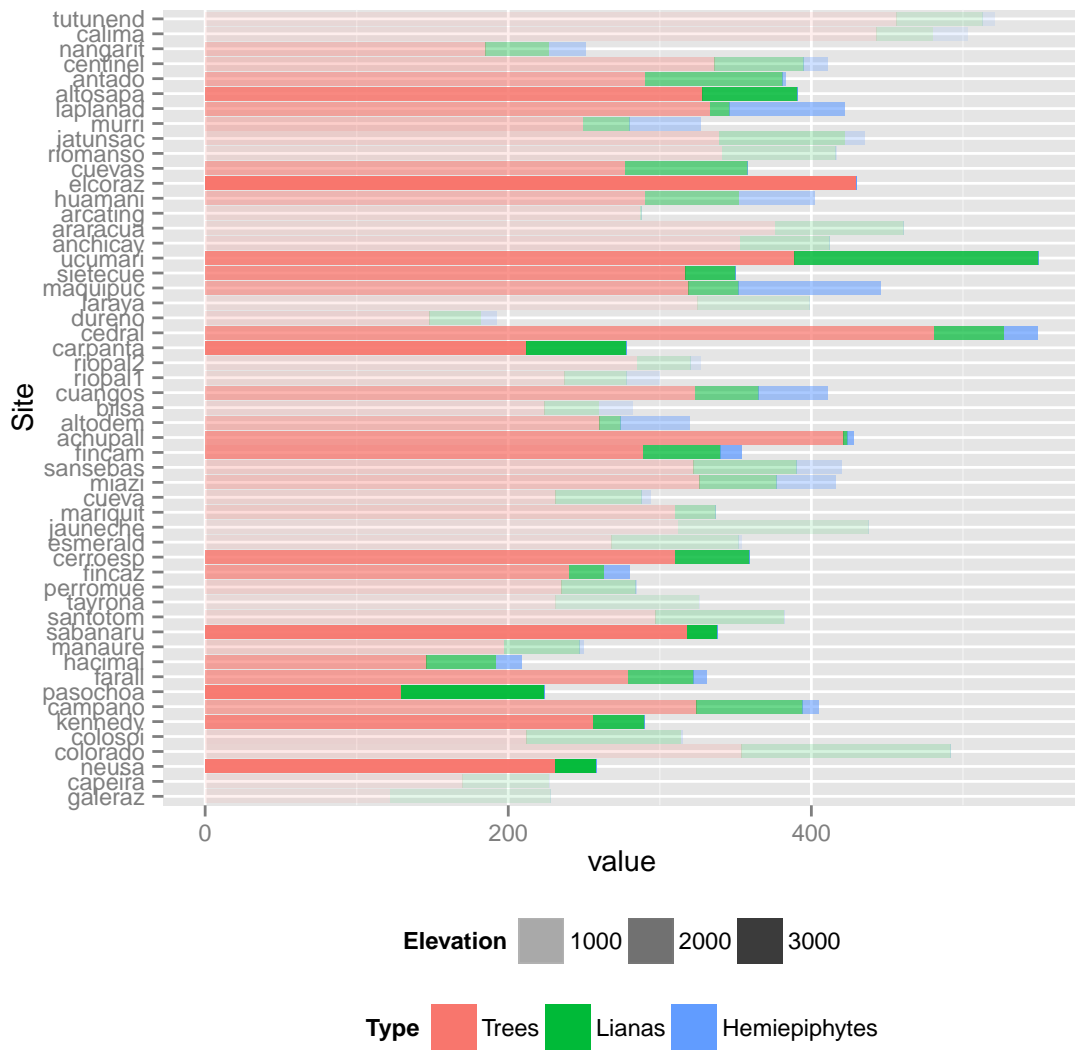


Figure 3: Plot of forest composition for all 53 forest sites. The length of the horizontal bars corresponds to the number of plants in the respective forest, with the colors denoting to which proportion this number is made up of *Trees*, *Lianas* and *Hemiepiphytes*. The forest sites have been sorted by their annual rainfall in descending order. The elevations of the forest sites are mapped to the alpha colors of the bars.

had been trees. The *Lianas* are the next common plant species, whereas *Hemiepiphytes* represent the rarest of the species (only 3.5% of the plants belong to this group). Another interesting pattern is that *Lianas* seem to compete with *Hemiepiphytes*: In most bars, there is either a larger blue or a larger green part. The combination is not encountered often. Concerning the impact of *Elevation*, the plot is not very insightful: There does not seem to be any difference in forest composition depending on how transparent the respective bar is. In the following, we investigate whether this impression is deceptive by applying statistical modelling.

3 Multinomial Regression

The logistic regression framework for binary outcomes can easily be extended to deal with response variables with more than two categories. These generalisations are then called multicategory or multinomial models. They assume that response variable Y follows a multinomial distribution, i.e.

$$Y_i \sim \text{multinomial}(\pi_i, N_i),$$

where N_i denotes the total number of counts for the i -th observation and π_i is the vector of class probabilities with length J , the number of different classes.

The most common function which is used to link π_i to the linear predictor is the *logit* link, which yields the multicategory logit model. Agresti (1990) points out that in this model, the log odds for all pairs of categories can be described by only $J - 1$ log-odds equations, although there exist $\binom{J}{2}$ of them.

Usually, the most common category is defined as the baseline. With *Trees* as the baseline category, the two logit equations determining the model can be written as

$$\begin{aligned} \log \frac{\pi_{i2}}{\pi_{i1}} &= \eta_{i2} = \beta_{20} + \beta_{21} \cdot \mathbf{R}_i + \beta_{22} \cdot \mathbf{E}_i + \beta_{23} \cdot \mathbf{RE}_i \\ \log \frac{\pi_{i3}}{\pi_{i1}} &= \eta_{i3} = \beta_{30} + \beta_{31} \cdot \mathbf{R}_i + \beta_{32} \cdot \mathbf{E}_i + \beta_{33} \cdot \mathbf{RE}_i \end{aligned} \quad (1)$$

where the two linear predictors η_{i2} and η_{i3} are related to the class probabilities π_{ik} via the generalized logit link function and R_i is the annual rainfall of the i -th observation, E_i the elevation and RE_i an interaction term between the two variables. For the i -th individual, π_{ik} gives the probability of belonging to class k , where we define $k = 2$ to denote *Lianas* and $k = 3$ as *Hemiepiphytes*. To see why the specification of those two equations suffices, notice that the log-odds between classes *Lianas* and *Hemiepiphytes* can be calculated as

$$\log \frac{\pi_{i2}}{\pi_{i3}} = \log \frac{\pi_{i2}}{\pi_{i1}} - \log \frac{\pi_{i3}}{\pi_{i1}} = (\beta_{20} - \beta_{30}) + (\beta_{21} - \beta_{31}) \cdot \mathbf{R}_i + (\beta_{22} - \beta_{32}) \cdot \mathbf{E}_i + (\beta_{23} - \beta_{33}) \cdot \mathbf{RE}_i$$

Before carrying out the regression, we change the measurement units of *Rainfall* from mm to meters and for *Elevation* from meters to km. This helps to make the regression coefficients more interpretable, as they otherwise take exceedingly small values.

4 Results

The estimated coefficients for Equation (1) are displayed in Table 1. In a multcategory logit model, interpreting the regression coefficients becomes increasingly cumbersome the more regressors the model contains and the more classes the response variable has. Yet, we can already note that all regressors exert a significant effect on the respective log-odds, with one exception: *Elevation* does not appear to influence the logarithm of the odds of belonging to *Hemiepiphytes* compared to *Trees*. Yet, our main interest is not on the effects on the odds but rather how the class proportions themselves change by varying the regressors. To calculate class probabilities, recall that we have

$$\pi_{ij} = \frac{\exp(\beta_{j0} + \beta_{j1} \cdot \mathbf{R}_i + \beta_{j2} \cdot \mathbf{E}_i + \beta_{j3} \cdot \mathbf{RE}_i)}{\sum_{h=1}^4 \exp(\beta_{h0} + \beta_{h1} \cdot \mathbf{R}_i + \beta_{h2} \cdot \mathbf{E}_i + \beta_{h3} \cdot \mathbf{RE}_i)}, \quad (2)$$

where the baseline coefficients are assumed to be $\beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = 0$. To see why this equation holds, notice that we have $\pi_{ij} = \exp(\eta_{ij}) \pi_{i1}$ from Equation (1). Summing over all values of j and rearranging, one obtains $\pi_{i1} = 1 / \sum_{j=1}^J \exp(\eta_{ij})$. Back-substitution then yields the general formula displayed in Equation (2).

Using this result, it is possible to calculate the estimated class proportions for any possible combination of the predictor values. One possibility to visualize the dependency of the estimated probabilities on the regressors is the type of plot depicted in 4, which was originally devised by Fox (2009).

	Estimate	Std. Error	z value	Pr(> z)
Lianas:(Intercept)	-0.888	0.054	-16.571	0.000
Lianas:Rainfall	-0.195	0.017	-11.286	0.000
Lianas:Elevation	-0.345	0.045	-7.713	0.000
Lianas:Rainfall:Elevation	0.061	0.016	3.859	0.000
Hemiepiphytes:(Intercept)	-3.271	0.108	-30.151	0.000
Hemiepiphytes:Rainfall	0.038	0.026	1.452	0.147
Hemiepiphytes:Elevation	-0.316	0.092	-3.458	0.001
Hemiepiphytes:Rainfall:Elevation	0.122	0.027	4.503	0.000

Table 1: Regression output for the multcategory logit model specified in Equation (1)

As we can see from the plot, there are some interesting things going on here: Firstly, for all forest sites with moderate rainfall (the first two plots from the left), a higher elevation is associated with a larger proportion of trees, with the effect most pronounced for the very dry forests where the estimated tree proportions rise from around 70% to over 80%. The pattern for the rain-forests is quite different: Here, a larger amount of *Elevation* is associated with an increase of *Hemiepiphytes*, which comes at the expense of trees and lianas: their combined proportion diminishes to roughly 90% for a forest site with an elevation of 3km and annual rainfall of 4,500 mm. The three considered values for *Rainfall* (1.5, 2.5 and 4.5) were chosen with the goal in mind that class probabilities should only be predicted for predictor values which are quite common in the data set at hand and do not constitute extreme cases.

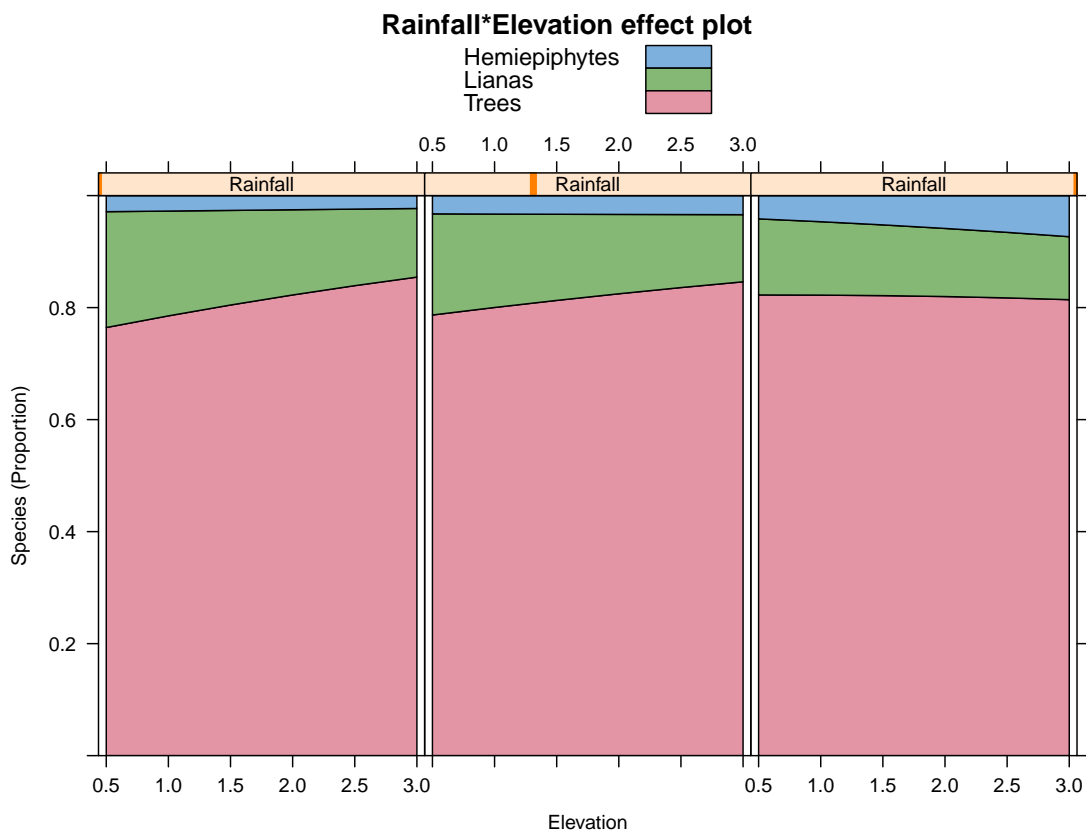


Figure 4: Stacked-area display for the effect of *Elevation* on forest composition at different levels of *Rainfall*. The different plots correspond to forests with annual rainfall of 1.5, 2.5 and 4.5 meters.

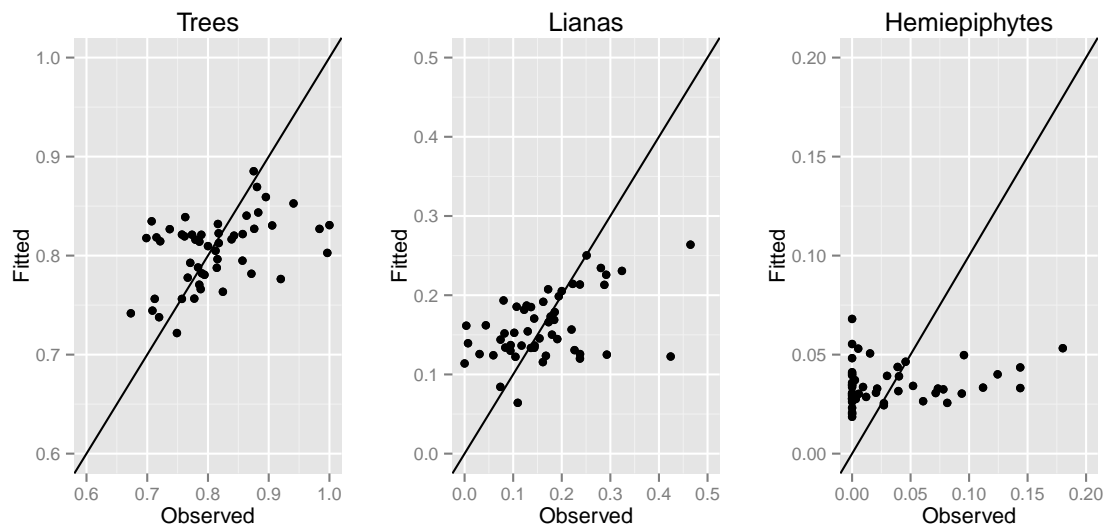


Figure 5: For all classes, the observed class proportions are plotted against the fitted values obtained from Equation (2). For a perfect fit, all data points would be located on the 45° line.

4.1 Diagnostics

The residual deviance of the model is overwhelmingly large (21,989). However, this does not indicate a bad fit, as this statistic is calculated by comparing the fit of the model to a saturated model which correctly predicts every individual plant and not each multinomial cell. To assess the goodness-of-fit, we instead compare the estimated with the empirical class frequencies. The plots in Figure 5 indicate that the model fits the proportions of the *Hemiepiphytes* and the *Lianas* quite poorly, with the tree predictions being on track most of the time. One way to improve model fit would be to collect more data points. Since this is not an option for us, we instead run a Bayesian analysis and compare the results.

5 Bayesian Analysis

5.1 Model Specification

In the Bayesian formulation of the regression problem, prior distributions have to be assigned to the regression coefficients. Posterior distributions are then found for the parameters by applying Bayes rule. Such an approach is not without its pitfalls, since the results can heavily depend on the chosen priors (e.g. the posterior will only have non-zero density on the support of the prior). A substantial but nonetheless fair critique of Bayesian inference is given by Cox (2007). For a comparison of Bayesian and frequentist inference, see Bickel (2012).

Since we do not possess information about effect sizes of rainfall and elevation (mainly due to a lack of knowledge in ecology), an objective Bayes approach was implemented by using non-informative priors. Specifically, the regression coefficients were assigned a multivariate normal distribution with mean zero and the identity matrix serving as the variance-covariance matrix. In preparing the model for R package *rjags*, we adapted a code snippet provided by Ntzoufras (2009).

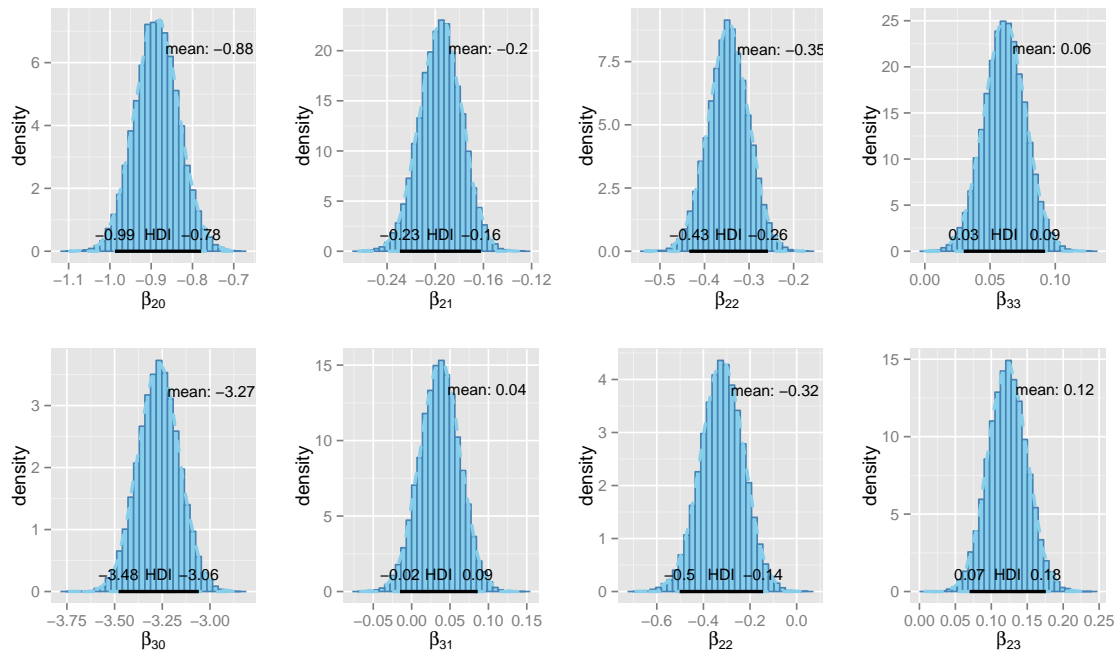


Figure 6: Posterior density plots for the regression coefficients in Equation (1) as obtained from performing MCMC

Figure 6 displays posterior densities for the parameters which were obtained by initializing a Markov Chain Monte Carlo (MCMC) algorithm.

Notice that the obtained point estimates are almost equal to the previously calculated estimates depicted in Table 1. This lends credibility to the outcomes of the frequentist model fitting result. Without any other data at hand which could influence model choice, we can conclude that the presently implemented model captures the relationships among the different variables adequately. To obtain better prediction results, one would need to either use a larger sample for model fitting or incorporate subject knowledge into the analysis.

5.2 Convergence Diagnostics

The posterior densities depicted in Figure 6 were obtained by initializing three Markov chains each running for 1,000,000 iterations. For this purpose, the general-purpose Gibbs sampler implemented in R package *rjags* was used. A burn-in period of 10,000 iterations was specified to limit the amount of samples which are drawn from a low-density region of the chain. In the limit, it does not make a difference at what point the chain was initialized. However, we are averaging a finite number of samples to obtain the estimates of the regression coefficients. To prevent that samples from a low-probability region are over-represented, it is advisable to specify a burn-in period in order to get rid of observations dependent on the starting values. Finally, we specified a thinning factor of 100 to reduce the amount of data which needs to be recorded. Trace plots of the obtained samples are displayed in Figure 7. It appears that all chains have converged in the sense that the empirical distribution of the MCMC samples reasonably approximates the target distribution. Statistical tests did not indicate lack of convergence either: The potential scale reduction factor of Brooks and Gelman (1998) is one for all parameters, as required.

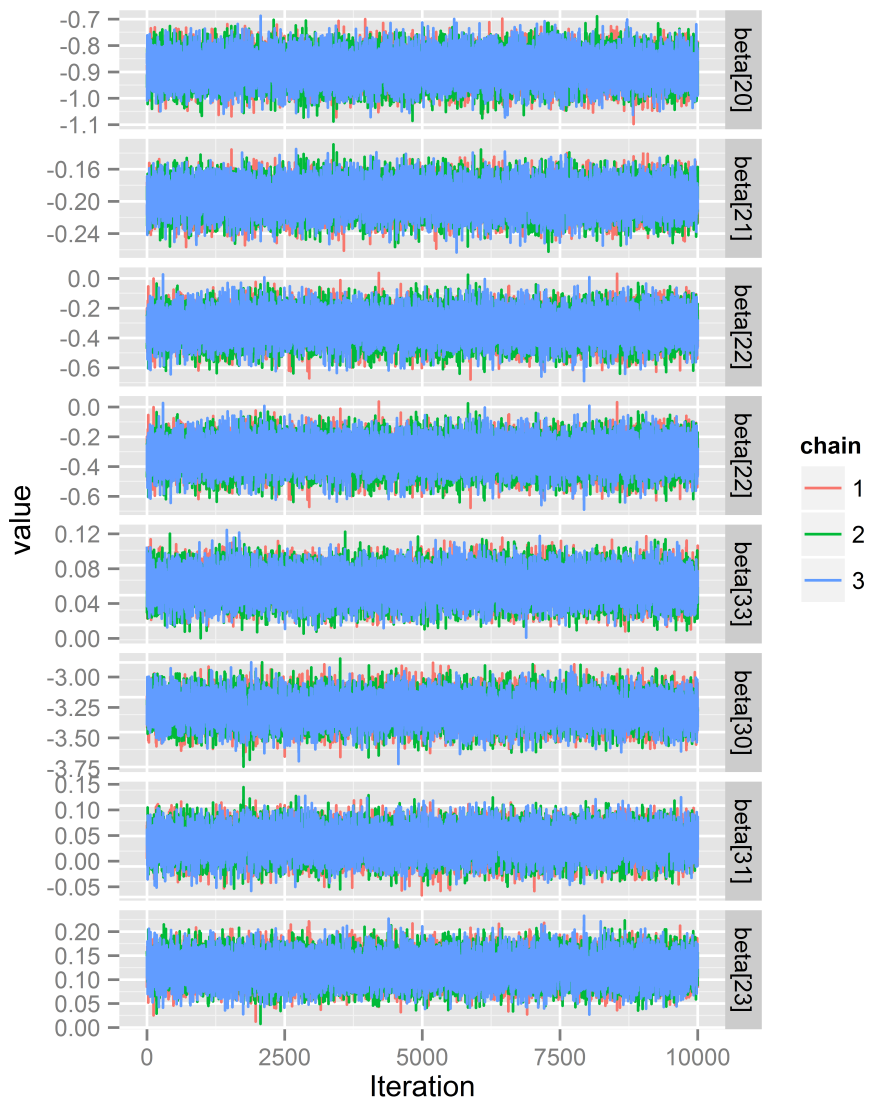


Figure 7: Trace plots for the estimated regression coefficients.

6 Conclusion

Using several visualization techniques, we investigated the structural differences among forest sites in Colombia and Ecuador. Equipped with information about the annual rainfall and the elevation of each site, we studied how these factors influence the amount of *Trees*, *Lianas* and *Hemiepiphytes* at each forest site. Visualization alone has proven to be insufficient here, and therefore we resorted to statistical modelling techniques.

We employed a multinomial logit regression model, using both fixed effects and an interaction term as regressors. This was implemented both using a frequentist and a Bayesian framework. The obtained results have been almost identical, reaffirming the notion that both approaches to inference yield in the limit the same results.

The present analysis could be extended by incorporating subject knowledge into the analysis and collecting more plant data to allow a more detailed description of forest structure.

References

- AGRESTI, A. (1990): *Categorical Data Analysis*, vol. 49 of *Wiley Series in Probability and Mathematical Statistics*, Wiley.
- BICKEL, P. (2012): “Frequentist Inference.”
- BRITANNICA, E. (2006): *Britannica Concise Encyclopedia*, vol. 440, Encyclopaedia Britannica.
- BROOKS, S. AND A. GELMAN (1998): “General methods for monitoring convergence of iterative simulations,” *Journal of computational and graphical . . .*, 7, 434–455.
- COX, D. R. (2007): *Principles of Statistical Inference*, vol. 63, Cambridge University Press.
- FOX, J. (2009): “Extensions to the effects Package,” 32.
- INGWELL, L. L., S. JOSEPH WRIGHT, K. K. BECKLUND, S. P. HUBBELL, AND S. A. SCHNITZER (2010): “The impact of lianas on 10 years of tree growth and mortality on Barro Colorado Island, Panama,” *Journal of Ecology*, 98, 879–887.
- NTZOUFRAS, I. (2009): *Bayesian Modeling Using WinBUGS*, vol. 698 of *Wiley Series in Computational Statistics*, Wiley.
- PHILLIPS, O. (2002): *Global patterns of plant diversity : Alwyn H. Gentry’s forest transect data set*, St. Louis Mo.: Missouri Botanical Press.
- SWAINE, M. D. AND J. GRACE (2007): “Lianas may be favoured by low rainfall: evidence from Ghana,” *Plant Ecology*, 192, 271–276.

Appendix

R Code

```
set.seed(1215) # for reproducibility

# use of knitr (Sweave
# successor) to create this
# report
opts_chunk$set(fig.align = "center",
  cache = FALSE, message = FALSE,
  background = "white", echo = FALSE,
  eval = TRUE)
options(replace.assign = TRUE,
  width = 85)

Plants <- read.csv(file = "plants.csv")

# clean data set replace NA
# with zero for obs 41
Plants[41, 5:6] <- 0

Props <- Plants[, c("Trees", "Lianas",
  "Hemiepiphytes)"]/Plants$Total
Props[is.na(Props)] <- 0
p <- ggplot(aes(x = Rainfall),
  data = Plants) + geom_histogram(binwidth = 2000,
  fill = "white", color = "black")
q <- ggplot(aes(x = Elevation),
  data = Plants) + geom_histogram(binwidth = 1000,
  fill = "white", color = "black")
grid.arrange(p, q, ncol = 2)
# more rainfall, more total
# plants
ggplot(Plants, aes(x = Rainfall,
  y = Total)) + geom_point() +
  stat_smooth(method = "lm")

df1 <- Plants
df1$Site <- reorder(df1$Site, df1$Rainfall)
df1.melted <- melt(df1, measure.vars = c("Trees",
  "Lianas", "Hemiepiphytes"))
colnames(df1.melted)[5] <- "Type"

p <- ggplot(df1.melted, aes(x = Site,
  y = value, fill = Type, alpha = Elevation)) +
  geom_bar(stat = "identity") +
  coord_flip() + theme(legend.position = "bottom")
p

Species <- with(data = Plants,
  cbind(Trees, Lianas, Hemiepiphytes))
Plants$Elevation <- Plants$Elevation/1000
Plants$Rainfall <- Plants$Rainfall/1000

multinom.fit <- multinom(Species ~
  Rainfall * Elevation, data = Plants)
multinom.fit$lev <- c("Trees",
  "Lianas", "Hemiepiphytes")

library(AER)
xtable(coeftest(multinom.fit),
  digits = 3, label = "tab:RegOutput",
  caption = paste0("Regression output for the multicategory ",
  "logit model specified in Equation (\\ref{eqmultinom})"))
plot(effect("Rainfall*Elevation",
  multinom.fit, xlevels = list(Rainfall = c(1.5,
  2.5, 4.5))), style = "stacked",
  rug = FALSE, par.strip.text = list(cex = 0.8),
```

```

xlab = list(fontsize = 10,
            label = "Elevation"), ylab = list(fontsize = 10,
            label = "Species (Proportion)")
Fitted <- fitted(multinom.fit)

p <- qplot(Props[, 1], Fitted[,
1]) + geom_point() + labs(x = "Observed",
y = "Fitted", title = colnames(Fitted)[1]) +
geom_abline(intercept = 0,
            slope = 1) + xlim(c(0.6,
1)) + ylim(c(0.6, 1))
q <- qplot(Props[, 2], Fitted[,
2]) + geom_point() + labs(x = "Observed",
y = "Fitted", title = colnames(Fitted)[2]) +
geom_abline(intercept = 0,
            slope = 1) + xlim(c(0,
0.5)) + ylim(c(0, 0.5))
r <- qplot(Props[, 3], Fitted[,
3]) + geom_point() + labs(x = "Observed",
y = "Fitted", title = colnames(Fitted)[3]) +
geom_abline(intercept = 0,
            slope = 1) + xlim(c(0,
0.2)) + ylim(c(0, 0.2))
grid.arrange(p, q, r, nrow = 1)
InteractionTerm <- Plants$Rainfall *
Plants$Elevation
jagsData <- list(y = Species, x1 = Plants$Rainfall,
n = 53, K = 3, b0 = c(-1, 0,
0, 0), N = Plants$Total,
B0 = diag(1e-04, 4), x2 = Plants$Elevation,
x3 = InteractionTerm)

my.mult <- jags.model(file = "bayesian.jags",
n.chains = 3, n.adapt = 10000,
data = jagsData)

codaSamples <- coda.samples(my.mult,
variable.names = c("beta"),
n.iter = 1e+06, thin = 100)

save(list = "codaSamples", file = "jagsResults.RData")
load(file = "jagsResults.RData")
mat.samples <- as.matrix(codaSamples)

## function from John Kruschke:
HDIofMCMC <- function(sampleVec,
credMass = 0.95) {
# Computes highest density
# interval from a sample of
# representative values,
# estimated as shortest
# credible interval.
# Arguments: sampleVec is a
# vector of representative
# values from a probability
# distribution. credMass is a
# scalar between 0 and 1,
# indicating the mass within
# the credible interval that is
# to be estimated. Value:
# HDIlim is a vector containing
# the limits of the HDI
sortedPts <- sort(sampleVec)
ciIdxInc <- floor(credMass *
length(sortedPts))
nCIs <- length(sortedPts) -
ciIdxInc
ciWidth <- rep(0, nCIs)
for (i in 1:nCIs) {
ciWidth[i] <- sortedPts[i +

```

```

        ciIdxInc] - sortedPts[i]
    }
    HDImin <- sortedPts[which.min(ciWidth)]
    HDImax <- sortedPts[which.min(ciWidth) +
        ciIdxInc]
    HDIlim <- c(HDImin, HDImax)
    return(HDIlim)
}
## custom function to produce
## posterior density plots

plotPosterior <- function(vector,
    param.name = "vector", showHDI = TRUE) {
    HDI <- HDIofMCMC(vector)
    mysize <- 6
    if (showHDI == TRUE) {
        r <- qplot(vector, geom = "blank") +
            geom_histogram(aes(x = vector,
                y = ..density..),
                fill = "#87CEEB",
                col = "steelblue") +
            theme_gray(base_size = 12) +
            geom_density(col = "#87CEEB",
                lty = 2, lwd = 1) +
            labs(x = param.name) +
            geom_segment(aes(x = HDI[1],
                y = 0, xend = HDI[2],
                yend = 0), lwd = 1) +
            geom_text(aes(x = HDI,
                y = 0, label = as.character(round(HDI,
                    digits = 2)),
                vjust = -1, size = 6)) +
            geom_text(aes(x = mean(HDI),
                y = 0, label = "HDI",
                size = 6), vjust = -1) +
            geom_text(aes(x = mean(vector),
                y = max(density(vector)$y),
                vjust = 3, hjust = -0.1,
                label = paste("mean:",
                    as.character(round(mean(HDI),
                        digits = 2))),
                size = 6)) + theme(legend.position = "none")
        return(r)
    } else {
        r <- qplot(vector, geom = "blank") +
            geom_histogram(aes(x = vector,
                y = ..density..),
                fill = "#87CEEB",
                col = "white") +
            theme_bw(base_size = 12) +
            geom_density(col = "#87CEEB",
                lty = 2, lwd = 1) +
            labs(x = param.name) +
            geom_text(aes(x = mean(vector),
                y = max(density(vector)$y),
                vjust = 3, hjust = -0.3,
                label = paste("mean:",
                    as.character(round(mean(HDI),
                        digits = 2))),
                size = 6)) + theme(legend.position = "none")
        return(r)
    }
}
coef.names <- paste("beta[", rep(c(2,
    3), each = 3), rep(0:3, times = 2),
    "]", sep = "")
coef.names2 <- lapply(coef.names,
    function(x) parse(text = x))

p <- plotPosterior(mat.samples[,

```

```

    5], param.name = coef.names2[[1]])
q <- plotPosterior(mat.samples[,
  6], param.name = coef.names2[[2]])
r <- plotPosterior(mat.samples[,
  7], param.name = coef.names2[[3]])
s <- plotPosterior(mat.samples[,
  8], param.name = coef.names2[[4]])
t <- plotPosterior(mat.samples[,
  9], param.name = coef.names2[[5]])
u <- plotPosterior(mat.samples[,
  10], param.name = coef.names2[[6]])
v <- plotPosterior(mat.samples[,
  11], param.name = coef.names2[[7]])
w <- plotPosterior(mat.samples[,
  12], param.name = coef.names2[[8]])
grid.arrange(p, q, r, s, t, u,
  v, w, ncol = 4, nrow = 2)
theme_set(theme_gray(base_size = 11))

# remove baseline
codaSamples2 <- lapply(codaSamples,
  function(x) as.matrix(x))
codaSamples2 <- lapply(codaSamples2,
  function(x) x[, -(1:4)])

# reshape for plotting
codaSamples.melted <- lapply(codaSamples2,
  melt)
codaSamples.melted <- lapply(codaSamples.melted,
  function(x) x[, -1])
codaSamples.melted <- melt(codaSamples.melted)

df4 <- data.frame(Iteration = rep(1:10000,
  24), value = codaSamples.melted$value,
  chain = codaSamples.melted$L1,
  coefficient = codaSamples.melted$Var2)
df4$chain <- as.factor(df4$chain)
df4$coefficient <- factor(df4$coefficient,
  labels = coef.names)

ggplot(aes(x = Iteration, y = value,
  group = chain, col = chain),
  data = df4) + geom_line() +
  facet_grid(facets = coefficient ~
    ., scales = "free_y")

## Gelman Diag:
## gelman.diag(lapply(codaSamples,
## function(x) x[, -(1:4)]))

```