Disentangling the effects of light, silvicultural management and soil on forest plant diversity

Lukas Müller

Bachelor thesis (Student ID 3934554)
submitted to
the Faculty of Environment & Natural Resources
at the Albert-Ludwigs-University Freiburg

Supervisor: Prof. Dr. Carsten Dormann, Department of Biometry and Environmental System Analysis. University of Freiburg
Co-supervisor: Prof. Dr. Florian Hartig, Head of the Group for Theoretical Ecology at the Faculty of Biology and Pre-Clinical Medicine. University of Regensburg

Freiburg, 10th of July 2018

Cover photograph: https://www.pixabay.com; released under the *Creative Commons CC0*.

Contents

Abstract

Biodiversity is often understood to be the main driver of forest ecosystem processes and services. Much research from different perspectives have been done on this matter. The underlying thesis embeds these perspective into a model, while focusing on the yet not to be analyzed relationship between light and species diversity in the herb layer over species' light demand. My studies are based on, and enabled by an extensive research project on biodiversity called *The Biodiversity Exploratories* (Fischer et al., 2010), from which I considered 55 monoculture forest plots in three different regions in Germany of 1 hectare each. Relevant data from them was extracted and put into context, using structural equation modeling (SEM) as main approach for analyzing and bayesian modeling for validation. I consecutively tested multicollinearity, over-fitting after modification of indices, while observing resulting explained variance and goodness of fit to reduce an initially complex model in order to find the simplest, yet still theory and data fitting one. The final model contains seven forest ecosystem properties: Soil type (ST), measured by sand content, main tree species as a measure for forest structure (FS), basal area share of forest canopy, indicating tree layer (TL), mean weighted Ellenberg light values representing properties of the herb layer (HL) as well as an index for forest management intensity *ForMI* (FM) (Kahl & Bauhus, 2014). Interrelating effects between these entities and on the environmental condition respectively main model variable biodiversity (BD) in the herb layer, assessed by effective number of species, especially from Light (LI), could be shown. My model is validated by its fit (P(χ^2): 0.065, RMSEA: 0.12, SRMSR: 0.17, CFI:0.96) and explained variance of dependent model variables (R^2 SEM: HL: 0.84, TL: 0.82, LI: 0.5, FS: 0.82, BD: 0.63). The latter is confirmed by the same model with a bayesian attempt (R^2 Bay.M.: HL: 0.7, TL: 0.81, LI: 0.45, FS: 0.42, BD: 0.57). It was performed to compensate for possible threads to SEM when using small sample sizes. According to my model, sand content in soil decreases biodiversity by 0.6 less plant species in the herb layer for every tenth of weight proportion. My results support the controversial assumption that forest management enhances plant diversity (4.8 more species for every *ForMI* index point). Furthermore, I discovered that for each percentage of photosynthetically active radiation of the key figure of light passing through the forest canopy, 0.78 more plant species grow in the herb layer. As expected, the link is mediated by species' light demand. This coherence, among other known effects, resulting from my model, has not been previously demonstrated. How these findings will be used in future studies or practice goes beyond the scope of this thesis.

Zusammenfassung

Als Hauptursache für funktionierende Prozesse und Leistungen in Waldökosystemen versteht man heute häufig Biodiversität. Viele Forschungen, die verschiedene Teilaspekte dieses Themas untersuchten wurden bereits unternommen. Die zugrunde liegende Arbeit bettet diese in ein Model ein und beleuchtet dabei vorrangig die bislang nicht analysierte Beziehung zwischen Licht und Artendiversität in der Krautschicht über deren Lichtbedarf. Meine Beobachtungen basieren auf Daten aus einem umfangreichen Forschungsprojekt, die *Biodiversity Exploratories* (Fischer et al., 2010), wovon ich 55 Waldparzellen in monokultur aus drei verschiedenen Regionen in Deutschland von je einem Hektar Größe in betracht zog. Ich benutzte Strukturgleichungsmodellierung (SEM) für Analyse und bayesische Modellierung zur Bestätigung um relevante Daten davon in Zusammenhang zu bringen. Ausgehend von einem komplexen Modell benutzte ich Modifikationsindizes und Tests auf Multikolliniarität und Überanpassung im Zusammenhang mit resultierender erklärter Varianz und Anpassungsgüte für die Reduzierung zum einfachsten, jedoch immernoch theorie- und datentreuen Modell. Dieses beinhaltet sieben Waldökosystemeigenschaften: Die Bodenart (ST), gemessen durch den Sandgehalt, die Waldstruktur (FS), welche durch die Hauptbaumart ausdrückt wird, die Baumschicht (TL), gemessen durch den Bedeckungsanteil des Kronendachs, der Krautschicht (HL), beziffert durch die mittleren gewichteten Ellenberg Lichtwerte, sowie ein Maß für die Intensität der Waldbewirtschftung *ForMI* (FM) (Kahl & Bauhus, 2014). Wechselseitige Effekte zwischen diesen und auf die Umweltbedingung und Hauptmodelvariable Biodiversität (BD) in der Krautschicht, bestimmt durch die effektive Artenzahl, vorallem durch Licht (LI), konnten gezeigt werden. Mein Model wird bestätigt durch gute Anpassungswerte (P(χ^2): 0.065, RMSEA: 0.12, SRMSR: 0.17, CFI:0.96) und erklärte Datenvarianz (R^2) abhängiger Modelvariablen (R^2 SEM: HL: 0.84, TL: 0.82, LI: 0.5, FS: 0.82, BD: 0.63). Letztere werden bestätigt durch Ergebnisse aus einem bayesischem Modelierungsansatz (R^2 Bay.M.: HL: 0.7, TL: 0.81, LI: 0.45, FS: 0.42, BD: 0.57). Er wurde durchgeführt, um mögliche Schwierigkeiten für SEM auszugleichen, welche von kleinem Stichprobenumfang ausgehen können. Nach meinem Model verringert Sandgehalt im Boden die Biodiversität um 0.6 weniger Pflanzenarten in der Krautschicht pro Zehntel Gewichtsanteil. Meine Resultate bekräftigen die kontrovers diskutierte Auffassung, dass Waldbewirtschaftung Pflanzenvielfalt erhöht (4.8 mehr Arten für jeden *ForMI* Indexpunkt). Außerdem fand ich heraus, dass für jedes Prozent photosyntetisch aktive Strahlung meiner Schlüsselvariable Licht, dass das Kronendach durchdringt, 0.78 mehr Pflanzenarten in der Krautschicht auftreten. Wie erwartet verläuft dieser Effekt über den Lichtbedarf der jeweiligen Arten. Dieser Zusammenhang, welcher unteren anderen bekannten Effekten von meinem Model erklärt wird, wurde meines Wissens bis heute noch nicht gezeigt. Wie diese Befunde in zukünftigen Studien oder Praxis benutzt werden könnten, geht über diese Arbeit hinaus.

List of abbreviations

BD Biodiversity in the herb layer

BE the *Biodiversity Exploratories*

CFI Comparative fit index

ELV Ellenberg light values

ENS Effective number of species

FEP Forest exploration plots

ForMI Forest Management Intensity Index

FS Forest structure

HL Herb layer

LEC Light extinction coefficient

LI Light

MTS Main tree species

PAR Photosynthetically active radiation

RMSEA Root mean square error of approximation

SEM Structural Equation Model

SMI Silvicultural Management Intensity Indicator

SRMSR Standardized root mean square residuals

ST Soil type

TL Tree layer

Chapter 1

Introduction

Forest ecosystems are complex and depend on numerous properties and entities. For analyzing them in human time scale, some are considered given. Soil components and type, created by bedrock and other geological processes are one example (Gradstein et al., 2012). Some modifications, typically anthropologically caused, take effect quicker and hence might change rapidly again. Silvicultural management in the present and recent past accounts for this case. Others are compounds, assembled from sometimes multi-directional relationships between given properties, temporal changes and anthropological interventions. Among them, biodiversity is a key factor of current changes (Sala et al., 2000). More specifically, biodiversity has decreased in recent times in Central Europe (Settele et al., 2010). Whether or not there is a causal link between this and a functioning ecosystem is a controversial topic among scientists (Loreau et al., 2001), of which most agree about this relation. Working processes, services and resulting benefits provided by a healthy ecosystem however are undoubtedly of high importance (Fischer et al., 2010). This is in of itself provides just cause for further investigation into the matter. Much research has been done on the relationship between forest management and biodiversity at various scale levels and evaluated from different perspectives (Hietel et al., 2005; Houghton, 1994). However, the literature on this topic presents contrary conclusion, depending on for example, locational scale (Mittelbach et al., 2001). Ultimately, the type of forest management determines whether and how it contributes to species richness. While former practices such as clear-cutting tend to decrease biodiversity, more recent strategies have had the opposite effect (Kuuluvainen, 2009). Hence the examined age-class forests are expected to have a positive impact. The regional scale of my study is set by measurements taken by Fischer et al. (2010) in three different regions of Germany. The diversity in structure and composition, captured in those plots, represents most of Germany‘s temperate forests and allows interpretation for at least the whole country. To compensate for complexity and aim for interpretability, only monocultural forests plots among them were considered. For the same reasons, vascular plant biodiversity was only examined in the herb layer.

The aim of this thesis is creating and discussing a model containing all necessary ecosystem entities that allows for interpretation while emphasizing the scarcely studied impact of light on biodiversity. Specifically in forests, where such research has not yet been carried out. The expectation is, that light effects species composition based on their light demand and the amount of light competition (Hautier et al., 2009) and thus impacts biodiversity (Fig. A.1). This could be due to a larger pool of species with medium to high light demand and, that even plant species with low Ellenberg light value benefit from more light. Generally, more energy input promotes plant individual abundance and thus biodiversity. I tested these assumptions and focused on both, human impact due to: 1) forest management; and more over 2) the role of light on plant diversity in theory and demonstrated by the model; 3) Change in soil properties with was given a minor character.

Chapter 2

Methods

2.1 From ecological theory and data to model

An ecological model that represents theoretical thinking for realistic interpretation is always threatened to diverge from its purpose as a result of a statistical model fitting pursuit (Fox et al., 2015). To avoid this, I built a primary model sketch, not containing any functional syntax or definite regression arrows (Fig. 2.1).

Model variables were found in the dataset (Tab. C.1), could be calculated from here or found elsewhere (e.g. from FlorWeb, BfN, 2018). I subsequently evaluated the model with regressions between a pair, or a set of variables deemed carrying theory-fitting information. These were primarily gained from correlations among measures (Fig. A.4). The concept was to explain recent forest structure by given soil conditions and anthropological impact due to forest management, each selected by different indicators and respective properties. Light is introduced as an effect

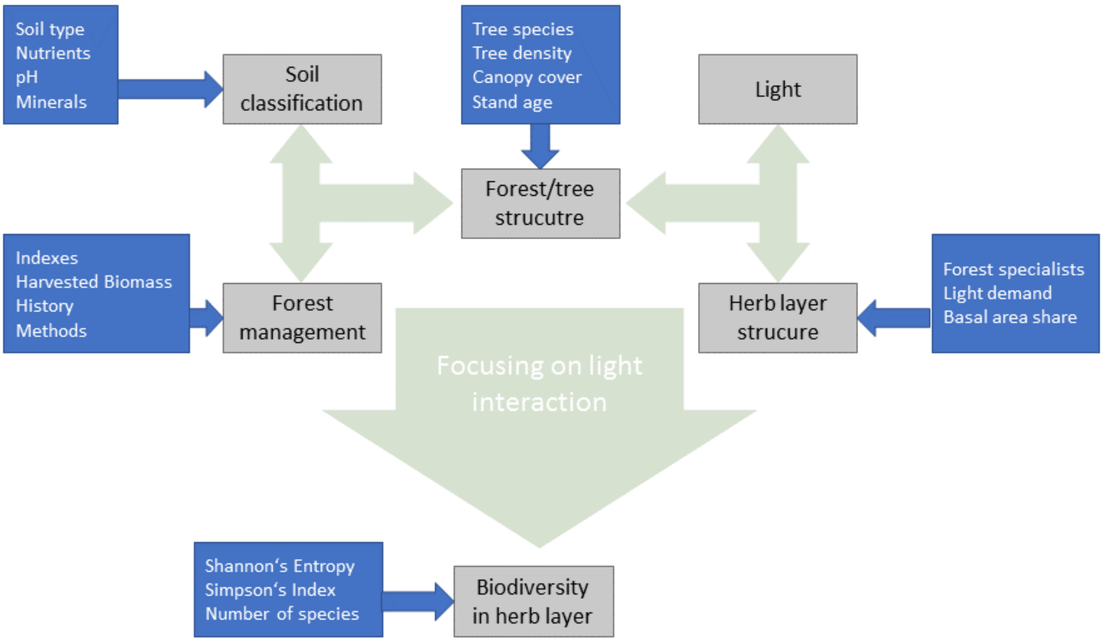


Figure 2.1: First model sketch (Blue: Input Ecosystem properties/Model indicators; Grey: Ecosystem entities/Model variables; Green: Effects)

and input variable for herb and tree layer. From this main structure, effects on biodiversity were expected to emerge.

The statistical programming language R (Ihaka & Gentleman, 2018) was used to model and analyze the underlying data. The software version used was 3.4.3.

R‘s ‘lavaan‘ package (Rosseel et al., 2018) as it provides an *open-source* solution for such models and was used for analyzing structural equation modeling (henceforth reffered to as SEM). The package contains tools for building, analyzing, improving and predicting SEM.

Another modeling technique, using a Bayesian approach with R‘s ‘jagsUI‘ package (Kellner, 2017) was introduced later to confirm obtain results from SEM.

2.2 Data origin

Data used for modeling, analyzing and interpreting was taken from a large-scale and long-term research on biodiversity undertaken by Fischer et al. (2010) called *The Biodiversity Exploratories* (henceforth BE). The data and according meta data is partially available to the public at BExIS (2018). See appendix C for detailed information.

In three different regions of Germany (biosphere reserve Schorfheide-Chorin, national park Hainich and surroundings and biosphere reserve Swabian Alb), the team gathered a vast number of measurements on each 50 forest and grassland plots of 1 hectare. Those regions mainly include pine (*Pinus sylvestris*), spruce (*Picea alba*) and European beech (*Fagus sylvatica*) in mixed and monoculture stands. They grow on different soils ranging from glacial shaped sand-rich sediment, to Cambisols, as well as limestone and karst forming Podzols. All varying greatly on a local scale.

For my analysis, I used half of forest exploration plots (hence named FEP),because of available light assessments. I further limited my scope by taking only pure-stand forests into consideration to 55 FEP in total.

2.3 Latent variables and indicators

Latent variables are not observed but inferred by variables, measured directly. The ones I used are forest management, soil and according soil type, tree species and resulting canopy, light and biodiversity in the first concept model and were reduced to the simplest, yet still data and theory-fitting model.

2.3.1 Light

Mean weighted Ellenberg Light values

Each vascular plant, found on FEP (Tab. C.1 20366) was recorded annually. I took only data from 2016 to achieve better coherence with used light data (Chap. 2.3.1) For each species, I assigned its according layer (herbs, shrubs, trees taller and smaller than 10 meters) and collected its *Ellenberg* light value (ELV; Ellenberg, 2003) from FloraWeb (BfN, 2018) using R‘s ‘distdrawr‘ (Friess, 2017) and ‘curl‘ (Jeroen Ooms, 2018) packages. Unknown species to FloraWeb were added manually with Seybold (2009). Mean values of species‘ ELV for each plot were taken for each layer and weighted for their respective relative basal cover share. Some species grow unaffected by surrounding light conditions (ELV = indifferent). Layers in plots where these unknown mean ELV take a share greater than 50% were taken out of consideration. This was also done for species that were unrecorded or had no sufficient taxonomic unit. Equation 2.1 displays the calculation of weighted ELV, exemplary for layer α with s = specie, C = cover and p = number of plots in respective layer.

$$\mathrm{wELV}_\alpha = \frac{1}{p}\sum_p \frac{\sum_s \mathrm{ELV}_{\alpha s}}{\sum_s C_{\alpha s}} \tag{2.1}$$

See Figure A.3 for ELV distribution among all species found on FloraWeb and on FEP. ELV were also recorded in 2010 (Tab. C.1 14410). I used them to compare them with my calculated values (*Pearson's* $\rho = 0.75$ and mean difference $= 0.97$) and chose the latter, since they are based on species data, collected in 2016.

Light Extinction Coefficient (LEC)

Light on FEP was recorded as the absolute values of photosynthetically active radiation (PAR), below (*LineData*) the canopy and on the nearest clearing or canopy gap as reference (*RefData*). According to Valladares (2003), these are the most plant-applicable light measures. These have been taken on 10 partially consecutive days with 25 measurements randomly across the surface of every FEP for 15 minutes, using the 'Quantum Li-COR 190R PAR' passive sensor. For reference measurements, a 'Quantum Li-COR 190R' sensor was used. The data was taken between August 2017 and September 2017. To avoid disturbances due to e.g. especially cloudy days for some particular plots, the LEC, (Eq. 2.2) was used as meaningful light parameter (Tab. C.1 22506).

Additionally, I took into consideration that some light, measured below the canopy results from diffusive light. The vegetation cover above the measure points in relation to the LEC, hitherto is calculated as

$$\mathrm{LEC} = \frac{\mathrm{LineData}}{\mathrm{RefData}} \tag{2.2}$$

suggests a exponential relation (Fig. 2.2a). The green curve shows aggregated basal area share ELV, described in section 2.3.1, the other two were computed from given values of tree and shrub cover share (Tab C.1 14410). *Beer–Lambert law* (Swinehart, 1962) applies in which light decay follows an exponential term when going through a different medium (such as leaf cover in this case).

By fitting a linear function of the form

$$\log \mathrm{LineData} = \log a + b\mathrm{RefData}, \tag{2.3}$$

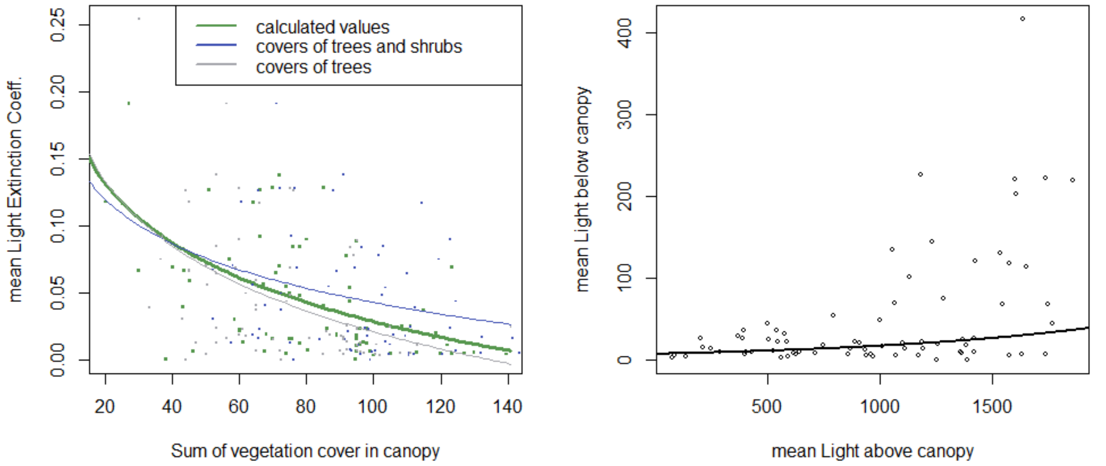
I calculated a function for LEC, shown on Figure 2.2b and equation 2.4, using R's 'nls2' package (Grothendieck, 2015).

$$\mathrm{LEC}_{\mathrm{corrected}} = \frac{\mathrm{LineData}}{6.76e^{0.0009RefData}} \tag{2.4}$$

Validating the computed corrected LEC reveals no pattern in residuals (Fig. 2.3a) and is proved by higher correlation (Fig. 2.3b).

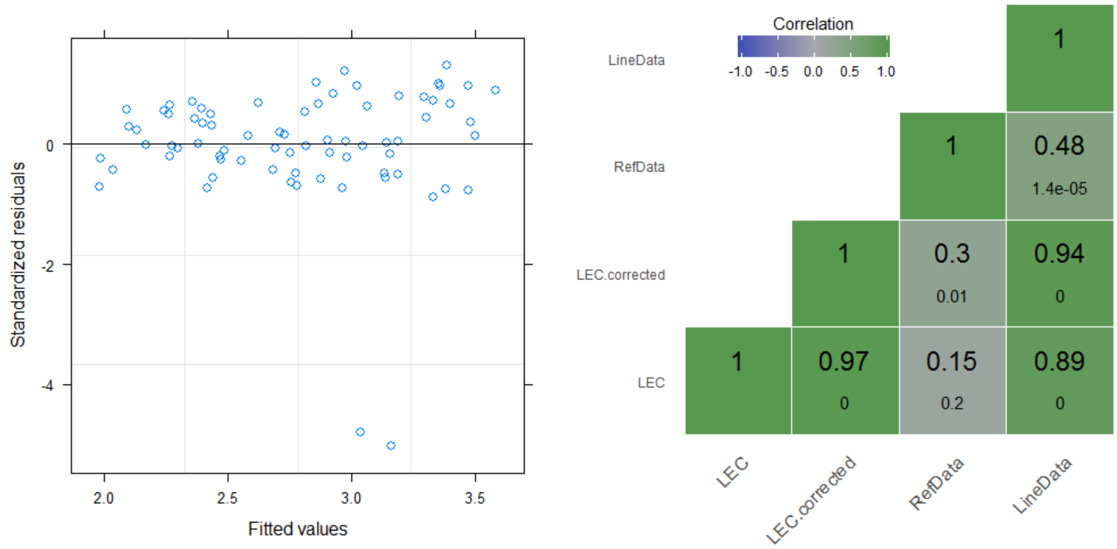
2.3.2 Biodiversity

For biodiversity, I only used the data from the herb layer as indicators because it accommodates the most unique species recorded in the underlying data (59 in shrub layer, 269 in herb layer). Therefore effects on biodiversity would be represented most reliably. This is important, when working with datasets limited by small sample sizes. I used two indices, presented in the following, measuring biodiversity. In FEP these were recorded in 2010 (Tab. C.1 14410). I could not calculate them for more recent years, as some needed variables for calculation are only available for 2010.



(a) Comparing light and vegetation cover (b) Exponential function for light extinction coefficient

Figure 2.2: Meaningful light parameter: *Light Extinction Coefficent*



(a) Observing residuals of corrected light extinction coefficient. Shown outliers were tested for overly impact.

(b) Correlations of corrected and given LEC. *Pearson's* ρ above respective p-value

Figure 2.3: Validating corrected light extinction coefficient

Effective number of species in herb layer

Rather than taking Shannon's entropy as an index of species diversity, I used its exponential for a comparable and linear parameter for measuring the effective number of species (ENS henceforth) (Hill, 1973) as suggested by Jost (2006). Equation 2.5 shows how it is calculated, where N is ENS, H Shannon's entropy, $\rho_i = \frac{n_i}{n}$ the share of specie *i* in a system of *n* individual plants.

$$N = e^{-\sum \rho_i \ln \rho_i} = e^H \tag{2.5}$$

Table 2.1: Main tree species (**Fagus sylvatica; Picea abies; Pinus sylvestris*)

Tree specie*	Beech	Spruce	Pine
Amount	36	9	10

Simpson‘s Index

Simpson‘s Index (Simpsons, 1949) is used as a measurement of diversity of an ecosystem in terms of its evenness. It is calculated by the probability of two randomly taken individuals, both belonging to different species. It took this index into consideration too, as taking only one might fail to capture diversity fully (Morris et al., 2014; Wilsey et al., 2005; Whittaker, 1972; Heino et al., 2008)

2.3.3 Herb layer

Basal area share of forest specialists

The forest specialist type of a species tells if it is primarily and typically found in forest habitats or not. This categorisation was taken from 'FloraWeb‘, using the same R methods as used for collecting ELV in Chapter 2.3.1. In the same way, the values were weighted for their species‘ basal area share. The species found on FEP also originate from Table. C.1 20366.

2.3.4 Forest structure

Main tree species MTS

As an indicator of how the forest is structured, I took its MTS (Tab. C.1 10580) as an estimating model variable. MTS were determined between 2008 and 2010. Table 2.1 shows their occurrences on FEP. To avoid distortion data, I only used forest plots with pure stand (Tab. C.1 17706). Another reason for excluding mixed stands results from relatively high proportion of missing data in admixed tree species (41% to 74%). Four different succession stages (thicket, immature timber, mature timber and pole wood, Tab. C.1 17706) are considered in the processed data to obtain a continuous tree age distribution which enhances result interpretability as suggested by Fischer et al. (2010).

2.3.5 Forest mangement

For quantifying the extend of silvicultural management, I took two indexes into consideration, as they account for different effects.

ForMI

Kahl & Bauhus (2014) propose their Forest Management Intensity Index (Tab. C.1 16466) based on three parts. The proportion of harvested tree volume, non native tree species (neophytes) and dead wood showing traces of saw cuts. This index was established on the same BE as the ones used in this thesis which accounts for usability in this context. The authors suggest that ForMI can be used for assessment covering retroactively for the past 30 to 40 years. The data used to calculate the index was collected between 2006 and 2012.

SMI

The Silvicultural Management Intensity Indicator (Tab. C.1 17746) is composed of two main subindices, each consisting of different aspects. A risk component (SMI_r), depending on tree

Table 2.2: Goodness of fit

Index or test	value	robust value
$P(\chi^2)$	1	0.21
CFI	1	0.95
RMSEA	0	0.06
$P(\text{RMSEA}) \leq 0.05$	1	0.39
SRMR	0.73	0.73

Table 2.3: R^2 for dependent latent variables

Variable	R^2
Herb layer	0.997
Tree layer	0.865
Light	0.651
Forest structure	0.760
Biodiversity	0.682

species and stand age and a density component (SMI_d) calculated with figures of stand age, tree species and silvicultural regime (Schall & Ammer, 2013).

2.3.6 Soil type

For suitable soil parameters: I took soil type, measured in gram per kilogram content of sand; silt, and clay (Tab. C.1 14686). The soil samples examined between May 2011 and March 2012 were taken from the upper 10 cm of the mineral soil. Dispersion of soil aggregates into discrete units and separation of soil particles of different size (sand: 2-0.063 mm, silt: 0.063-0.002 mm and clay: <0.002 mm) was done by sieving and sedimentation according to DIN-ISO 11277. This also included soil drying and destruction of soil organic matter with hydrogen peroxide. The proportion of each type found in soil, alter plant species abundance by providing different amounts of reactive surface for microbes (Berg & Smalla, 2009) and ions, as well as a range of diverse drainage properties (Blume et al., 2009). It also accounts for different types of bedrock, soil depth (Blume et al., 2009) and according favored tree species growing on it. Other soil parameters such as pH-value or microbiotic activity were only available for 33 plots and thus not used (Tab. C.1 10574).

2.3.7 Tree layer

I took the total percentage of basal area cover of trees as indicator for tree layer. Only those taller than 10 meters were recorded for this figure. Rather than taking available data from 2010 (Tab C.1 14410), I gathered and calculated this figure in the same way as ELV (Chap. 2.3.1) and basal area share of forest specialists (Chap. 2.3.3) to ensure for temporal consistency.

2.4 Model simplification and improvement

The first model (Fig. 2.4), involving all described latent variables, shows acceptable data fit.

Table 2.2 displays tests and indexes, used to demonstrate goodness of fit. Since most data is not normally distributed (Fig. A.2) and can not be transformed to do so, robust values are taken into consideration. Ruling out this as possible source of error, my Bayesian model, conducted later, assigned better fitting distributions to each variable. A non-significant χ^2 test, implies that the empirical variance matrix Σ, calculated from the data and the one computed by the model $\Sigma(\theta)$ did not differ systematically (Steinmetz, 2015). Besides this, fit indices such as RMSEA, (Steiger, 1998), SRMSR, (Joreskog & Sorbom, 1986) and CFI (Bentler & Peter, 1990) were established to quantify the extent of misspecification of the model. Hu & Bentler (1999) proposes that models with RMSEA < 0.06, SRMSR < 0.08 and CFI close to, or greater than 0.98, to be acceptable, which are all met by the underlie model. After validating the model, the percentage of explained variance of depended (latent) variables (R^2) reveals the validity of later interpretation.

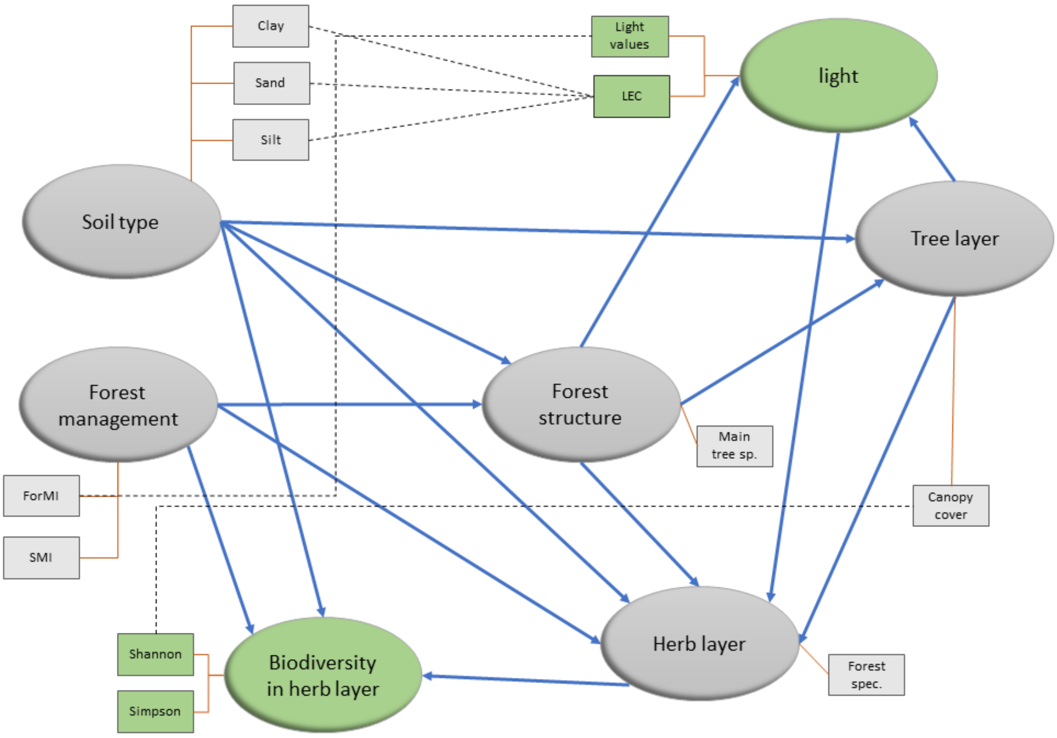


Figure 2.4: Blue arrows indicate regressions. Black dashed lines reveal covariances between indicators. Brown lines show measuring indicators for latent variables.

All these criteria however disregard multicollinearity and over-fitting to be a possible source of distortion (Steinmetz, 2015; Bentler & Peter, 1990). Over-fitting in practical terms happens when a model contains many regressions in ratio to model variables (in an extreme case: more parameters than and the model can sustain) and therefore corresponds too closely the data (Definition: Pearsall, 2018).

Multicollinearity occurs when one indicator of a latent variable can be linearly predicted by another indicator of the same latent variable. A set of tests from R‘package ‘mctest‘ (Ullah & Aslam, 2018; Imdadullah & Aslam, 2016) exposed multicollinearity between sand and silt, measuring soil type for their effects on latent variables shown on Figure 2.4. The package involves the methods ‘Determinant‘ (Cooley & Lohnes, 1971) ,‘R-Squared‘ (Gujarati & Porter, 1999), ‘Farrar-χ^2‘ (Farrar & Glauber, 1967), ‘Condition index‘ (Belsley et al., 2005), ‘Sum of reciprocal of eigenvalues‘ (Ahrens, 1987), ‘Theil‘s indicator‘ (Theil, 1971) and ‘Red indicator‘ Kovàcs et al. (2005), which all proved multicollinearity for soil type indicators. Both sand and silt account for similar ecological effects, such as soil drainage (Cosby et al., 1984), whereas clay is an important element for creating reactive surface. Another reason for removing silt as a of soil type is that each of its three indicators add up to 1, which makes at least one of them redundant. Significant tests thereafter result in only leaving sand as the most powerful soil indicator. The same procedure reveals, that each latent variable, in this set of data can only be determined by a single indicator, which rejects the concept of latent variables for the model. For each other tested latent variable, Farrars χ^2, Red indicator and Theil‘s method showed multicollinearity. This is confirmed by (Kolenikov & Bollen, 2012) as all model attempts created negative variances in both observed and latent variables, so-called *Heywood Cases*, (Rindskopf, 1984). Another reason for those cases is small sample size, which always threatens SEM goodness of fit indices and interpretability

Table 2.4: Model variable with fitting distribution.
* Gradient-log-normalization of categorical normal probability distribution

Variable	Distribution
Light	Gamma
Tree layer	Beta
Herb layer	Normal
Biodiversity	Gamma
Forest structure	Softmax*

Steinmetz (2015). Only having 55 FEP as samples, I had to select one variable each. For finding an indicator for forest management, I chose ForMI rather than SMI, as it was established that the former, studied the same set of forest plots as analyzed with this model (Kahl & Bauhus, 2014). Simpson‘s Index, focusing on species evenness in comparison to the effective number of species, calculated from Shannon‘s Entropy (Jost, 2006) showing richness more clearly, led to choosing the latter. Light is best represented by the percentage passing through the canopy, while mean ELV in the herb layer is a comparatively indirect measure, altered by shadows casted by shrub layer plants and indifferent ELV.

Table 2.3 leads to the assumption that the especially high amount of explained variance of herb layer might result of over-fitting (Stone, 2016). While further investigating using R‘s ‘caret package‘ (Kuhn et al., 2018), I noticed that leaving out only 1% of data already drops R^2 values by over 10% of the particular over-fitting-presumed herb layer variable. Leaving out 9%, results in an unidentified and not converging model. This can be interpreted as the model being on the edge of over-parameterization (Whittaker et al., 2010). Furthermore, the effects of e.g. forest management on forest structure and tree layer improves fit and R^2 but do not necessarily serve theoretical thinking. Either the direct effect should be neglected or the mediator between the two, forest structure, would be ineffective in this regression. This is backed by observing significance levels of regressions within the model (Fig. A.6).

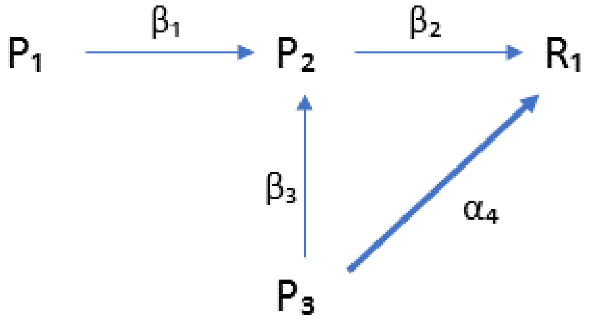
Conducting these practices to all regressions and correlations, I designed a simplified model with an effect of light on biodiversity (Fig. 3.1). Modification indices were used subsequently to make predictions of parameter change in regression or correlation drop (Hox & Bechger, 1998). This method can only be used so long as the structure of the model supports its theory and fits (Steinmetz, 2015), which was the case for the simplified model.

2.5 Bayesian approach

To verify my model, obtained from theoretical thinking and analysis with SEM, I used R‘s ‘jagsUI‘ Package Kellner (2017), which implements the ‘JAGS‘ (Plummer, 2017) program. Similar to ‘BUGS‘ (Thomas, 2014), it uses Markov Monte Carlo chain sampling (Gilks et al., 1995) for Bayesian regression analysis. A major difference to SEM with ‘lavaan‘ is that distributions, other than normal, can be fitted to each variable, before calculating regression parameters. This contributes to the model‘s accuracy and, in my case, to its needed validation. As stated earlier, most data used, is not normally distributed. I used **R**‘s ‘fitdistrplus‘ (Delignette-Muller & Dutang, 2017) package to find suitable distributions for each of the dependent variables (Tab. 2.4). The respective likelihood function and corresponding parameterisation, as well as deparameterisation used to calculate R^2 (Eq. 2.6), can be found in appendix B.4.

$$R^2 = 1 - \frac{\sum(X_{obs} - X_{pred})^2}{\sum(X_{obs} - X_{null})^2} \tag{2.6}$$

Figure 2.5: Sketch of manually calculating total effect of predictors on response variable



with X_{obs} = Observed values of variable X, X_{pred} = model predicted values of variable X and X_{null} = model precited values of variable X with null model.

Resulting R^2 of depended variables can be compared with the ones calculated from the same model with 'lavaan'. I conducted this additional step to ensure, that low sample size does not impact plausibility of the model, as this is not the case in Bayesian regression analysis.

2.6 Model analysis

Unbalanced impact of outliers were inspected, using R's 'faoutlier' Chalmers (2017) and 'influence.SEM' (Pastore & Altoe, 2018) package. In order to assess the model's robustness, I conducted a perturbation analysis by adding a normally distributed error of 10% on each variable, repeatedly for 1000 samples and observed how explained variances of depended variables changed (Fig. A.5).

Subsequently I examined the model's effect strength. The idea was to predict biodiversity using regression paths and altered input data. For example, multiplying light and forest management with the same factor would reveal the ratio of effect strength. Lavaan (developers version: 0.6-1.1230, respectively 0.5-23.1097) offers *lavPredict*-function for this purpose but is not designed for handling models without latent variables at the time of writing. It is promised to do so in future releases (Rossel, 2017). The function *sem.predict* of R's 'piecewiese.SEM' package (Lefcheck, 2016) failed in predicting the entire model and was only able to forecast local regressions, due to the way it is programmed. I extended the function to support multinomial logistic regressions to include MTS, because of its categorical data structure. I exposed that the function merely wraps R's predict method, and therefore is not capable of making predictions for regressions interacting with each other. One would need to feed obtained results from a regression consecutively into its next path branch. Conclusively the model's sensitivity could only be assessed by manually summating all (in)direct effect strengths shown as an example in Figure 2.5. As illustrated, the effect η of the regression of P$_3$ on R$_1$ (equation 2.7) can be calculated by equation 2.8.

$$R_1 \sim P_3 = \alpha_4 P_3 + \beta_3 P_3 * \beta_2 P_2 \tag{2.7}$$

$$\eta = \alpha_4 + \beta_3 * \beta_2 \tag{2.8}$$

Chapter 3

Results

After various simplifications and improvements, I developed a forest ecosystem model (Fig. 3.1). The model supports the introductory formulated hypothesis of an effect of light on biodiversity through plants' light demand, estimated by ELV. It contains 7 variables.

Light	measured by its percentage of PAR, passing the forest canopy (LEC).
Soil	indicated by the soil type property of sand content which is strongly correlated with other soil types and therefore summarizing their interpretations.
Forest structure	estimated by the main tree species growing in respective plots.
Tree layer	measured by the basal area share of trees taller than 10 meters.
Herb layer	figured by the weighted mean of ELV.
Forest Management	quantified by ForMI, an index established on BE as well.
Biodiversity	in herb layer calculated with ENS .

Choosing only one indicating variable for each environmental entity resulted from multicollinearity in variables, when adding more indicators.

The model involves no unbalanced impact of outliers and is not further improvable by modification of variables (indices), which is underlined by fit indices (Tab. 3.1) and explained variance of depended variables (Tab. 3.2). Since fit indices are slightly lower than suggested (Chap.2.4), I conducted a Bayesian modeling approach, which is unaffected by small sample size. Its comparative R^2 values are also presented in Table 3.2. They show great similarity, except for explained variance of forest structure respectively among main tree species.

Adding a 10% error to each variable revealed no big impact, underlining the robustness of the model (Figure A.5).

Table 3.1: Goodness of fit

Index or test	value
$P(\chi^2)$	0.65
CFI	0.96
RMSEA	0.11
P(RMSEA) $\leq$	1
SRMR	0.124

Table 3.2: R^2 of dependent variables

Variable	R^2 (SEM)	R^2 (bayesian)
Herb layer	0.842	0.701
Tree layer	0.821	0.812
Light	0.503	0.446
Forest structure	0.825	0.423
Biodiversity	0.632	0.568

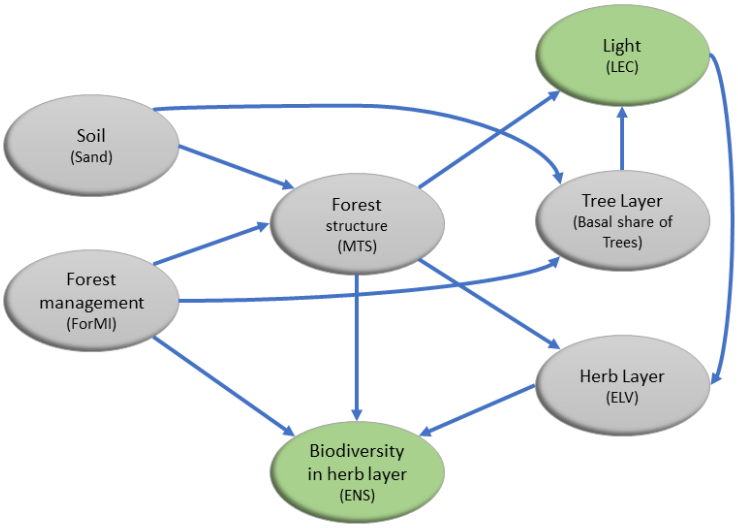
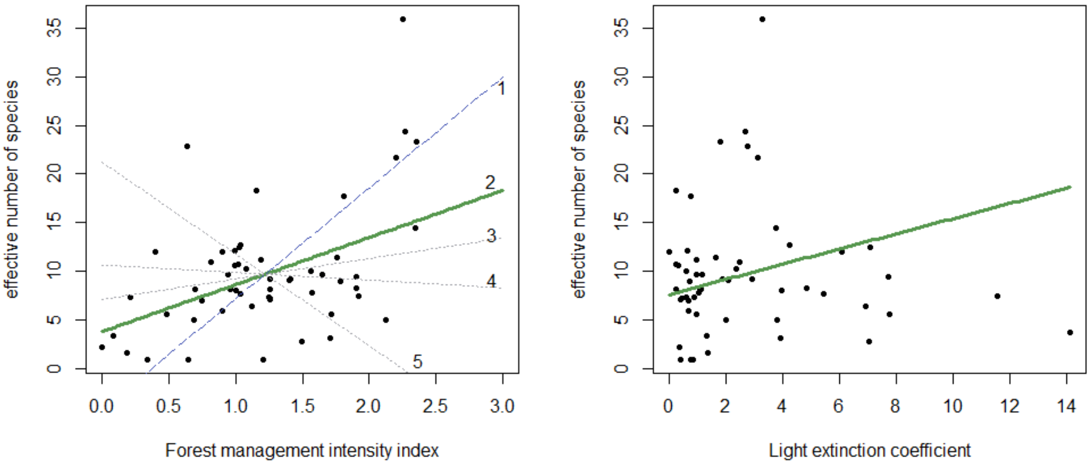


Figure 3.1: Model with indirect effect of light on biodiversity.Blue arrows indicate regressions. Abbreviations are explained in chapter 2.3



(a) Effect strength of forest management on biodiversity. Bold green line is total, dashed blue direct and dotted gray indirect effects. For explanation, see Figure A.7

(b) Indirect effect of light on biodiversity

Figure 3.2: Effects of forest management and light on biodiversity in herb layer

Investigating the hypothesis of testing the role of light in having an influence on biodiversity reveals an indirect impact. This is shown in Figure 3.2, compared to my findings of the also significant role that forest management and soil type have on biodiversity (Fig. A.8). The latter two are stronger in absolute values with 0.44 (Forest management) and -0.39 (Soil type) over 0.35 (all scaled to be comparable). This results in 4.8 more species per ForMI point, 0.6 less with every 100g per kg more sand content in soil and 0.78 respectively for every percent of light, passed through the forest canopy.

Chapter 4

Discussion

The obtained results from the examined forest plots with structural equation model (Fig. 3.1) are discussed separately in the following.

4.1 Forest management intensity and biodiversity

My findings show a rise of the effective number of species in the herb layer when intensifying forest management, as suggested by (Hansen et al., 1991) and (Kahl & Bauhus, 2014) who believe it to be the main driver for biodiversity in forest ecosystems. This may be the result of a development in managing forests in a biodiversity sustaining way (Fedrowitz et al., 2014) in central Europe as compared to Germany. The vast majority of forest plots, considered in my analysis, were managed age-class forests, which support species richness for understorey plants (Raison et al., 2001) by stand structural complexity, e.g. achieved by multiple age classes and spatial arrangement (Franklin & Pelt R, 2004). This holds even considering that only monoculture forests were tested, which would classically thought of, cause the opposite. Former techniques such as clear-cutting, that superficially seem to simulate nature-close events into silvicultural methods proved to be wrong (Esseen et al., 1997) and were abandoned (Mann, 2011; Naturschutzgesetz, 2011) in the examined BE (Fischer et al., 2010). However, this still is apparent in older studies, stands and other regions and may seem surprising to undifferentiated reader. Other known parameters in management towards the goal of biodiversity, for instance is maintaining the stability of hydrological systems (Lindenmayer et al., 2006), could not be judged by the data available. Much research has been done on this matter (Heller & Zavaleta, 2009) that helped to develop recommendations, used during the last decades to make improvements and seem to be showing their effect now. The importance of biodiversity interacting with ecosystem productivity in comparison to other factors (Tilman et al., 2012; Zhang et al., 2012) is successfully implemented in forest management for the analyzed regions.

Within the range of assessed management intensities, the effective number of species rose steadily. If, and at which point this might shrink or stagnate is still to be determined. This could be computed in future releases (Rossel, 2017) of prediction-method containing software for this particular purpose.

4.2 The role of light

Examined as one of the first studies, my findings support the hypothesis of light having an impact on plant diversity in forest ecosystems. For each percentage more of light passing through the canopy, a rise in effective number of species in the herb layer by 0.77 could be significantly modeled. The more light that is available on the ground, the greater the abundance of plants with higher light demand grow. This results in more species at ground level. One explanation is, that smaller trees compete with shrubs in forest plots with more light, where less tall trees grow. This competition costs energy and causes draw backs for larger trees while favoring smaller trees and leads to greater biodiversity in the underlie herb layer. This is due to less competition between herb and shrub layer and more effective space. Another idea is that higher light availability, as major energy input, promotes a greater number of individual plants. This, for example, among other factors enriches plant abundance in tropic forests. Adaption to darker habitats as always results in less species succeeding. To confirm this, I surveyed the plant database of FloraWeb (BfN, 2018), which revealed that there are far more recorded species with higher light demand. Light availability therefore rises the overall capacity of species richness. The significance of this suggestion has to be relativised, since not all species are effectively able to grow in the studied environment. One quarter of forest specialist species have indifferent light values and thus, this factor in general, can not be understood as a direct measure of which species actually grow. Therefore change of light does not significantly alter (Fig. A.11) the composition of forest specialist plants, even though non-forest types prefer higher light intensities on average. Why this is the case has not been studied so far, to my knowledge. One explanation for this could be that clearings or other patches of light are only temporal or too far inside forests, which diminish chances of seed dispersal vectors to succeed.

4.3 Soil type and effective number of species

Sand content, as indicating variable for soil type showed a negative effect on plant species richness in the herb layer. This is due to several reasons. Sandy soils have lower water holding capacities, less dissolved ions in soil solution, quicker draining and accordingly lower pH (Blume et al., 2009). The same capacity limitation is true for nutrients in the soil matrix. This makes only less water and nutrient-demanding, as well as acidity adapted plants able to grow. Sandier soils also consist of less finer components ranging from coarse silt to clay, which results in less reactive surface by some magnitudes. This decreases the amount of microbial biomass in the soil layer (Fig. A.10), which results in less micro scale bioturbation and bacterial functions e.g. nitrogen-fixing (Van Der Heijden et al., 2008). I found that forest plots with sandier soils support less trees >10 and more <10 meters with great effect strength. They do so because of decreased trunk stability and thus resulting in improved light conditions for smaller trees. Subjacent shrub layer area share and effective number of species are decreasing less strong from this than in the underlying herb layer (Fig. A.9). This might be due to some tall growing shrubs, specialized for limited nutrient and water resources, casting more shadow and thus worsening plant conditions on the ground. Limited resources generally support highly specialized plants and therefore decrease biodiversity (Schulze & Chapin, 1987).

4.4 Model predictive power and uncertainty

Although the model‘s predicting variables for biodiversity, forest management, forest structure and herb layer are correlated (Pearson $\rho = 0.38$ in average), they show robustness, when tested (Fig. A.5) on sensitivity. However this collinearity still possibly poses some caution when interpreting obtained results, even when stated tests are done (Dormann et al., 2013).

The data for effective number of species and main tree species were taken in 2010. While the latter is of minor importance, since MTS doesn’t change that fast, the former may have an influence, as light data was collected in 2017 and the rest was taken in 2016 or calculated from it.

One of the possible reasons that led to rejecting latent variables with SEM was the limited sample size of dataset I used of 55. This lowers validity and trust in goodness of fit indices and tests. My simplified model therefore has slightly worse SRMSR, RMSEA and CFI than suggested by Hu & Bentler (1999). Since all of those measures are not independent of sample size, I adduced comparative R^2 values from the same model with a Bayesian approach. Another reason to do so, was that models, not containing any latent variables, are not yet fully implemented in the program package I used for running my model Rossel (2017). Not knowing to what extent this is the case in the program version, released at the time of writing, I could only verify my model's explained variance with this additional approach. As SEM, with the used software, is only capable of dealing with normal distributions, this method yields additional correction in using different distributions. The interpretation of the only significant diverge of both model's results in R^2 of forest structure goes beyond the scope and statistical depth of this thesis. I rate the overall small difference despite different statistical approaches however, as strong confirmation of my results.

Acknowledgements

I thank my supervisor Prof. Dr. Carsten Dormann and co-supervisor Prof. Dr. Florian Hartig for their support. Especially the former provided me with insight and expertise that greatly assisted my thesis.

Among other useful literature I'd like to single out Steinmetz (2015) book 'Lineare Strukturgleichungsmodelle - Eine Einführung mit R' as very helpful.

Furthermore I thank Florian Keppeler for advise in use of JAGS and Daniel Di Marzo for language assistance.

Bibliography

Ahrens, H. (1987). Multivariate Analysis - Methods and Applications. *Biometrical Journal*, 29(6), 755–756.

Belsley, D., Edwin, A., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*, volume 571. John Wiley & Sons.

Bentler, B. & Peter, M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, 107(2), 238.

Berg, G. & Smalla, K. (2009). Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS microbiology ecology*, 68(1), 1–13.

BExIS (2018). Biodiversity Exploratories Information System. Internet Database.

BfN (2018). FloraWeb. https://www.floraweb.de. Bundesamt für Naturschutz Bonn.

Blume, H., Brümmer, G., Scheffer, F., Horn, R., Kandeler, E., Schachtschabel, P., Kögel-Knabner, I., Welp, G., Kretzschmar, R., & Thiele-Bruhn, S. (2009). *Lehrbuch der Bodenkunde*. Spektrum Lehrbuch. Spektrum Akademischer Verlag.

Chalmers, P. (2017). R Package 'faoutlier'. https://cran.r-project.org/web/packages/faoutlier/faoutlier.pdf.

Cooley, W. & Lohnes, P. R. (1971). *Multivariate Data Analysis*. John Wiley & Sons Inc.

Cosby, B., Hornberger, G., Clapp, R., & Ginn, T. (1984). A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water resources research*, 20(6), 682–690.

Delignette-Muller, M.-L. & Dutang, C. (2017). R Package 'fitdistrplus'. https://cran.r-project.org/web/packages/fitdistrplus/fitdistrplus.pdf.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46.

Ellenberg, H. (2003). *Zeigerwerte von Pflanzen in Mitteleuropa*. Scripta Geobotanica. Erich Goltze GmbH & Company.

Esseen, P.-A., Ehnstrøm, B., Ericson, L., & Sjøberg, K. (1997). Boreal Forests. *Ecological Bulletins*, (pp. 16–47).

Farrar, D. & Glauber, R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, (pp. 92–107).

Fedrowitz, K., Koricheva, J., Baker, S. C., Lindenmayer, D. B., Palik, B., Rosenvald, R., Beese, W., Franklin, J. F., Kouki, J., & Macdonald, E. (2014). Can retention forestry help conserve biodiversity? A meta-analysis. *Journal of Applied Ecology*, 51(6), 1669–1679.

Fischer, M., Bossdorf, O., Gockel, S., Hänsel, F., Hemp, A., Hessenmöller, D., Korte, G., Nieschulze, J., Pfeiffer, S., Prati, D., et al. (2010). Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. *Basic and Applied Ecology*, 11(6), 473–485.

Fox, G., Negrete-Yankelevich, S., & Sosa, V. (2015). *Ecological Statistics: Contemporary Theory and Application*. Oxford University Press.

Franklin, J. & Pelt R, V. (2004). Spatial Aspects of Structural Complexity in Old-Growth Forests. *Journal of Forestry*, 102, 22–28.

Friess, N. (2017). R Package 'distdrawr'. https://cran.r-project.org/web/packages/distdrawr/distdrawr.pdf.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. CRC press.

Gradstein, F. M., Ogg, J. G., Schmitz, M., & Ogg, G. (2012). *The geologic time scale 2012*. elsevier.

Grothendieck, G. (2015). R Package 'nls2'. https://cran.r-project.org/web/packages/nls2/nls2.pdf.

Gujarati, D. N. & Porter, D. C. (1999). Essentials of econometrics.

Hansen, A. J., Spies, T. A., Swanson, F. J., & Ohmann, J. L. (1991). Conserving biodiversity in managed forests. *BioScience*, 41(6), 382–392.

Hautier, Y., Niklaus, P. A., & Hector, A. (2009). Competition for light causes plant biodiversity loss after eutrophication. *Science*, 324(5927), 636–638.

Heino, J., Mykrä, H., & Kotanen, J. (2008). Weak relationships between landscape characteristics and multiple facets of stream macroinvertebrate biodiversity in a boreal drainage basin. *Landscape Ecology*, 23(4), 417–426.

Heller, N. E. & Zavaleta, E. S. (2009). Biodiversity management in the face of climate change: A review of 22 years of recommendations. *Biological Conservation*, 142(1), 14 – 32.

Hietel, E., Waldhardt, R., & Otte, A. (2005). Linking socio-economic factors, environment and land cover in the German Highlands, 1945–1999. *Journal of Environmental Management*, 75(2), 133 – 143.

Hill, M. O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54(2), 427–432.

Houghton, R. A. (1994). The Worldwide Extent of Land-Use Change. *BioScience*, 44(5), 305–313.

Hox, J. J. & Bechger, T. M. (1998). An introduction to structural equation modeling. *Family Science Review*.

Hu, L.-t. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1–55.

Ihaka, R. & Gentleman, R. (2018). The R Project for Statistical Computing. https://www.r-project.org/.

Imdadullah, M. & Aslam, M. (2016). Mctest: Multicollinearity Diagnostic Measures,. *URL https://CRAN. R-project. org/package= mctest. R package version*, 1, 495.

Jeroen Ooms, H. W. (2018). R Package 'curl'. https://cran.r-project.org/web/packages/curl/curl.pdf.

Joreskog, K. & Sorbom, D. (1986). *LISREL VI, analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods.* Scientific Software.

Jost, L. (2006). Entropy and diversity. *Oikos*, 113(2), 363–375.

Kahl, T. & Bauhus, J. (2014). An index of forest management intensity based on assessment of harvested tree volume, tree species composition and dead wood origin. *Nature Conservation*, 7, 15–27.

Kellner, K. (2017). R Package 'jagsui'. https://cran.r-project.org/web/packages/jagsUI/jagsUI.pdf.

Kolenikov, S. & Bollen, K. A. (2012). Testing negative error variances: Is a Heywood case a symptom of misspecification? *Sociological Methods & Research*, 41(1), 124–167.

Kovàcs, P., Petres, T., & Tóth, L. (2005). A new measure of multicollinearity in linear regression models. *International statistical review*, 73(3), 405–412.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2018). R Package 'caret'. https://cran.r-project.org/web/packages/caret/caret.pdf.

Kuuluvainen, T. (2009). Forest management and biodiversity conservation based on natural ecosystem dynamics in northern Europe: the complexity challenge. *AMBIO: A Journal of the Human Environment*, 38(6), 309–315.

Lefcheck, J. (2016). R Package 'piecewiseSEM'. https://cran.r-project.org/web/packages/piecewiseSEM/piecewiseSEM.pdf.

Lindenmayer, D., Franklin, J., & Fischer, J. (2006). General management principles and a checklist of strategies to guide forest biodiversity conservation. *Biological Conservation*, 131(3), 433 – 445.

Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J., Hector, A., Hooper, D., Huston, M., Raffaelli, D., & Schmid, B. (2001). Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science*, 294(5543), 804–808.

Mann, S. (2011). Forest Protection and Sustainable Forest Management in Germany and the P.R. China – A Comparative Assessment. *BfN.* Federal Agency of Nature Conservation.

Mittelbach, G. G., Steiner, C. F., Scheiner, S. M., Gross, K. L., Reynolds, H. L., Waide, R. B., Willig, M. R., Dodson, S. I., & Gough, L. (2001). What is the observed relationship bwtween species richness and productivity? *Ecology*, 82(9), 2381–2396.

Morris, E. K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T. S., Meiners, T., Müller, C., Obermaier, E., Prati, D., et al. (2014). Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecology and evolution*, 4(18), 3514–3524.

Naturschutzgesetz (2011). Federal Nature Conservation Act, Bundesnaturschutzgesetz §5 Land-, Forst- und Fischereiwirtschaft (3). German Government.

Pastore, M. & Altoe, G. (2018). R Package 'influence.SEM'. https://cran.r-project.org/web/packages/influence.SEM/influence.SEM.pdf.

Pearsall, J. (2018). *Oxford Dictionary*. Oxford.

Plummer, M. (2017). Just Another Gibbs Sampler. http://mcmc-jags.sourceforge.net/.

Raison, R. J., Brown, A. G., & Flinn, D. W. (2001). *Criteria and indicators for sustainable forest management*, volume 7. CABI.

Rindskopf, D. (1984). Structural equation models: Empirical identification, Heywood cases, and related problems. *Sociological Methods & Research*, 13(1), 109–119.

Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., tulla, M., Katsikatsou, M., Barendse, M., Chow, M., & Jorgensen, T. (2018). R Package 'lavaan'. https://cran.r-project.org/web/packages/lavaan/lavaan.pdf.

Rossel, Y. (2017). Future plans for lavaan. http://lavaan.ugent.be/development.html.

Sala, O. E., Stuart, F. C. I., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L. F., Jackson, R. B., Kinzig, A., Leemans, R., Lodge, D. M., Mooney, H. A., Oesterheld, M., Poff, N. L., Sykes, M. T., Walker, B. H., Walker, M., & Wall, D. H. (2000). Global Biodiversity Scenarios for the Year 2100. *Science*, 287(5459), 1770–1774.

Schall, P. & Ammer, C. (2013). How to quantify forest management intensity in Central European forests. *European Journal of Forest Research*, 132(2), 379–396.

Schulze, E. & Chapin, F. (1987). Plant specialization to environments of different resource availability. In *Potentials and limitations of ecosystem analysis* (pp. 120–148). Springer.

Settele, J., Penev, L., Georgiev, T., Grabaum, R., Grobelnik, V., Hammen, V., Klotz, S., Kotarac, M., & Kühn, I. (2010). *Atlas of Biodiversity Risk. Editorial.* Pensoft.

Seybold, S. (2009). Schmeil-Fitschen: Flora von Deutschland und angrenzender Länder. *Quelle andMeyer, Heidelberg, Wiesbaden.*

Simpsons, E. H. (1949). Measurement of Diversity. *Nature*, 163, 688.

Steiger, J. H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling: A Multidisciplinary Journal*, 5(4), 411–419.

Steinmetz, H. (2015). *Lineare Strukturgleichungsmodelle: Eine Einführung mit R.* Sozialwissenschaftliche Forschungsmethoden. Rainer Hampp Verlag.

Stone, B. K. (2016). Five Reasons Why Your R-squared Can Be Too High. http://blog.minitab.com/blog/adventures-in-statistics-2/five-reasons-why-your-r-squared-can-be-too-high.

Swinehart, D. F. (1962). The Beer-Lambert Law. *Journal of Chemical Education*, 39(7), 333.

Theil, H. (1971). *Principles of econometrics*. Journal of Marketing Research.

Thomas, A. (2014). Bayesian inference Using Gibbs Sampling. http://openbugs.net/w/FrontPage.

Tilman, D., Reich, P. B., & Isbell, F. (2012). Biodiversity impacts ecosystem productivity as much as resources, disturbance, or herbivory. *Proceedings of the National Academy of Sciences*, 109(26), 10394–10397.

Ullah, D. M. I. & Aslam, D. M. (2018). R Package ”mctest”. https://cran.r-project.org/web/packages/mctest/mctest.pdf.

Valladares, F. (2003). Light heterogeneity and plants: from ecophysiology to species coexistence and biodiversity. In *Progress in botany* (pp. 439–471). Springer.

Van Der Heijden, M. G., Bardgett, R. D., & Van Straalen, N. M. (2008). The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology letters*, 11(3), 296–310.

Whittaker, G., Confesor, R., Di Luzio, M., Arnold, J., et al. (2010). Detection of overparameterization and overfitting in an automatic calibration of SWAT. *Transactions of the ASABE*, 53(5), 1487–1499.

Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon*, (pp. 213–251).

Wilsey, B. J., Chalcraft, D. R., Bowles, C. M., & Willig, M. R. (2005). Relationships among indices suggest that richness is an incomplete surrogate for grassland biodiversity. *Ecology*, 86(5), 1178–1184.

Zhang, Y., Chen, H. Y., & Reich, P. B. (2012). Forest productivity increases with evenness, species richness and trait variation: a global meta-analysis. *Journal of ecology*, 100(3), 742–749.

Appendix A

Graphs

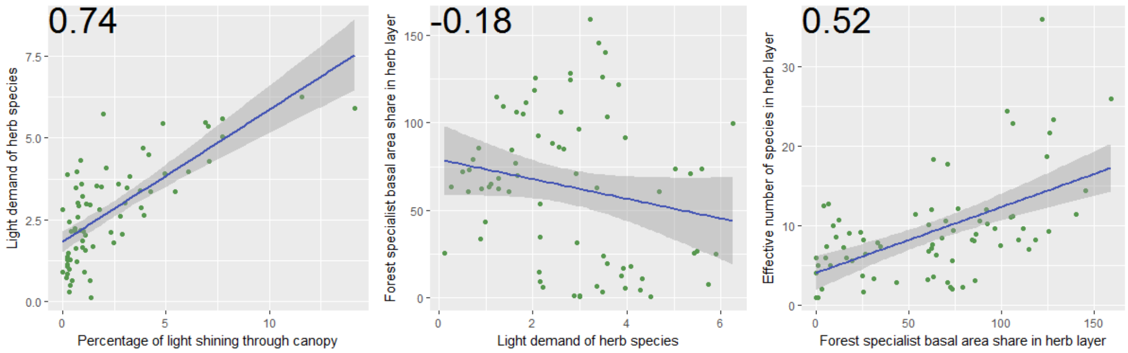


Figure A.1: Cascading effects of light on herb species composition and its diversity with Pearsons ρ correlation.

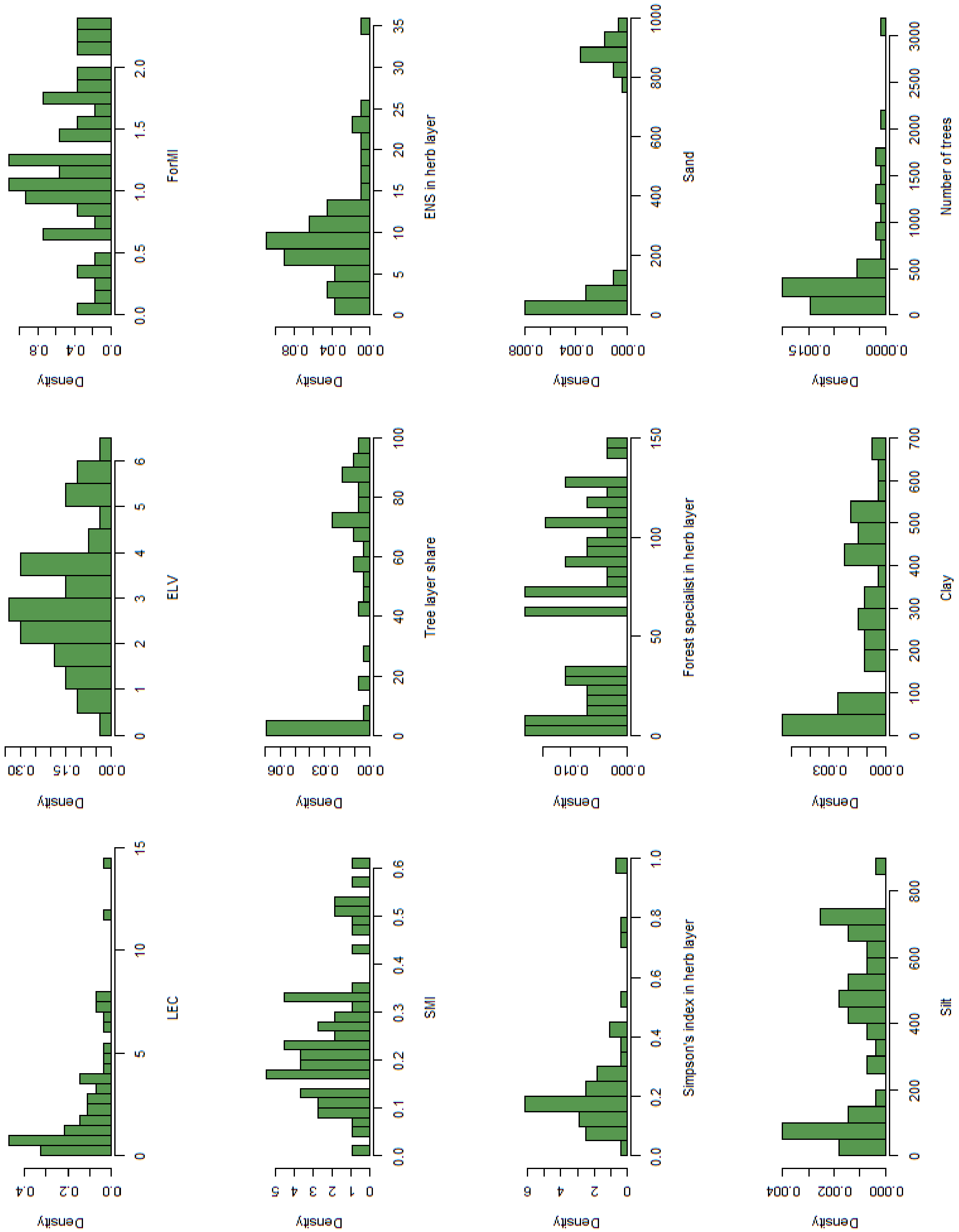
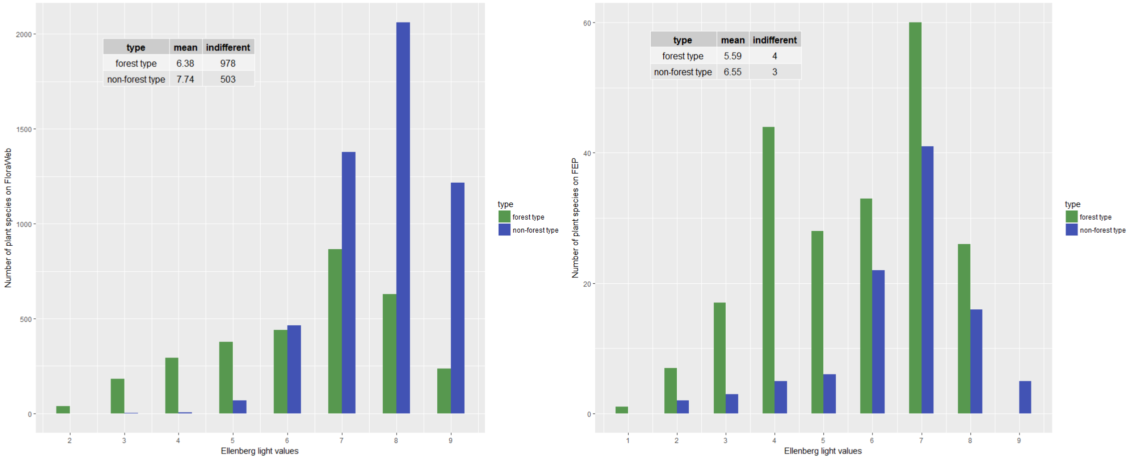


Figure A.2: Histogram of all used variables.



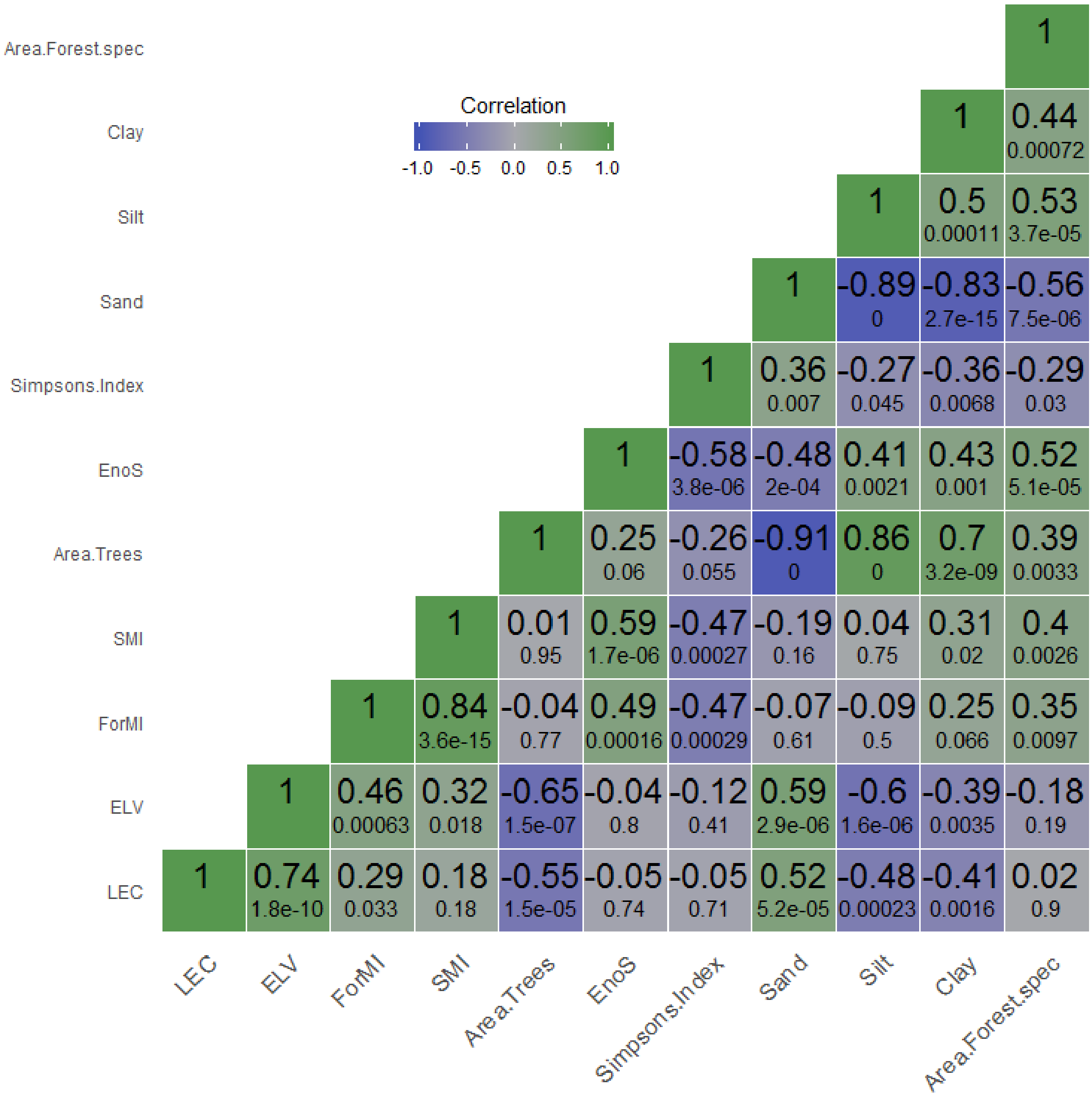
type	mean	indifferent
forest type	6.38	978
non-forest type	7.74	503

type	mean	indifferent
forest type	5.59	4
non-forest type	6.55	3

(a) Distribution of ELV of all plant species found on FloraWeb

(b) Distribution of ELV of plant species found on FEP

Figure A.3: distinguished by forest specialist type.



	LEC	ELV	ForMI	SMI	Area.Trees	EnoS	Simpsons.Index	Sand	Silt	Clay	Area.Forest.spec
Area.Forest.spec											1
Clay										1	0.44 0.00072
Silt									1	0.5 0.00011	0.53 3.7e-05
Sand								1	-0.89 0	-0.83 2.7e-15	-0.56 7.5e-06
Simpsons.Index							1	0.36 0.007	-0.27 0.045	-0.36 0.0068	-0.29 0.03
EnoS						1	-0.58 3.8e-06	-0.48 2e-04	0.41 0.0021	0.43 0.001	0.52 5.1e-05
Area.Trees					1	0.25 0.06	-0.26 0.055	-0.91 0	0.86 0	0.7 3.2e-09	0.39 0.0033
SMI				1	0.01 0.95	0.59 1.7e-06	-0.47 0.00027	-0.19 0.16	0.04 0.75	0.31 0.02	0.4 0.0026
ForMI			1	0.84 3.6e-15	-0.04 0.77	0.49 0.00016	-0.47 0.00029	-0.07 0.61	-0.09 0.5	0.25 0.066	0.35 0.0097
ELV		1	0.46 0.00063	0.32 0.018	-0.65 1.5e-07	-0.04 0.8	-0.12 0.41	0.59 2.9e-06	-0.6 1.6e-06	-0.39 0.0035	-0.18 0.19
LEC	1	0.74 1.8e-10	0.29 0.033	0.18 0.18	-0.55 1.5e-05	-0.05 0.74	-0.05 0.71	0.52 5.2e-05	-0.48 0.00023	-0.41 0.0016	0.02 0.9

Figure A.4: Correlation matrix of all used continuous variables. Respective *Pearson‘s r* above p-value.
Explanation: LEC (Ch. 2.3.1), ELV (Ch. 2.3.1), ForMI (Ch. 2.3.5), SMI (Ch. 2.3.5), Area trees (Ch. 2.3.7), ENS (Ch. 2.3.2), Simpson‘s index (Ch. 2.3.2), Sand, Silt, Clay (Ch. 2.3.6), Area Forest spec (Ch. 2.3.3)

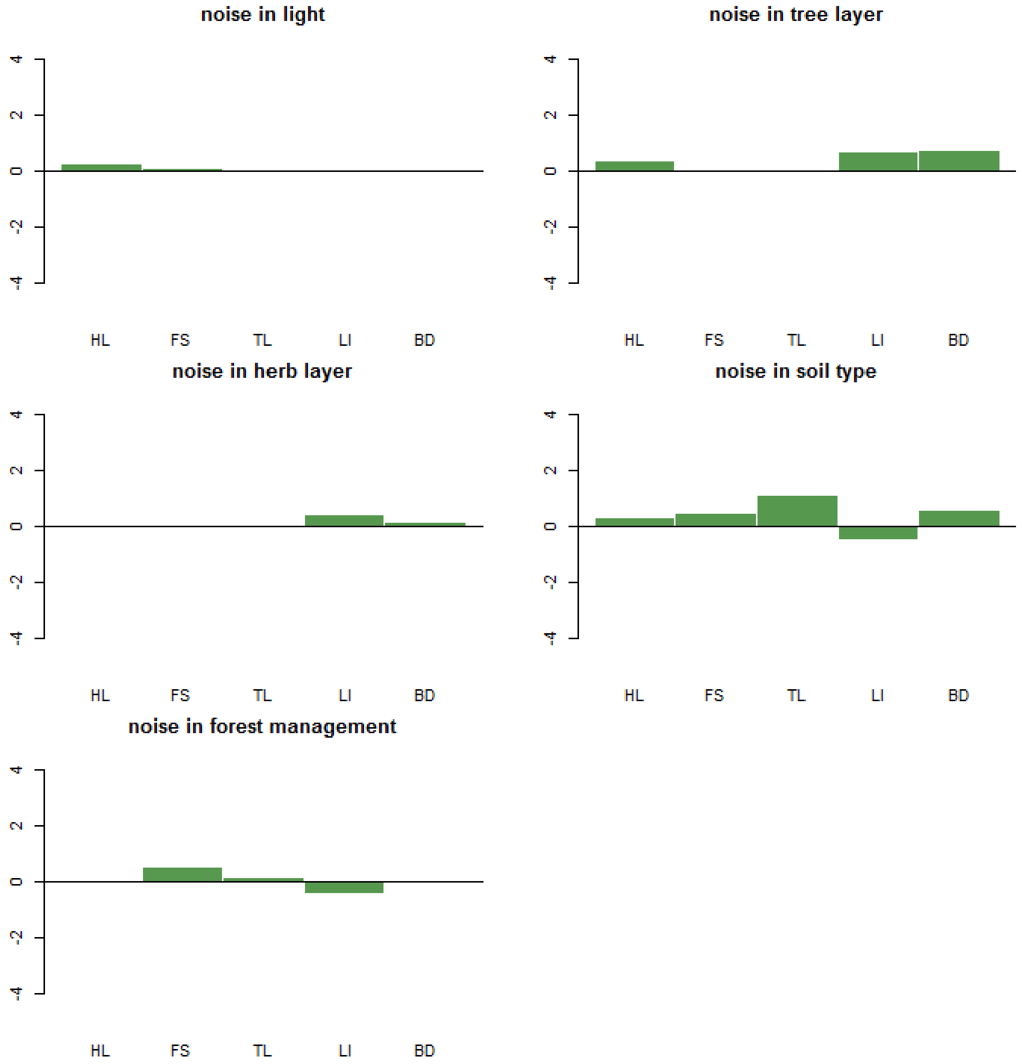
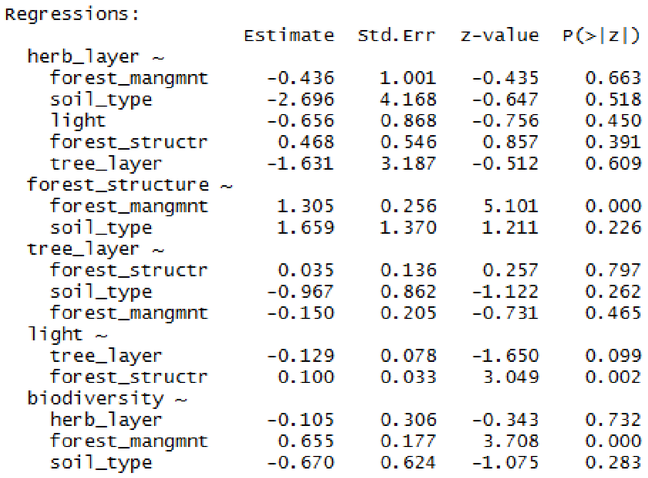


Figure A.5: Barplots of the effect of the respective variable on all depended variables.

Figure A.6: Regression for complex model



```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  herb_layer ~
    forest_mangmnt   -0.436    1.001   -0.435    0.663
    soil_type        -2.696    4.168   -0.647    0.518
    light            -0.656    0.868   -0.756    0.450
    forest_structr    0.468    0.546    0.857    0.391
    tree_layer       -1.631    3.187   -0.512    0.609
  forest_structure ~
    forest_mangmnt    1.305    0.256    5.101    0.000
    soil_type         1.659    1.370    1.211    0.226
  tree_layer ~
    forest_structr    0.035    0.136    0.257    0.797
    soil_type        -0.967    0.862   -1.122    0.262
    forest_mangmnt   -0.150    0.205   -0.731    0.465
  light ~
    tree_layer       -0.129    0.078   -1.650    0.099
    forest_structr    0.100    0.033    3.049    0.002
  biodiversity ~
    herb_layer       -0.105    0.306   -0.343    0.732
    forest_mangmnt    0.655    0.177    3.708    0.000
    soil_type        -0.670    0.624   -1.075    0.283
```

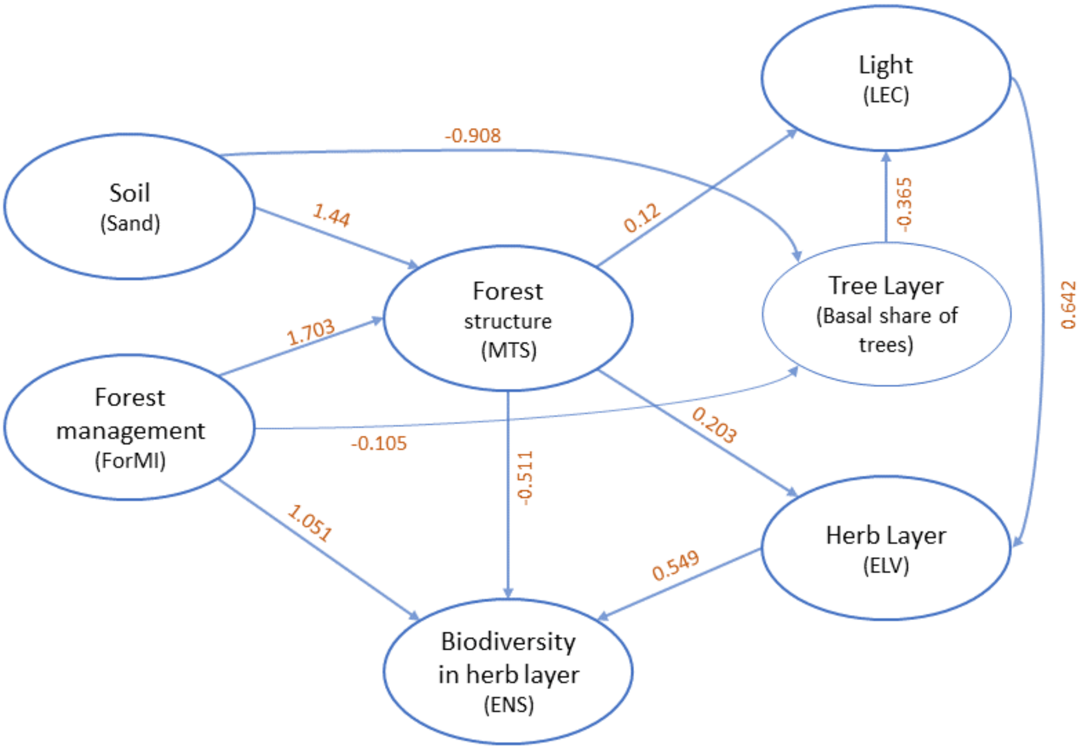


Figure A.7: model, with effect strengths

Forest management on biodiversity

Effect 1: direct effect

Effect 2: total effect

Indirect effects:

3: FM ⇒ FS ⇒ HL ⇒ BD

4: FM ⇒ FS ⇒ LI ⇒ HL ⇒ BD

5: FM ⇒ FS ⇒ BD

Soil type on biodiversity

Effect 4: total effect.

Indirect effects:

1: ST ⇒ FS ⇒ HL ⇒ BD

2: ST ⇒ FS ⇒ LI ⇒ HL ⇒ BD

3: ST ⇒ FS ⇒ TL ⇒ LI ⇒ HL ⇒ BD

5: ST ⇒ FS ⇒ BD

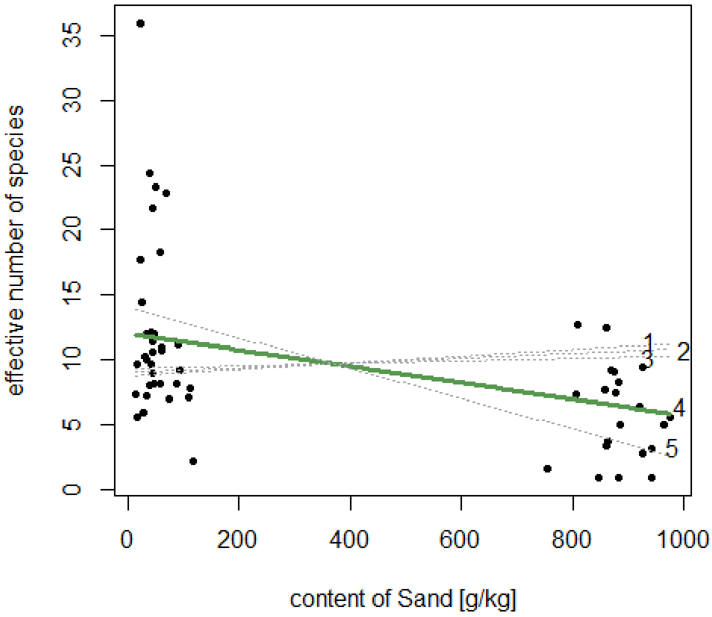


Figure A.8: Effect of soil type on biodiversity. For explanation, see Figure A.7. Bold green lines are total, dashed blue direct and dotted gray indirect effects

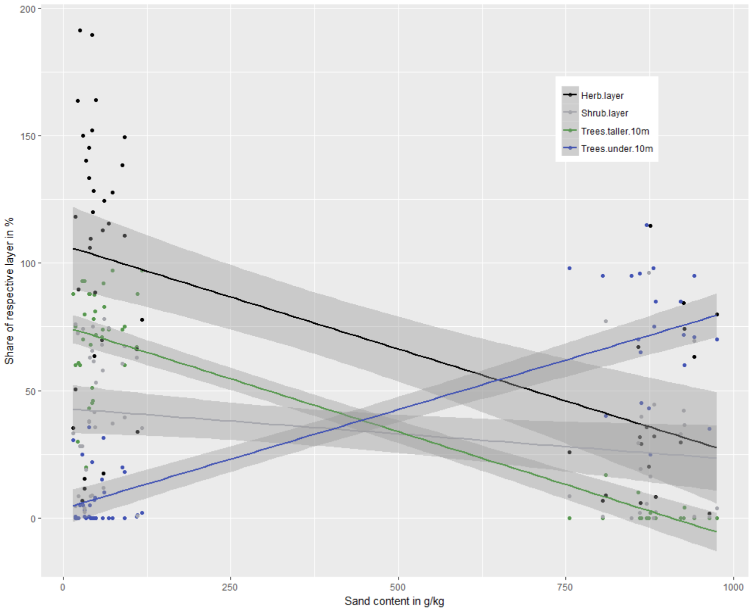
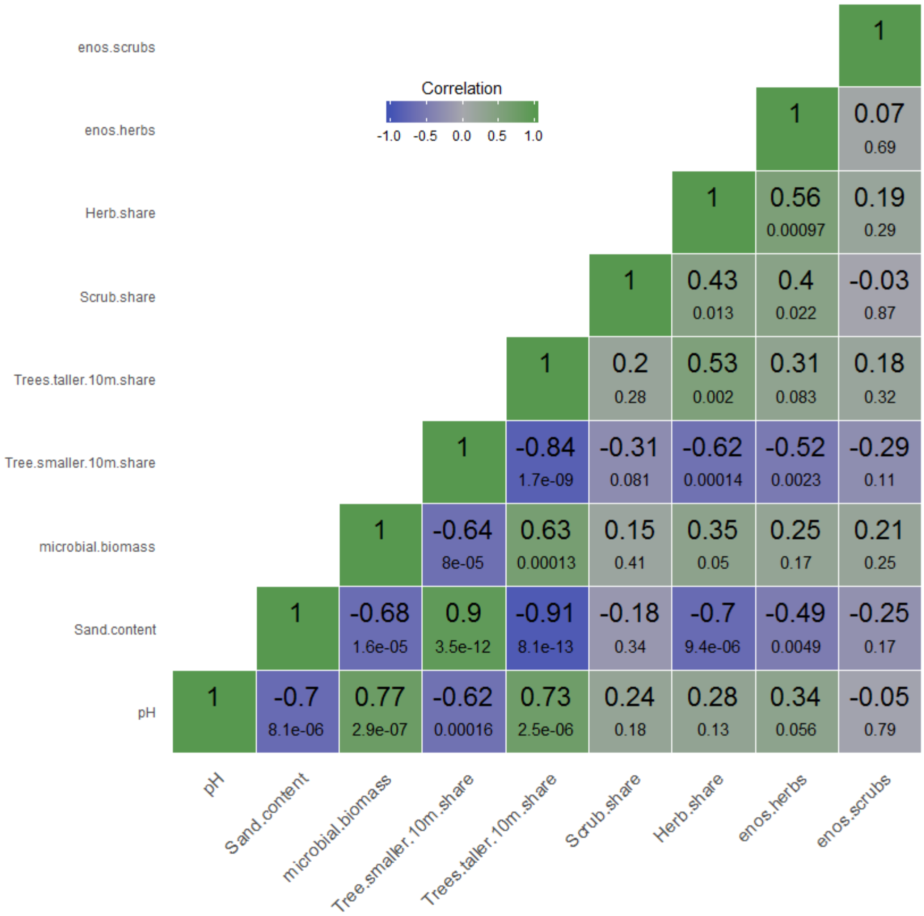


Figure A.9: Effects of sand content in soil on forest layer share.



	pH	Sand.content	microbial.biomass	Tree.smaller.10m.share	Trees.taller.10m.share	Scrub.share	Herb.share	enos.herbs	enos.scrubs
enos.scrubs									1
enos.herbs								1	0.07 0.69
Herb.share							1	0.56 0.00097	0.19 0.29
Scrub.share						1	0.43 0.013	0.4 0.022	-0.03 0.87
Trees.taller.10m.share					1	0.2 0.28	0.53 0.002	0.31 0.083	0.18 0.32
Tree.smaller.10m.share				1	-0.84 1.7e-09	-0.31 0.081	-0.62 0.00014	-0.52 0.0023	-0.29 0.11
microbial.biomass			1	-0.64 8e-05	0.63 0.00013	0.15 0.41	0.35 0.05	0.25 0.17	0.21 0.25
Sand.content		1	-0.68 1.6e-05	0.9 3.5e-12	-0.91 8.1e-13	-0.18 0.34	-0.7 9.4e-06	-0.49 0.0049	-0.25 0.17
pH	1	-0.7 8.1e-06	0.77 2.9e-07	-0.62 0.00016	0.73 2.5e-06	0.24 0.18	0.28 0.13	0.34 0.056	-0.05 0.79

Figure A.10: Correlation matrix of soil type and parameter and ENS.

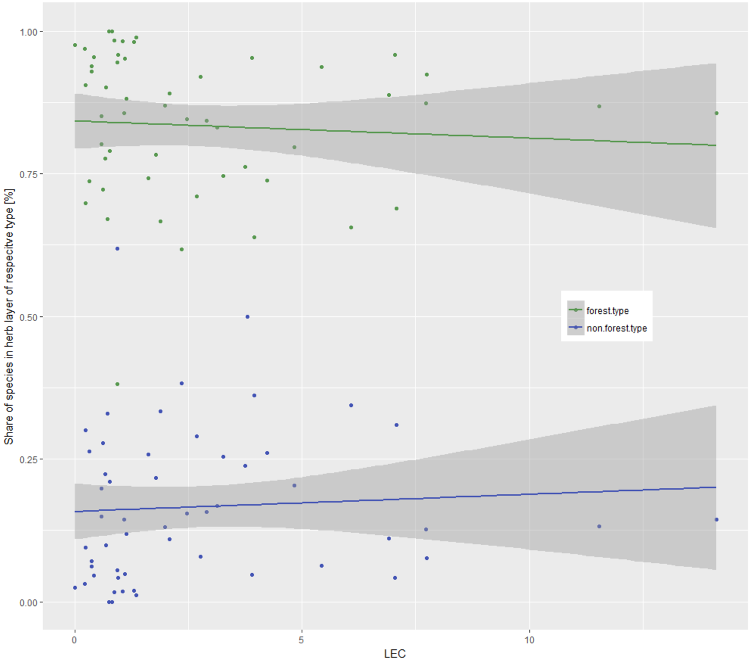


Figure A.11: Non-Significant regressions of LEC on non/forest specialist share among plant species in herb layer.

Appendix B

R-Code

B.1 Initialisation

```
############
## 1.INIT ##
############

#### 1.1 Packages ####

library(Amelia)
library(GGally)
library(Hmisc)
library(MASS)
library(car)
library(caret)
library(curl)
library(data.table)
library(distdrawr)
library(effects)
library(faoutlier)
library(gdata)
library(ggplot2)
library(ggrepel)
library(grid)
library(gridExtra)
library(influence.SEM)
library(jtools)
library(klaR)
library(lavaan)
library(lubridate)
library(mctest)
library(mgcv)
library(minpack.lm)
library(nlme)
library(nls2)
library(outliers)
library(piecewiseSEM)
library(plotly)
library(polycor)
library(psych)
library(readxl)
library(reshape2)
library(semPlot)
library(semTools)
library(svMisc)
library(tibble)
library(tidyverse)
library(vcd)
library(xlsx)

#### 1.2 Used functions and scripts  ####
```

```

# not in - operator
'%!in%'<-function(x,y)!('%in%'(x,y))

# get unix path from windows path
writeClipboard(gsub(”““““”, ”/”, readClipboard()))

get_cor_heatmap<-function(df,remove_c=NULL,remove_p=NULL, debug=F){

  ##################################################
  # df = dataframe of correlations between variables
  # remove_c = correlations (no.) to be dropped
  # remove_p = p-values to be dropped
  # debug = prints mat. to examine drops
  # Patchwork function from different howtos (links)
  # creates triangle shaped matrix of correlations
  # in heatmap style with p-values
  # commented loop to rename p values doesn't work
  ##################################################

  #http://www.sthda.com/english/wiki/ggplot2-quick-correlation-matrix-heatmap-r-
  #software-and-data-visualization

  get_upper_tri<-function(cormat){
    # function to get upper triangle of the cormat
    cormat[lower.tri(cormat)]<-NA
    return(cormat)
  }

  get_upper_tri_p<-function(cormat){
    # function to get upper triangle of the cormat
    cormat[lower.tri(cormat)]<-NA
    return(cormat)
  }

  #extract it
  upper_tri<-round(get_upper_tri(df$r),digits=2)
  upper_tri_p<-signif(get_upper_tri_p(df$P),digits=2)
  rownames(upper_tri_p)<-rownames(upper_tri)
  colnames(upper_tri_p)<-colnames(upper_tri)

  #melt
  melted_cormat<-melt(upper_tri, na.rm = TRUE)
  melted_cormat_p<-melt(upper_tri_p, na.rm = TRUE)
  if(debug==T){
    print(melted_cormat)
    print(melted_cormat_p)
  }

  #remove irrelevant
  if(!(is.null(remove_c))) melted_cormat<-melted_cormat[-remove_c,]
  if(!(is.null(remove_p))) melted_cormat_p<-melted_cormat_p[-remove_p,]

  #heatmap of correlations
  ggheatmap<-ggplot(melted_cormat, aes(Var2, Var1, fill = value))+
    geom_tile(color = ”white”)+
    scale_fill_gradient2(low = ”#4253b4”, high = ”#57984f”, mid = ”#a6a7ad”,
                         midpoint = 0, limit = c(-1,1), space = ”Lab”,
                         name=”Correlation”) +
    theme_minimal()+ # minimal theme
    theme(axis.text.x = element_text(angle = 45, vjust = 1,
                                     size = 12, hjust = 1))+
    coord_fixed()

  #fill with values
  ggheatmap +
    geom_text(aes(Var2, Var1, label = value), color = ”black”, size = 6,
              position=position_nudge(y=0.2)) +
    geom_text(data=melted_cormat_p ,aes(label = value), color = ”black”,
              size = 4, position=position_nudge(y=-0.2))+
```

```
    theme(
      axis.title.x = element_blank(),
      axis.title.y = element_blank(),
      panel.grid.major = element_blank(),
      panel.border = element_blank(),
      panel.background = element_blank(),
      axis.ticks = element_blank(),
      legend.justification = c(1, 0),
      legend.position = c(0.5,0.8),
      legend.direction = ”horizontal”)+
    guides(fill = guide_colorbar(barwidth = 7, barheight = 1,
                                  title.position = ”top”, title.hjust = 0.5))
}

multiplot<-function(..., plotlist=NULL, file, cols=1, layout=NULL) {

  #http://www.cookbook-r.com/Graphs/Multiple_graphs_on_one_page_(ggplot2)/

  # Make a list from the ... arguments and plotlist
  plots<-c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use ’cols’ to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout<-matrix(seq(1, cols * ceiling(numPlots/cols)),
                      ncol = cols, nrow = ceiling(numPlots/cols))
  }

  if (numPlots==1) {
    print(plots[[1]])

  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx<-as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                       layout.pos.col = matchidx$col))
    }
  }
}

# reader function

ids<-read.csv(paste(getwd(),”/id.csv”,sep=””))[,2:3]
wd<-paste(getwd(),”/2 merge/”,sep=””)

exclude<-c(”exp.xlsx”,”1580_allgemeine Daten Exploratorien.xlsx”,
             ”6240_MatrixData.xlsx”,
             ”22506_Light measurements in MIP forest plots_1.1.1.xlsx”,
             ”species_forest_specialist_FLORAWEB.xlsx”,”20366.veg_light.xlsx”,
             ”22506_Light measurements in MIP forest plots_1.1.1.xlsx”,
             ”20366_Vegetation Records for Forest EPs, 2016.xlsx”,
             ”df_floraweb.xlsx”,”df_floraweb_MANUALLY_ADDED_MISSING.xlsx”,
             ”species_forest_specialist_FLORAWEB.xlsx”,
             ”species_forest_specialist_FLORAWEB_MANUALLY_ADDED_MISSING.xlsx”)

drop<-c(”Exploratory”, ”EP”, ”SeasonYear”, ”Sampling_from”, ”Sampling_till”,
          ”Sampling_by”, ”Forest_PI”, ”Forest_type_in_detail”,
          ”Editor_summer_2009”, ”Editor_spring_2009”, ”Date_summer_2009”,
          ”Date_spring_2009”, ”HEW”, ”SEW”,”AEW”, ”Plot”, ”Plotid”,
```

```
          ”Type”,”id”,”rwGP”,”hwGP”,”Bemerkungen”, ”Schutzgebietsstatus”,
          ”Duengung”, ”Schnitte”, ”Tiere_aggr”, ”LandnutzungGrld”,
          ”commentsForest”,”BSS”,”BST”,”BES”,”commentsSoil”,”HW”,”RW”,”Area”,
          ”VIP”,”Exploratorium”, ”StComp_quality”, ”StComp_Area”, ”TYP”,
          ”commentsSoil”, ”V_EP”, ”MTS_V”,”NA.”,NA,”NA”,”¡NA¿”,”EP_ID”,
          ”Date_spring_2010”,”Date_summer_2010”,”Editor_spring_2010”,
          ”Editor_summer_2010”,”ATS2_d50”,”ATS2_d100”,”ATS2_dg”,”ATS2_G”,
          ”ATS2_N”,”ATS2_V”,”mean_ell_B2”,”mean_ell_B1”, ”mean_ell_sum”,
          ”ATS2_pG”,”ATS2_pArea”,”ATS21”,”ATS1_d50”,”ATS1_d100”,”ATS1_dg”,
          ”ATS1_V”, ”mean_ell_S”, ”Forest_Type”,”ATS1_V”,”ATS1_G”,”ATS1_N”,
          ”ATS1_pG”,”ATS1_pArea”,”ATS11”,”MTS_d50”,”MTS_d100”,”MTS_dg”)

# Explanation:
# Used files must be in .xlsx format
# Used files must be in subfolder in wd
# if no metadata is within file, it should be added
# in exp.xlsx in respecive folder


get_table<-function(wd=getwd(), ids=ids, exclude=exclude, drop, fill_na=T,
                      save=”file”){
  #################################################
  # wd: string of path to dir of dirs of files
  # ids: data.frame of ep id translation
  # exclude: vector of tables to drop
  # drop: columns to be dropped
  # fill_na: fill missing values with nas
  # save: save to ’file’, in ’object’ or ’both’
  #################################################

  #################################################
  ### get file paths
  files<-numeric()
  manual_exp<-logical()
  c<-1
  dirs<-list.dirs(wd)
  for (i in 1:length(dirs)){
    if(grepl(”temp output”,dirs[i])){
      dirs<-dirs[-i]
    }
  }
  for (i in 2:length(dirs)){

    #iterate through dirs in wd
    #check if meta data is in file or external (manual_exp)

    dirfiles<-list.files(dirs[i], pattern=”.xlsx”)
    ifelse((”exp.xlsx” %in% dirfiles), manual_exp[i-1]<-TRUE,
           manual_exp[i-1]<-FALSE)
    for (j in 1:length(dirfiles)){

      #iterate through files in current dir
      #ignore excluded (exclude)
      #write full path in list (files)

      if(dirfiles[j] %!in% exclude){
        files[c]<-paste(dirs[i],’/’,dirfiles[j],sep=””)
        c<-c+1
      }
    }
  }
  #################################################

  #################################################
  #### read files
  for (f in 1:length(files)){

    #iterate through files
    #check if meta data is manual or within the file
    #read file
```

```
    #add EP_Plotid if missing
    #drop unwanted columns
    #merge
    #print progress
    #fill nas (empty, and ”X”)

    si<-2
    if(manual_exp[f]==TRUE) si=1
    new_df<-read.xlsx(files[f],sheetIndex=si,stringsAsFactors = F)
    if(”EP_Plotid” %!in% colnames(new_df)){
      new_df<-merge(ids, new_df, by.x=”Plotid”, by.y=”Plot_ID”)
    }
    new_df<-new_df[, !(names(new_df) %in% drop)]
    ifelse(f¿1, df<-merge(new_df, df, by=”EP_Plotid”), df<-new_df)
    print(paste(”Table”,f,”of”,length(files),”loaded”,
                round(f/length(files)*100,digits=1),”%”))
    print(paste(”[”,paste(rep(”#”,f),collapse=””),
                paste(rep(”-”,length(files)-f),collapse=””),”]”,sep=””))
  }
  if(fill_na==T){
    df<-data.frame(apply(df,1:2,function(x) if( x %in% c(””,” ”,”X”))
      return(NA) else return(x)))
  }
  #################################################

  #################################################
  ### output ######################################
  if(save==”file” — save==”both”){
    rownames(df)<-df[,1]
    df[,1]<-NULL
    write.xlsx(df[-1,],paste(wd,”/”,”Forest_EPs.xlsx”,sep=””))
    print(paste(”Table saved in”,wd))
  }
  if(save==”object” — save==”both”){
    #test
    df<-as.data.frame(lapply(df, Hmisc::all.is.numeric, what = ”vector”,
                               extras = NA))
    #test
    assign(”Forest_EPs”,df,envir=globalenv())
    print(”Table stored/attached in object ’Forest_EPs’”)
    attach(Forest_EPs)
  }
  #################################################
}

# meta reader

exclude_r<-c(”exp.xls”,”1580_allgemeine Daten Exploratorien.xlsx”,
               ”6240_MatrixData.xlsx”)
drop_r<-drop

# Explanation:
# Used files must be in .xlsx format
# Used files must be in subfolder in wd
# if no metadata is within file, exp.xlsx in
# respecive folder manually from website

get_meta<-function(wd=getwd(), exclude_r, drop_r, save=T){
  #################################################
  # wd: string of path to dir of dirs of files
  # exclude: vector of tables to drop
  # drop_r: rows to be dropped
  # save: T: save output as csv in wd, F: output in console
  #################################################

  #################################################
  ### get file paths
  files<-numeric()
  c<-1
  dirs<-list.dirs(wd)
```

```
  for(i in 2:length(dirs)){

    #iterate through dirs in wd
    #check if meta data is in file or external
    #save respecitive file path/paths

    dirfiles<-list.files(dirs[i], pattern=”.xlsx”)
    if(”exp.xlsx” %in% dirfiles){
      files[c]<-paste(dirs[i],”/exp.xlsx”,sep=””)
      c<-c+1
    }
    else{
      for(j in 1:length(dirfiles)){

        #iterate through files in current dir
        #ignore excluded (exclude)
        #write full path in list (files)

        if(dirfiles[j] %!in% exclude){
          files[c]<-paste(dirs[i],’/’,dirfiles[j],sep=””)
          c<-c+1
        }
      }
    }
  }
  #################################################

  #################################################
  #### read files
  for (f in 1:length(files)){

    #iterate through files
    #read file
    #only keep wanted columns
    #remove NA rows
    #drop unwated rows
    #stack data frames
    #drop duplicates

    new_exp<-read.xlsx(files[f],sheetIndex=1,stringsAsFactors = F)
    new_exp<-new_exp[, names(new_exp) %in% c(”name”,”typeOfVariable”,”units”
                                              ,”description”)]
    if(is.na(new_exp[nrow(new_exp),2])) new_exp<-new_exp[-(nrow(new_exp)),]
    rownames(new_exp)<-new_exp[,1]
    new_exp[,1]<-NULL
    new_exp<-new_exp[which(rownames(new_exp) %!in% drop_r), ]
    ifelse(f¿1, exp<-rbind(exp,new_exp) , exp<-new_exp)
  }
  exp<-exp[!duplicated(exp),]
  #################################################

  #################################################
  ### output #####################################
  if(save==T){
    write.xlsx(exp,paste(wd,”/”,”Forest_EPs_meta.xlsx”,sep=””))
    print(paste(”Table saved in”,wd))
  }
  else{
    return(exp)
  }
  #################################################
}

# call reader
get_table(wd, ids, exclude, drop, save=”object”)
get_meta(wd, exclude_r, drop_r, save=T)
```

code/6_1_final_init.R

B.2 Data

```
############
## 2.DATA ##
############

#### 2.1 LEC correted ####

# uncorrected mean LEC and raw light data

light<-fread(paste0(getwd(),”/22506.txt”))
light<-light[order(PlotID)]

attach(light)

mean_LEC<-aggregate(abs(LEC) PlotID,light,mean, na.action = na.omit)
mean_RefData<-aggregate(abs(RefData) PlotID,light,mean, na.action = na.omit)
mean_LineData<-aggregate(abs(LineData) PlotID,light,mean, na.action = na.omit)
light<-merge(mean_LEC,merge(mean_RefData,mean_LineData))
names(light)<-c(”EP_Plotid”,”mean_LEC”,”mean_RefData”,”mean_LineData”)

light[,2:4]<-round(light[,2:4],8)

write.xlsx(light,paste(wd_temp,”/22506.light.xlsx”,sep=””))

# corrected LEC

# tree cover plot Plot

Tree_cover<-(B2_fs+B2_nfs+B1_fs+B1_nfs)

df_nls<-data.frame(Tree_cover, cover_sum_woody_species_S_B1_B2,
                     cover_sum_trees_B1_B2, mean_LEC)
x<-seq(0:140)

plot(mean_LEC Tree_cover, pch=”.”, col=”#57984f”, cex=3,
     xlab=”Sum of vegetation cover in canopy”,
     ylab=”mean Light Extinction Coeff.”)
nls(mean_LEC -b*log(Tree_cover)+a, start=list(a=1,b=1), data=df_nls)
fit1<--0.06409*log(x)+0.32360
lines(fit1, lwd=3, col=”#57984f”)

points(mean_LEC cover_sum_woody_species_S_B1_B2, pch=”.”, col=”#4253b4”, cex=2)
nls(mean_LEC -b*log(cover_sum_woody_species_S_B1_B2)+a, start=list(a=1,b=1),
    data=df_nls)
fit2<--0.04779*log(x)+0.26275
lines(fit2, col=”#4253b4”)

points(mean_LEC cover_sum_trees_B1_B2, pch=”.”, col=”#a6a7ad”, cex=2)
nls(mean_LEC -b*log(cover_sum_trees_B1_B2)+a, start=list(a=1,b=1), data=df_nls)
fit3<--0.06991*log(x)+0.34263
lines(fit3, col=”#a6a7ad”)

legend(x=”topright”,legend=c(”calculated values”,
                             ”given covers of trees and shrubs”,
                             ”given values of trees”),
       col=c(”#57984f”,”#4253b4”,”#a6a7ad”), lwd = 5)

#model of corrected fit

df_nls<-data.frame(mean_LineData, mean_RefData)

x<-seq(0:2000)

plot(mean_LineData mean_RefData, pch=1, cex=0.5, xlab=”mean Light above canopy”,
     ylab=”mean Light below canopy”)

mod<-nls(log(mean_LineData) log(a)+b*mean_RefData, start=list(a=1, b=1),
         data=df_nls)
fit8<-coef(mod)[1]*exp(x*coef(mod)[2])
```

```

lines(fit8, lwd=2)
mean_LEC_corr<-mean_LineData/(coef(mod)[1]*exp(mean_RefData*coef(mod)[2]))


#validating fit

plot(mod)

df_cor<-data.frame(”LEC”=mean_LEC, ”LEC corrected”=mean_LEC_corr,
                     ”RefData”=mean_RefData, ”LineData”=mean_LineData)
cor_mat<-rcorr(as.matrix(df_cor), type=”pearson”)
get_cor_heatmap(cor_mat)

write.xlsx(data.frame(EP_Plotid,mean_LEC_corr),
           paste(wd,”/00001_LEC_corr/mean_LEC_corr.xlsx”,sep=””))
#### 2.2 Forest specialist species and ELV ####

’######################################################’
’# STEP 1
’# Get unique species and FloraWebIndex via url
’######################################################’

species<-fread(paste0(getwd(),”/20366.txt”))
species_unique<-unique(species$Species)
species_list<-list()
number<-list()
for(i in 1:length(species_unique)){

  ################################
  # quick and very very dirty
  # iterate through all species
  # and use distdrawr::check_species
  # to find floraweb number
  # (dealt with many exceptions)
  ################################

  if(grepl(”sp.”,species_unique[i])){
    species_list[i]<-species_unique[i]
    number[i]<-NA
    next
  }
  if(grepl(”agg.”,species_unique[i])){
    if(grepl(”aggr.”,species_unique[i])){
      species_list[i]<- strsplit(species_unique[i], split=’_aggr.’,
                                   fixed=TRUE)
      number[i]<-check_species(data.frame(species_list[i]),
                                 level=”species”)[1,1]
      next
    }
    species_list[i]<- strsplit(species_unique[i], split=’_agg.’, fixed=TRUE)
    if(grepl(”_fruticosus”,species_unique[i])){
      species_list[i]<-gsub(”_fruticosus”,””,species_list[i])
      number[i]<-check_species(data.frame(species_list[i]),
                                 level=”species”)[1,1]
      next
    }
    if(grepl(”_riviniana”,species_unique[i])){
      species_list[i]<-gsub(”_riviniana”,””,species_list[i])
      number[i]<-check_species(data.frame(species_list[i]),
                                 level=”species”)[1,1]
      next
    }
    if(grepl(”_veris”,species_unique[i])){
      species_list[i]<-gsub(”_veris”,””,species_list[i])
      number[i]<-check_species(data.frame(species_list[i]),
                                 level=”species”)[1,1]
      next
    }
    number[i]<-check_species(data.frame(species_list[i]),
```

```
                level=”species”)[1,1]
  next
 }
 if(grepl(”_x_”,species_unique[i])){
  species_list[i]<- gsub(”_x_”,”_”,species_unique[i])
  number[i]<-check_species(data.frame(species_list[i]),
                level=”species”)[1,1]
  next
 }
 species_list[i]<-species_unique[i]
 number[i]<-check_species(data.frame(species_list[i]),
               level=”species”)[1,1]
}

#voila: species with FloraWebIndex
df_floraweb<-data.frame(”Species”=unlist(species_list), ”FWI”=unlist(number))

print(paste(round(sum(is.na(df_floraweb$FWI))/nrow(df_floraweb)*100,digits=1),
      ”% of Species not found in FloraWeb”,sep=””))

'##############################'
'# adding missing FWI manually #
'##############################'

write.xlsx(df_floraweb, paste0(getwd(),”/df_floraweb.xlsx”))
df_floraweb<-read.xlsx(paste0(getwd(),
                ”/df_floraweb_MANUALLY_ADDED_MISSING.xlsx”),1)

'######################################################'
'# STEP 2:
'# get forest habitat data from curl
'######################################################'

Wald<-list()
options(warn=-1)
for(j in 1:nrow(df_floraweb)){

 ##############################
 # again quick and very very dirty
 # iterate through previous df
 # find and manipulate html string
 # list it up
 ##############################

 tmp<-tempfile()
 ifelse(is.na(df_floraweb$FWI[j]),
     Wald[j]<-NA,
     tmp_file<-curl_download(paste(
      ”http://floraweb.de/pflanzenarten/oekologie.xsql?suchnr=”,
      df_floraweb$FWI[j],”&”,sep=””),tmp)
 )
 temp_output<-readLines(tmp_file)

 if(!(is.na(Wald[j]))){
  for(i in 1:length(temp_output)){
   if(grepl(”Bindung an Wald”,temp_output[i])){
    if(grepl(”BERGLAND : fehlt”,temp_output[i])){
     Wald[j]<-gsub(”s”,”ss”,
             gsub(”¡/p¿”,””,
                gsub(”ALPEN : ”,””,
                  unlist(strsplit(temp_output[i],”¡br¿”))[2])))
    } else if(grepl(”TIEFLAND : fehlt”,temp_output[i])){
     Wald[j]<-gsub(”s”,”ss”,
             gsub(”¡/p¿”,””,
                gsub(”BERGLAND : ”,””,
                  unlist(strsplit(temp_output[i],”¡br¿”))[3])))
    } else if(grepl(”keine”,temp_output[i+1])){
     Wald[j]<-”keine Waldart”
    } else{
     Wald[j]<-gsub(”¡/p¿”,””,
```

```
                          gsub(”TIEFLAND : ”,””,
                               unlist(strsplit(temp_output[i],”¡br¿”))[4]))
        }
      }
    }
  }
}
#Don’t forget to turn warnings back on!
options(warn=0)

#voila
df_floraweb$Beschreibung <- unlist(Wald)

’#####################################################’
’# STEP 3:
’# format forest habitat type
’# output
’# write to file
’#####################################################’

Typ <- list()
Schicht <- list()
Bemerkung <- list()
for(i in 1:nrow(df_floraweb)){
  if(!(is.na(df_floraweb$Beschreibung[i]))){

    if(df_floraweb$Beschreibung[i]==”keine Waldart”){
      Typ[i] <- ”kein Waldspezialist”
      Bemerkung[i] <- NA
    }else{
      Typ[i] <- ”Waldspezialist”
    }


    if(grepl(”Krautschicht”,df_floraweb$Beschreibung[i])){
      Schicht[i] <- ”Krautschicht”
      Bemerkung[i] <- gsub(”Krautschicht: ”,””,df_floraweb$Beschreibung[i])
    } else if(grepl(”Strauchschicht”,df_floraweb$Beschreibung[i])){
      Schicht[i] <- ”Strauchschicht”
      Bemerkung[i] <- gsub(”Strauchschicht: ”,””,df_floraweb$Beschreibung[i])
    } else if(grepl(”Baumschicht”,df_floraweb$Beschreibung[i])){
      Schicht[i] <- ”Baumschicht”
      Bemerkung[i] <- gsub(”Baumschicht: ”,””,df_floraweb$Beschreibung[i])
    } else{
      Schicht[i] <- NA
    }
  } else{
    Bemerkung[i] <- NA
    Typ[i] <- NA
    Schicht[i] <- NA
  }
}

df_floraweb$Bemerkung <- unlist(Bemerkung)
df_floraweb$Typ <- unlist(Typ)
df_floraweb$Schicht <- unlist(Schicht)

df_floraweb <- subset(df_floraweb, select = -c(Beschreibung) )

write.xlsx(df_floraweb, paste0(getwd(),
                               ”/species_forest_specialist_FLORAWEB.xlsx”))


’#####################################################’
’# ADDITIONAL ELLENBERG CHECK:
’# Same as Step 2
’# collecting Ellenberg lightvalue instead
’#####################################################’
Lichtzahl <- list()
options(warn=-1)
```

```
for(j in 1:nrow(df_floraweb)){

  ################################
  # again quick and very very dirty
  # iterate through previous df
  # find and manipulate html string
  # list it up
  ################################

  tmp<-tempfile()
  ifelse(is.na(df_floraweb$FWI[j]),
         Lichtzahl[j]<-NA,
         tmp_file<-curl_download(paste(
           ”http://floraweb.de/pflanzenarten/oekologie.xsql?suchnr=”,
           df_floraweb$FWI[j],”&”,sep=””),tmp)
  )
  temp_output<-readLines(tmp_file)

  if(!(is.na(Lichtzahl[j]))){
    for(i in 1:length(temp_output)){
      if(grepl(”Lichtzahl”,temp_output[i])){
        temp_output[i]<-strsplit(temp_output[i],”¡a href”)[[1]][2]
        if(grepl(”keine Angaben”, temp_output[i])){
          Lichtzahl[j]<-NA
        } else if(grepl(”indifferent”, temp_output[i])){
          Lichtzahl[j]<-NA
        } else{
          Lichtzahl[j]<-regmatches(unlist(temp_output[i]),
                                     gregexpr(”[[:digit:]]+”, unlist(temp_output[i])))
        }
      }
    }
  }
}
#Don’t forget to turn warnings back on!
options(warn=0)

Lichtzahl<-unlist(Lichtzahl)

df_floraweb<-cbind(df_floraweb,as.data.frame(Lichtzahl))

write.xlsx(df_floraweb, paste0(getwd(),
                               ”/species_forest_specialist_FLORAWEB.xlsx”))

’########################################’
’# adding missing light values manually #’
’########################################’
sum(is.na(df_floraweb$Lichtzahl))
df_Species<-read.xlsx(paste0(
  getwd(),”/species_forest_specialist_FLORAWEB_MANUALLY_ADDED_MISSING.xlsx”),1)
sum(is.na(df_Species$Lichtzahl))

’#######################################################’
’# STEP 4:
’# Data ->EPs
’#######################################################’

veg_data<-fread(paste0(getwd(),”/20366.txt”))

veg_data<-merge(veg_data[veg_data$Year==2016], df_Species, by=”Species”)

colnames(veg_data)[6]<-”EP_Plotid”

EP_Plotid<-unique(veg_data$EP_Plotid)
B2_fs<-numeric()
B1_fs<-numeric()
S_fs<-numeric()
H_fs<-numeric()
sum_fs<-numeric()
B2_nfs<-numeric()
```

```
B1_nfs<-numeric()
S_nfs<-numeric()
H_nfs<-numeric()
sum_nfs<-numeric()
B2_ell_weighted<-numeric()
B1_ell_weighted<-numeric()
S_ell_weighted<-numeric()
H_ell_weighted<-numeric()


for(i in 1:length(EP_Plotid)){

  id = EP_Plotid[i]
  B2_fs[i]<-sum(veg_data[veg_data$EP_Plotid==id &
                             veg_data$Typ==”Waldspezialist” &
                             veg_data$Layer==”B2”]$cover)
  B1_fs[i]<-sum(veg_data[veg_data$EP_Plotid==id &
                             veg_data$Typ==”Waldspezialist” &
                             veg_data$Layer==”B1”]$cover)
  S_fs[i]<-sum(veg_data[veg_data$EP_Plotid==id &
                            veg_data$Typ==”Waldspezialist” &
                            veg_data$Layer==”S”]$cover)
  H_fs[i]<-sum(veg_data[veg_data$EP_Plotid==id &
                            veg_data$Typ==”Waldspezialist” &
                            veg_data$Layer==”H”]$cover)
  sum_fs[i]<-sum(B2_fs[i],B1_fs[i],S_fs[i],H_fs[i])


  B2_nfs[i]<-sum(veg_data[veg_data$EP_Plotid==id &
                              veg_data$Typ==”kein Waldspezialist” &
                              veg_data$Layer==”B2”]$cover)
  B1_nfs[i]<-sum(veg_data[veg_data$EP_Plotid==id &
                              veg_data$Typ==”kein Waldspezialist” &
                              veg_data$Layer==”B1”]$cover)
  S_nfs[i]<-sum(veg_data[veg_data$EP_Plotid==id &
                             veg_data$Typ==”kein Waldspezialist” &
                             veg_data$Layer==”S”]$cover)
  H_nfs[i]<-sum(veg_data[veg_data$EP_Plotid==id &
                             veg_data$Typ==”kein Waldspezialist” &
                             veg_data$Layer==”H”]$cover)
  sum_nfs[i]<-sum(B2_nfs[i],B1_nfs[i],S_nfs[i],H_nfs[i])


  B2_ell_weighted[i]<-sum(as.numeric(veg_data[EP_Plotid==id &
                                                  Layer==”B2” &
                                                  cover¿0]$Lichtzahl)*
                              (veg_data[EP_Plotid==id & Layer==”B2” &
                                          cover¿0]$cover/
                                 sum(veg_data[EP_Plotid==id &
                                                Layer==”B2”]$cover)),na.rm=T)
  B1_ell_weighted[i]<-sum(as.numeric(veg_data[EP_Plotid==id & Layer==”B1” &
                                                  cover¿0]$Lichtzahl)*
                              (veg_data[EP_Plotid==id &
                                          Layer==”B1” & cover¿0]$cover/
                                 sum(veg_data[EP_Plotid==id &
                                                Layer==”B1”]$cover)),na.rm=T)
  S_ell_weighted[i]<-sum(as.numeric(veg_data[EP_Plotid==id & Layer==”S” &
                                                 cover¿0]$Lichtzahl)*
                             (veg_data[EP_Plotid==id &
                                         Layer==”S” & cover¿0]$cover/
                                sum(veg_data[EP_Plotid==id &
                                               Layer==”S”]$cover)),na.rm=T)
  H_ell_weighted[i]<-sum(as.numeric(veg_data[EP_Plotid==id &
                                                 Layer==”H” & cover¿0]$Lichtzahl)*
                             (veg_data[EP_Plotid==id & Layer==”H” &
                                         cover¿0]$cover/
                                sum(veg_data[EP_Plotid==id &
                                               Layer==”H”]$cover)),na.rm=T)
  #set to NA, if 0
  if(B2_ell_weighted[i]==0) B2_ell_weighted[i]<-NA
```

```
  if(B1_ell_weighted[i]==0) B1_ell_weighted[i]<-NA
  if(S_ell_weighted[i]==0) S_ell_weighted[i]<-NA
  if(H_ell_weighted[i]==0) H_ell_weighted[i]<-NA

  #check if NA’s are covering more then 50%, if so, assign NA
  if(sum(veg_data[EP_Plotid==id & Layer == ”H”]$cover¿0)){
    if(sum(veg_data[EP_Plotid==id & Layer==”H” & cover¿0][which(is.na(
      veg_data[EP_Plotid==id & Layer==”H” & cover¿0]$Lichtzahl))]$cover)/
      sum(veg_data[EP_Plotid==id & Layer==”H”]$cover) ¿ 0.5){
      H_ell_weighted[i]<-NA
    }
  }
  if(sum(veg_data[EP_Plotid==id & Layer == ”S”]$cover¿0)){
    if(sum(veg_data[EP_Plotid==id & Layer==”S” & cover¿0][which(is.na(
      veg_data[EP_Plotid==id & Layer==”S” & cover¿0]$Lichtzahl))]$cover)/
      sum(veg_data[EP_Plotid==id & Layer==”S”]$cover) ¿ 0.5){
      S_ell_weighted[i]<-NA
    }
  }
  if(sum(veg_data[EP_Plotid==id & Layer == ”B1”]$cover¿0)){
    if(sum(veg_data[EP_Plotid==id & Layer==”B1” & cover¿0][which(is.na(
      veg_data[EP_Plotid==id & Layer==”B1” & cover¿0]$Lichtzahl))]$cover)/
      sum(veg_data[EP_Plotid==id & Layer==”B1”]$cover) ¿ 0.5){
      B1_ell_weighted[i]<-NA
    }
  }
  if(sum(veg_data[EP_Plotid==id & Layer == ”B2”]$cover¿0)){
    if(sum(veg_data[EP_Plotid==id & Layer==”B2” & cover¿0][which(is.na(
      veg_data[EP_Plotid==id & Layer==”B2” & cover¿0]$Lichtzahl))]$cover)/
      sum(veg_data[EP_Plotid==id & Layer==”B2”]$cover) ¿ 0.5){
      B2_ell_weighted[i]<-NA
    }
  }
}

Plot_Forest_specialist<-data.frame(”EP_Plotid”=EP_Plotid, ”B2_fs”=B2_fs,
                                   ”B1_fs”=B1_fs, ”S_fs”=S_fs, ”H_fs”=H_fs,
                                   ”sum_fs”=sum_fs, ”B2_nfs”=B2_nfs,
                                   ”B1_nfs”=B1_nfs, ”S_nfs”=S_nfs,
                                   ”H_nfs”=H_nfs, ”sum_nfs”=sum_nfs,
                                   ”B2_ell_weighted”=B2_ell_weighted,
                                   ”B1_ell_weighted”=B1_ell_weighted,
                                   ”S_ell_weighted”=S_ell_weighted,
                                   ”H_ell_weighted”=H_ell_weighted)


write.xlsx(Plot_Forest_specialist,paste0(getwd(),”/EP_specilist_share.xlsx”))

# Cleanup
rm(species, species_unique,number,df_floraweb,wald,Typ,Schicht,Bemerkung,
   Lichtzahl,df_Species,wd_temp,wd_temp2,wd_temp3,veg_data,
   Plot_Forest_specialist)
rm(EP_Plotid,B2_fs,B1_fs,S_fs,H_fs,sum_fs,B2_nfs,B1_nfs,S_nfs,H_nfs,sum_nfs,
   B2_ell_weighted,B1_ell_weighted,S_ell_weighted,H_ell_weighted)
# temp files moved to /temp output
```

code/6_2_final_data.R

B.3 SEM

```
###########
## 3.SEM ##
###########

#### 3.1 init ####

Forest_EPs_pure<-subset(Forest_EPs, Mixture==”pure”)
```

```
detach(Forest_EPs)
attach(Forest_EPs_pure)
ForMI[39]<-NA


#### 3.2 complex model ####

df_sem_complex<-data.frame(”mean_LEC_corr”=mean_LEC_corr,
                             ”H_ell_weighted”=H_ell_weighted, ForMI,SMI,
                             ”B2”=B2_fs, ”shannon_herbs”=exp(shannon_herbs),
                             Simpsons_herbs, ”H_fs”=H_fs,
                             ”mts”=ordered(mainTreeSpecies),
                             ”Sand”=(Fine_Sand+Coarse_Sand+Medium_Sand),
                             ”Silt”=(Fine_Silt+Medium_Silt+Coarse_Silt), Clay)

df_sem_complex_f<-data.frame(scale(df_sem_complex[names(
  df_sem_complex)[!names(df_sem_complex) == ”mts”]]))

df_sem_complex_f[”mts”]<-df_sem_complex$mts


model_sem_complex<-’
#latent variables
forest_management =  ForMI+SMI
biodiversity =  shannon_herbs + Simpsons_herbs
forest_structure =  mts
light =  mean_LEC_corr+H_ell_weighted
soil_type =  Sand+Clay+Silt
tree_layer =  B2
herb_layer =  H_fs

#regressions
herb_layer   forest_management + soil_type + light + forest_structure +
tree_layer
forest_structure   forest_management + soil_type
tree_layer   forest_structure + soil_type #+ forest_management
light   tree_layer + forest_structure
biodiversity   herb_layer + forest_management + soil_type

#co(variances)
Sand    mean_LEC_corr
Silt    mean_LEC_corr
Clay    mean_LEC_corr
B2    shannon_herbs
ForMI    H_ell_weighted
’

sem_complex<-sem(model_sem_complex, data=df_sem_complex_f, ordered = c(”mts”))
summary(sem_complex,fit.measures=T,rsquare=T)


#### 3.3 simplified model ####

df_sem_simple<-data.frame( ”forest_management”=ForMI,
                             ”tree_layer”=MTS_N,
                             ”biodiversity”=exp(shannon_herbs),
                             ”forest_structure”=ordered(mainTreeSpecies),
                             ”soil_type”=(Fine_Sand+Coarse_Sand+Medium_Sand),
                             ”herb_layer”=H_ell_weighted,
                             ”light”=mean_LEC_corr)
df_sem_simple$forest_management[39]<-NA
df_sem_simple_f<-data.frame(scale(df_sem_simple[names(
  df_sem_simple)[!names(df_sem_simple) == ”forest_structure”]]))
df_sem_simple_f[”forest_structure”]<-df_sem_simple$forest_structure


model_sem_simple<-’
#regressions
herb_layer    forest_structure + a*light
forest_structure   forest_management + soil_type
```

```
tree_layer   forest_structure + soil_type
light   forest_structure + tree_layer
biodiversity    forest_management + forest_structure + b*herb_layer
ab := a*b
'
sem_simple<-sem(model_sem_simple, data=df_sem_simple_f,
                  ordered = c(”forest_structure”))
summary(sem_simple,fit.measures=T,rsquare=T)

#### 3.4 Inspecting Outliers ####

# simple model

inf.deltaci<-Deltachi(model_sem_simple, df_sem_simple_f, scaled=T)
inf.cook<-genCookDist(model_sem_simple, df_sem_simple_f)
inf.tlirmsea<-fitinfluence(c(”tli”,”rmsea”), model_sem_simple,
                             df_sem_simple_f)

explore.influence(inf.tlirmsea$Dind$tli)
which_tli<-which(inf.tlirmsea$Dind$tli ¿ 0.015)
# −>7, 37, 52
inf.tlirmsea$Dind$tli[which(inf.tlirmsea$Dind$tli ¿ 0.015)]

explore.influence(inf.tlirmsea$Dind$rmsea)
which_rmsea<-which(inf.tlirmsea$Dind$rmsea ¡ -0.05)
# −>52

explore.influence(inf.deltaci)
which_deltaci<-which(inf.deltaci ¡ -1 — inf.deltaci ¿ 1)
# −>7, 36, 37, 43, 45, 52
inf.deltaci[which(inf.deltaci ¡ -1 — inf.deltaci ¿ 1)]

explore.influence(inf.cook)
which_cook<-which(inf.cook ¿ 5)
# −>7, 12, 36, 37, 43
inf.cook[which(inf.cook ¿ 5)]

#better plot
gof<-GOF(df_sem_simple_f, model_sem_simple, progress=T)
gcd<-gCD(df_sem_simple_f, model_sem_simple, progress=T)

plot(gof)
plot(gcd)

#inspecting outliers
suspects<-sort(unique(c(which_deltaci, which_rmsea, which_tli, which_cook)))

df_rm_outlier<-df_sem_simple_f
for (i in suspects){
  df_rm_outlier<-df_rm_outlier[-i,]
  sem_rm_outlier<-sem(model_sem_simple, data=df_rm_outlier,
                        ordered = c(”forest_structure”))
  summary(sem_rm_outlier,fit.measures=T,rsquare=T)
}

# complex model

modificationindices(sem_complex)

get_cor_heatmap(rcorr(as.matrix(dplyr::select_if(df_sem_complex, is.numeric))))

inf.deltaci<-Deltachi(model_sem_complex, df_sem_complex_f, scaled=T)
inf.cook<-genCookDist(model_sem_complex, df_sem_complex_f)
inf.tlirmsea<-fitinfluence(c(”tli”,”rmsea”), model_sem_complex,
                             df_sem_complex_f)

explore.influence(inf.tlirmsea$Dind$tli)
which(inf.tlirmsea$Dind$tli ¿ 0.01)
# −>7, 13, 14, 46
inf.tlirmsea$Dind$tli[which(inf.tlirmsea$Dind$tli ¿ 0.01)]
```

```

explore.influence(inf.tlirmsea$Dind$rmsea)
# ->no outliers

explore.influence(inf.deltaci)
which(inf.deltaci ¡ -0.5)
# ->14, 36, 37, 46, 52
inf.deltaci[which(inf.deltaci ¡ -0.5)]

explore.influence(inf.cook)
which(inf.cook ¿ 3)
# ->7,12,36,37,43
inf.cook[which(inf.cook ¿ 3)]

# different approach
gof<-GOF(df_sem_complex_f, model_sem_complex, progress=T)
gcd<-gCD(df_sem_complex_f, model_sem_complex, progress=T)

plot(gof)
plot(gcd)

#inspecting outliers
df_outlier<-df_sem_complex_f[-c(36,37,46)]
sem_outlier<-sem(model_sem_complex, data=df_outlier,
                   ordered = c(”forest_structure”))
summary(sem_outlier,fit.measures=T,rsquare=T)

#### 3.5 Pertubation ####

sem_pertub<-function(noise, n, model, data, progress=T, variable){
  ###################################
  # Function to test variable noise
  # on r 2 of dependend variables
  # noise: normally distributed noise
  # n: number of iterations
  # model : charakter string od model
  #         in laavan sem syntax
  # data: data frame of variables
  # progress: print progress of iter
  # variable: variable to test
  ###################################


  herb_layer_r2<-numeric()
  forest_structure_r2<-numeric()
  tree_layer_r2<-numeric()
  light_r2<-numeric()
  biodiversity_r2<-numeric()

  for (i in 1:n){
    # get inital data.frame
    df_ini<-data
    df_s<-data.frame(scale(df_ini[names(df_ini)[!names(df_ini) ==
                                                    ”forest_structure”]]))
    df_s[”forest_structure”]<-df_ini$forest_structure
    df<-df_s

    # put noise
    if(variable == ”forest_structure”){
      prob<-sample(c(TRUE,FALSE), 55, replace=T, prob=c(noise,1-noise))
      for (j in 1:length(df$forest_structure)){
        if(prob[j]){
          not<-df$forest_structure[j]
          df$forest_structure[j]<-sample(
            levels(df$forest_structure)[
              (levels(df$forest_structure) %!in% not)],1)
        }
      }
    }else{
      df[[variable]]<-df[[variable]]+rnorm(n=length(df[[variable]]),
```

```
                                             mean=0, sd=noise)
    }

    # assign values
    fit<-sem(model, data=df, ordered = c(”forest_structure”))
    herb_layer_r2[i]<-inspect(fit,”r2”)[1]
    forest_structure_r2[i]<-inspect(fit,”r2”)[2]
    tree_layer_r2[i]<-inspect(fit,”r2”)[3]
    light_r2[i]<-inspect(fit,”r2”)[4]
    biodiversity_r2[i]<-inspect(fit,”r2”)[5]

    #print progress
    if(progress==T){
      svMisc::progress(i)
    }
  }
  output<-data.frame(”herb_layer_r2”=herb_layer_r2,
                       ”forest_structure_r2”=forest_structure_r2,
                       ”tree_layer_r2”=tree_layer_r2,
                       ”light_r2”=light_r2,”biodiversity_r2”=biodiversity_r2)
  return(output)
}

# computing reference and pertubations
ref<-inspect(sem_11.2, ”rsquare”)
light_sens_r2<-sem_pertub(noise=0.1, n=100, model=model_11.2,
                            data=df_sem11.2_f, variable=”light”)
tree_layer_sens_r2<-sem_pertub(noise=0.1, n=100, model=model_11.2,
                                 data=df_sem11.2_f, variable=”tree_layer”)
herb_layer_sens_r2<-sem_pertub(noise=0.1, n=100, model=model_11.2,
                                 data=df_sem11.2_f, variable=”herb_layer”)
soil_type_sens_r2<-sem_pertub(noise=0.1, n=100, model=model_11.2,
                                data=df_sem11.2_f, variable=”soil_type”)
forest_management_sens_r2<-sem_pertub(noise=0.1, n=100, model=model_11.2,
                                        data=df_sem11.2_f,
                                        variable=”forest_management”)


# matrix of percentage of effect of pertubation
effectmat<-matrix(0,5,5)
for(i in 1:length(ref)){
  effectmat[1,i]<-(1-mean(light_sens_r2[,i])/ref[i])*100
  effectmat[2,i]<-(1-mean(tree_layer_sens_r2[,i])/ref[i])*100
  effectmat[3,i]<-(1-mean(herb_layer_sens_r2 [,i])/ref[i])*100
  effectmat[4,i]<-(1-mean(soil_type_sens_r2[,i])/ref[i])*100
  effectmat[5,i]<-(1-mean(forest_management_sens_r2[,i])/ref[i])*100
}

# aking avay effects from variable to itself
effectmat[1,4]<-NA
effectmat[2,3]<-NA
effectmat[3,1]<-NA

# plot
par(mfrow=c(3,2), mar = c(1,2,1,1)+ 1,oma = c(1.5,1.5,0,0) + 1)
main_name<-c(”light”, ”tree layer”, ”herb layer”, ”soil type”,
             ”forest management”)
for (i in 1:nrow(effectmat)){
  barplot(effectmat[i,], ylim=c(-5,5),  col=c(”#57984f”),
          border=0, main=paste(”noise in”,main_name[i]), space=0,
          names.arg = c(”HL”,”FS”,”TL”,”LI”,”BD”),
          beside=T)
  abline(h=0)
}
mtext(”Effects strength in % change on explained variance” , side=2,
      line=3, at=10)
plot.new()
legend(”topleft”,legend=c(”model b”, ”model a”), col=c(”#57984f”,”#4253b4”),
       pch=15)

```

```
#### 3.6 Prediction ####

# approach with sem.predict()

# altered sem.predict() function, that allows regressions from nnet:multinom()
sem.predict1 = function(object, newdata, sefit = FALSE, ...) {

  # If not a model object, then coerce to a list
  if(class(object) != ”list”) object = list(object)

  # Isolate model(s) in the model list that contain the predictors in newdata
  new.x = colnames(newdata)

  x.vars = suppressWarnings(sapply(object, function(i) any(
    all.vars(formula(i))[-1] %in% new.x)))

  if(!any(x.vars))

    stop(”No variables in new data are found in model(s)!”) else

      x.modelList = object[x.vars]

  # Send newdata to each model in the model list and return output
  # as a data.frame
  predict.df = do.call(

    cbind, lapply(x.modelList, function(i) {

      # Get model predictions
      if(any(class(i) %in% c(”lm”, ”rq”, ”glm”, ”neg.bin”, ”gls”, ”pgls”,
                             ”multinom”)))

        predict.df = predict(i, newdata, se.fit = sefit, ...) else

          if(any(class(i) %in% c(”lme”, ”glmmPQL”)))

            predict.df = predict(i, newdata, level = 0, ...) else

              if(any(class(i) %in% c(”lmerMod”, ”glmerMod”, ”merModTest”)))

                predict.df = predict(i, newdata, re.form = NA, ...)

              # If se.fit = TRUE for mixed models, calculate standard
              # errors based on fixed-effects only
              if(sefit == TRUE & any(class(i) %in% c(”lme”, ”glmmPQL”,
                                                     ”lmerMod”, ”glmerMod”,
                                                     ”merModTest”))) {

                # Bind in predictions to new data
                newdata = data.frame(newdata, predict.df)

                colnames(newdata)[ncol(newdata)] = all.vars(formula(i))[1]

                if(any(class(i) %in% c(”lme”, ”glmmPQL”))) {

                  # Dmat.lme = model.matrix(formula(i)[-2], newdata)

                  pvar = sqrt(diag(model.matrix(formula(i)[-2], newdata)
                                   %*% vcov(i) %*% t(model.matrix(
                                     formula(i)[-2], newdata) )))

                } else {

                  # Dmat.lmer = model.matrix(terms(i), newdata)

                  pvar = sqrt(diag(model.matrix(terms(i), newdata) %*%
                                     tcrossprod(vcov(i), model.matrix(
                                       terms(i), newdata))))

                }
```

```

                # Return list with predicted errors
                predict.df = list(fit = predict.df, se.fit = pvar)

              }

              # If predictions are stored in a list, bind columns
              if(class(predict.df) == ”list”) predict.df =
                  do.call(data.frame, predict.df[1:2]) else

                    predict.df = data.frame(predict.df)

                  # Name columns
                  if(ncol(predict.df) == 1)

                    colnames(predict.df) = paste(all.vars(formula(i))[1],
                                                 ”fit”, sep = ”.”) else
                    colnames(predict.df) = paste(all.vars(formula(i))[1],
                    colnames(predict.df), sep = ”.”)

                    # Return predictions
          return(predict.df)

    } )

  )

  # Bind in newdata
  cbind(newdata, predict.df)

}

# df without scaling

modlist11.2 = list(
  glm(herb_layer   forest_structure + light, na.action=na.omit,
      data=df_sem_simple),
  nnet::multinom(forest_structure   forest_management + soil_type,
                 data=df_sem_simple),
  glm(tree_layer   forest_structure + soil_type, na.action=na.omit,
      data=df_sem_simple),
  glm(light   tree_layer + forest_structure, na.action=na.omit,
      data=df_sem_simple),
  glm(biodiversity   herb_layer + forest_management + forest_structure,
      na.action=na.omit, data=df_sem_simple)
)

df_sem11.2_a<-df_sem_simple
df_sem11.2_a$tree_layer<-df_sem11.2_a$tree_layer*1.5

test1<-sem.predict1(modlist11.2, df_sem11.2_a)

mean(test1$biodiversity, na.rm=T)
mean(test1$biodiversity.fit, na.rm=T)

df_sem11.2_a<-df_sem_simple
df_sem11.2_a$forest_management<-df_sem11.2_a$forest_management*1.5

test1<-sem.predict1(modlist11.2, df_sem11.2_a)

mean(test1$biodiversity, na.rm=T)
mean(test1$biodiversity.fit, na.rm=T)

df_sem11.2_a<-df_sem_simple
df_sem11.2_a$light<-df_sem11.2_a$light*1.5

test1<-sem.predict1(modlist11.2, df_sem11.2_a)

mean(test1$biodiversity, na.rm=T)
mean(test1$biodiversity.fit, na.rm=T)
```

```

df_sem11.2_a<-df_sem_simple
df_sem11.2_a$soil_type<-df_sem11.2_a$soil_type*1.5

test1<-sem.predict1(modlist11.2, df_sem11.2_a)

mean(test1$biodiversity, na.rm=T)
mean(test1$biodiversity.fit, na.rm=T)

# direct approach

summary(sem_simple, fit.measures=T, rsquare=T)
light_on_biodiv<-0.418*0.547
# 0.228646
# indirects:
# li->hl->bd: 0.228646
fm_on_biodiv<-0.857 + 1.433*-0.464 +1.433*0.267*0.547 +
  1.433*0.368*-0.369*0.418*0.547 + 1.433*0.302*0.418*0.547
# 0.4558342
# direct: 0.857
# indirects:
# fm->fs->bd: -0.664912
# fm->fs->hl->bd: 0.2092882
# fm->fs->tl->li->hl->bd: -0.04449221
# fm->fs->li->hl->bd: 0.09895021
soil_on_biodiv<-1.626*-0.464 + 1.626*0.267*0.547 +
  1.626*0.368*-0.369*0.418*0.547 + 1.626*0.302*0.418*0.547 +
  -0.509*-0.369*0.418*0.547
# -0.4122513
# indirects:
# st->fs->bd: -0.754464
# st->fs->hl->bd: 0.2374757
# st->fs->tl->li->hl->bd: -0.05048453
# st->fs->li->hl->bd: 0.1122771
# st->tr->li->hl->bd: 0.04294452

#fm on biodiv plot
scaleling<-scale(df_sem_simple[names(df_sem_simple)[!names(
  df_sem_simple) == ”forest_structure”]])
y_<-attr(scaleling, ”scaled:center”)[3]
S_y<-attr(scaleling, ”scaled:scale”)[3]
x_<-attr(scaleling, ”scaled:center”)[1]
S_x<-attr(scaleling, ”scaled:scale”)[1]
m<-0.4558342
m_d<-0.857 #direct
mi_1<--0.664912
mi_2<-0.2092882
mi_3<--0.04449221
mi_4<-0.09895021
plot(df_sem_simple$forest_management, df_sem_simple$biodiversity,
     xlab=”Forest management intensity index”,
     ylab=”effective number of species”, pch=20, xlim=c(0, 3))
curve(m*S_y/S_x*x+(-S_y*x_*m/S_x+y_), add=T, lwd=3, col=”#57984f”)
curve(m_d*S_y/S_x*x+(-S_y*x_*m_d/S_x+y_), add=T, lwd=1, lty=5, col=”#4253b4”)
curve(mi_1*S_y/S_x*x+(-S_y*x_*mi_1/S_x+y_), add=T, lwd=1, lty=3, col=”#a6a7ad”)
curve(mi_2*S_y/S_x*x+(-S_y*x_*mi_2/S_x+y_), add=T, lwd=1, lty=3, col=”#a6a7ad”)
curve(mi_3*S_y/S_x*x+(-S_y*x_*mi_3/S_x+y_), add=T, lwd=1, lty=3, col=”#a6a7ad”)
curve(mi_4*S_y/S_x*x+(-S_y*x_*mi_4/S_x+y_), add=T, lwd=1, lty=3, col=”#a6a7ad”)
text(locator(), labels=c(”1”,”2”,”3”,”4”,”5”,”6”))
m_s<-m*S_y/S_x #4.934273

#light on biodiv plot
y_<-attr(scaleling, ”scaled:center”)[3]
S_y<-attr(scaleling, ”scaled:scale”)[3]
x_<-attr(scaleling, ”scaled:center”)[6]
S_x<-attr(scaleling, ”scaled:scale”)[6]
m<-0.228646
plot(df_sem_simple$light, df_sem_simple$biodiversity,
     xlab=”Light extinction coefficient”,
     ylab=”effective number of species”, pch=20)
```

```
curve(m*S_y/S_x*x+(-S_y*x_*m/S_x+y_), add=T, lwd=3, col=”#57984f”)
m_s<-m*S_y/S_x # 0.5049382

#soil on biodiv plot
scaleling<-scale(df_sem_simple[names(df_sem_simple)[!names(df_sem_simple) ==
                                                        ”forest_structure”]])
y_<-attr(scaleling, ”scaled:center”)[3]
S_y<-attr(scaleling, ”scaled:scale”)[3]
x_<-attr(scaleling, ”scaled:center”)[4]
S_x<-attr(scaleling, ”scaled:scale”)[4]
m<--0.4122513
mi_1<--0.754464
mi_2<-0.2374757
mi_3<--0.05048453
mi_4<-0.1122771
mi_5<-0.04294452
plot(df_sem_simple$soil_type, df_sem_simple$biodiversity,
     xlab=”content of Sand [g/kg]”,
     ylab=”effective number of species”, pch=20)
curve(m*S_y/S_x*x+(-S_y*x_*m/S_x+y_), add=T, lwd=3, col=”#57984f”)
curve(mi_1*S_y/S_x*x+(-S_y*x_*mi_1/S_x+y_), add=T, lwd=1, lty=3, col=”#a6a7ad”)
curve(mi_2*S_y/S_x*x+(-S_y*x_*mi_2/S_x+y_), add=T, lwd=1, lty=3, col=”#a6a7ad”)
curve(mi_3*S_y/S_x*x+(-S_y*x_*mi_3/S_x+y_), add=T, lwd=1, lty=3, col=”#a6a7ad”)
curve(mi_4*S_y/S_x*x+(-S_y*x_*mi_4/S_x+y_), add=T, lwd=1, lty=3, col=”#a6a7ad”)
curve(mi_5*S_y/S_x*x+(-S_y*x_*mi_5/S_x+y_), add=T, lwd=1, lty=3, col=”#a6a7ad”)
text(locator(), labels=c(”1”,”2”,”3”,”4”,”5”,”6”))
m_s<-m*S_y/S_x #-0.006538488
```

code/6_3_final_sem.R

B.4 JAGS

```
# model
model{
  for(i in 1:n){

    Light[i]   dgamma(shape_light, shape_light/lambda_light[i])
    lambda_light[i]<-exp(mu_light[i])
    mu_light[i]<- alpha_light[1]*FS_bu[i] + alpha_light[2]*FS_fi[i] +
                    alpha_light[3]*FS_ki[i] + beta_light*TL[i]


    TL[i]   dbeta(a[i],b[i])
    a[i]<-p[i]*tau[4]
    b[i]<-(1-p[i])*tau[4]
    p[i]<-1/(1+exp(-mu_tl[i]))
    mu_tl[i]<-alpha_tl[1]*FM[i] + beta_tl*Soil[i]


    HL[i]   dnorm(mu_hl[i], tau[2])
    mu_hl[i]<-alpha_hl[1]*FS_bu[i] + alpha_hl[2]*FS_fi[i] +
      alpha_hl[3]*FS_ki[i] + beta_hl*Light[i]

    BD[i]   dgamma(shape_bd, shape_bd/lambda_bd[i])
    lambda_bd[i]<-exp(mu_bd[i])
    mu_bd[i]<-alpha_bd[1]*FS_bu[i] + alpha_bd[2]*FS_fi[i] +
      alpha_bd[3]*FS_ki[i] + beta_bd[1]*HL[i] + beta_bd[2]*FM[i]

    FS[i,1:3]   dsample(p_fs[i, 1:3],1)
    for (j in 1:3){
      p_fs[i,j]<-exp(mu_fs[i,j])/sum(exp(mu_fs[i,1]), exp(mu_fs[i,2]),
                                        exp(mu_fs[i,3]))
      mu_fs[i,j]<-beta_fs1[j]*Soil[i] + beta_fs2[j]*FM[i]
    }


  }
```

```
  for(k in 1:3){
    alpha_light[k]   dnorm(0,0.5)
    alpha_tl[k]   dnorm(0,0.0001)
    alpha_hl[k]   dnorm(0,0.0001)
    alpha_bd[k]   dnorm(0,0.0001)
    beta_fs1[k]   dnorm(0,0.5)
    beta_fs2[k]   dnorm(0,0.5)
  }
  for(k in 1:2){
    beta_bd[k]   dnorm(0,0.0001)
  }
  for(k in 1:4){
    sigma[k]   dunif(0,100)
    tau[k]<-pow(sigma[k],-2)
  }
  beta_light   ddexp(0,0.5)
  beta_tl   dnorm(0,0.0001)
  beta_hl   dnorm(0,0.0001)
  shape_light   dunif(0,100)
  shape_bd   dunif(0,100)
}

# null model

model{
  for(i in 1:n){

    Light[i]   dgamma(shape_light, shape_light/lambda_light[i])
    lambda_light[i]<-exp(mu_light[i])
    mu_light[i]<- alpha_light[1]


    TL[i]   dbeta(a[i],b[i])
    a[i]<-p[i]*tau[4]
    b[i]<-(1-p[i])*tau[4]
    p[i]<-1/(1+exp(-mu_tl[i]))
    mu_tl[i]<-alpha_tl[1]

    HL[i]   dnorm(mu_hl[i], tau[2])
    mu_hl[i]<-alpha_hl[1]

    BD[i]   dgamma(shape_bd, shape_bd/lambda_bd[i])
    lambda_bd[i]<-exp(mu_bd[i])
    mu_bd[i]<-alpha_bd[1]

    for (j in 1:3){
      FS[i,j]   dbern(p_fs[i,j])
      p_fs[i,j]<-1/(1+exp(-mu_fs[i,j]))
      mu_fs[i,j]<-alpha_fs[j]
    }
  }
  for(k in 1:3){
    alpha_light[k]   dnorm(0,0.5)
    alpha_tl[k]   dnorm(0,0.0001)
    alpha_hl[k]   dnorm(0,0.0001)
    alpha_bd[k]   dnorm(0,0.0001)
    alpha_fs[k]   dnorm(0,0.0001)
  }
  for(k in 1:4){
    sigma[k]   dunif(0,100)
    tau[k]<-pow(sigma[k],-2)
  }
  r_tl   dunif(0,100)
  shape_light   dunif(0,100)
  shape_bd   dunif(0,100)
}

range01<-function(x){(x-min(x, na.rm=T)+0.001)/(max(x, na.rm=T)-min(x, na.rm=T)+0.002)}
```

```
# Data

Sand<-(Fine_Sand+Coarse_Sand+Medium_Sand)

cat_mts<-model.matrix(ForMI+Sand mainTreeSpecies)

datalist<-list(n=(nrow(Forest_EPs_pure)-1),
                 Soil=Sand[-39],
                 FM=ForMI[-39],
                 FS_bu=cat_mts[,1],
                 FS_fi=cat_mts[,2],
                 FS_ki=cat_mts[,3],
                 Light=mean_LEC_corr[-39],
                 TL=B2_fs[-39]/100+0.0001,
                 HL=H_ell_weighted[-39],
                 BD=exp(shannon_herbs)[-39])

datalist_scale<-lapply(datalist, range01)
datalist_scale$n<-datalist$n

datalist_scale$FS_fi<-datalist$FS_fi
datalist_scale$FS_ki<-datalist$FS_ki
datalist_scale$FS_bu<-rep(1,54)-datalist$FS_ki - datalist$FS_fi
datalist_scale$TL<-datalist$TL
datalist_scale$FS<-cbind(datalist_scale$FS_bu, datalist_scale$FS_fi,
                           datalist_scale$FS_ki)


# Jags

out<-jags(data = datalist_scale,
            parameters.to.save = c(”alpha_light”, ”alpha_tl”,
                                   ”alpha_hl”, ”alpha_bd”,
                                   ”beta_fs”, ”beta_bd”,
                                   ”sigma”, ”beta_light”,
                                   ”beta_tl”, ”beta_hl”,
                                   ”alpha_fs”, ”beta_fs1”, ”beta_fs2”),
            model.file = paste0(wd, ”/model.txt”),
            n.chains = 3,
            n.adapt = 5000,
            n.iter = 20000,
            n.burnin = 5000,
            parallel = TRUE)

out_null<-jags(data = datalist_scale,
                 parameters.to.save = c(”alpha_light”, ”alpha_tl”,
                                        ”alpha_hl”, ”alpha_bd”,
                                        ”beta_fs”, ”beta_bd”,
                                        ”sigma”, ”beta_light”,
                                        ”beta_tl”, ”beta_hl”,
                                        ”alpha_fs”, ”beta_fs1”, ”beta_fs2”),
                 model.file = paste0(getwd(), ”/nullmodel.txt”),
                 n.chains = 3,
                 n.adapt = 1000,
                 n.iter = 6000,
                 n.burnin = 1000,
                 parallel = TRUE)


# R 2  light

X<-datalist_scale$Light

pred<-exp(out$q50$alpha_light[1]*datalist_scale$FS_bu +
              out$q50$alpha_light[2]*datalist_scale$FS_fi +
              out$q50$alpha_light[3]*datalist_scale$FS_ki +
              out$q50$beta_light*datalist_scale$TL)

null<-exp(out_null$q50$alpha_light[1])

```

```
(Rsq<-1 - sum((X - pred) 2 )/sum((X - null) 2 ) ) #0.4443166


# R 2 Biodiversity

X<-datalist_scale$BD

pred<-exp(out$q50$alpha_bd[1]*datalist_scale$FS_bu +
              out$q50$alpha_bd[2]*datalist_scale$FS_fi +
              out$q50$alpha_bd[3]*datalist_scale$FS_ki +
              out$q50$beta_bd[1]*datalist_scale$HL +
              out$q50$beta_bd[2]*datalist_scale$FM)

null<-exp(out_null$q50$alpha_bd[1])

(Rsq<-1 - sum((X - pred) 2 , na.rm=T)/sum((X - null) 2 , na.rm=T)) #0.5666024


# R 2 Herb Layer

X<-datalist_scale$HL

pred<-out$q50$alpha_hl[1]*datalist_scale$FS_bu +
  out$q50$alpha_hl[2]*datalist_scale$FS_fi +
  out$q50$alpha_hl[3]*datalist_scale$FS_ki +
  out$q50$beta_hl*datalist_scale$Light

null<-out_null$q50$alpha_hl[1]

(Rsq<-1 - sum((X - pred) 2 , na.rm=T)/sum((X - null) 2 , na.rm=T)) #0.7010546

# R 2 Forest Structure

X<-datalist_scale$FS

predTemp<-matrix(0, ncol=3, nrow=length(datalist_scale$Soil))

pred<-matrix(0, ncol=3, nrow=length(datalist_scale$Soil))


for(j in 1:3){

  predTemp[,j]<-exp(out$q50$beta_fs1[j]*datalist_scale$Soil +
                        out$q50$beta_fs2[j]*datalist_scale$FM)

}

for(j in 1:3){

  pred[,j]<-predTemp[,j]/apply(predTemp,1,sum)

}

null<-1/(1+exp(-(out_null$q50$alpha_fs[1])))

(Rsq<-1 - sum((X - pred) 2 , na.rm=T)/sum((X - null) 2 , na.rm=T)) #0.422625


# R 2 Tree Layer

X<-datalist_scale$TL

pred<-1/(1+exp(-(out$q50$alpha_tl[1]*datalist_scale$FS_bu +
                     out$q50$alpha_tl[2]*datalist_scale$FS_fi +
                     out$q50$alpha_tl[3]*datalist_scale$FS_ki +
                     out$q50$beta_tl*datalist_scale$Soil)))

null<-1/(1+exp(-(out$q50$alpha_tl[1])))

(Rsq<-1 - sum((X - pred) 2 , na.rm=T)/sum((X - null) 2 , na.rm=T)) #0.8117061
```

```



# Distribution

hist(out$sims.list$alpha_light)
hist(out$sims.list$alpha_tl)
hist(out$sims.list$alpha_hl)
hist(out$sims.list$alpha_bd)
hist(out$sims.list$alpha_fs)
hist(out$sims.list$beta_fs)
hist(out$sims.list$beta_bd)
hist(out$sims.list$beta_tl)
hist(out$sims.list$beta_hl)
hist(out$sims.list$sigma)
hist(out$sims.list$beta_light)

descdist(mean_LEC_corr, boot=500)
fit.gamma<-fitdist(mean_LEC_corr, distr=”gamma”)
plot(fit.gamma)

descdist(exp(shannon_herbs), boot=500)
fit.gamma<-fitdist(exp(shannon_herbs), distr=”gamma”)
plot(fit.gamma)

descdist(B2_fs, boot=500)
B2_fs_scale<-range01(B2_fs)
fit.beta<-fitdist(B2_fs_scale, distr=”beta”)
plot(fit.beta)

descdist(as.numeric(na.omit(H_ell_weighted)), boot=500)
fit.norm<-fitdist(as.numeric(na.omit(H_ell_weighted)), distr=”norm”)
plot(fit.norm)
```

code/6_3_1_final_jags.R

B.5 Text

```
############
## 4.TEXT ##
############

#### 4.1 inits & misc ####

# colors:
# 57984f (darkgreen)
# 4253b4 (blue)
# a6a7ad (grey)
Sand<-(Fine_Sand+Coarse_Sand+Medium_Sand)

#### 4.2 Introduction ####

# first glimpse at regression & correlation
f1<-ggplot(Forest_EPs, aes(x=mean_LEC_corr, y=H_ell_weighted))+
  geom_point(color=”#57984f”)+
  geom_smooth(method=”glm”, color=”#4253b4”)+
  scale_x_continuous(”Percentage of light shining through canopy”)+
  scale_y_continuous(”Light demand of herb species”)+
  annotate(”text”, -Inf, Inf, label = round(
    cor.test(mean_LEC_corr,H_ell_weighted)$estimate,2), size=10, hjust = 0,
    vjust = 1)

f2<-ggplot(Forest_EPs, aes(x=H_ell_weighted, y=H_fs))+
  geom_point(color=”#57984f”)+
  geom_smooth(method=”glm”, color=”#4253b4”)+
  scale_x_continuous(”Light demand of herb species”)+
  scale_y_continuous(”Forest specialist basal area share in herb layer”)+
```

```
  annotate(”text”,-Inf, Inf, label = round(
    cor.test(H_ell_weighted,H_fs)$estimate,2), size=10, hjust = 0,
    vjust = 1)

f3<-ggplot(Forest_EPs, aes(x=H_fs, y=exp(shannon_herbs)))+
  geom_point(color=”#57984f”)+
  geom_smooth(method=”glm”, color=”#4253b4”)+
  scale_x_continuous(”Forest specialist basal area share in herb layer”)+
  scale_y_continuous(”Effective number of species in herb layer”)+
  annotate(”text”, -Inf, Inf, label = round(
    cor.test(exp(shannon_herbs),H_fs)$estimate,2), size=10, hjust = 0,
    vjust = 1)

multiplot(f1,f2,f3, cols=3)

#### 4.3 Methods ####

# get unique species count
num_spec_S<-length(unique(veg_data$Species[veg_data$Layer==”S” &
                                               veg_data$cover¿0]))
num_spec_H<-length(unique(veg_data$Species[veg_data$Layer==”H” &
                                               veg_data$cover¿0]))

# percentage of missing data in admixed tree species
sum(is.na(ATS1.x))/74
sum(is.na(ATS2.x))/74
sum(is.na(ATS1.y))/74
sum(is.na(ATS2.y))/74

# table of main tree species
table(mainTreeSpecies)

# check for multicolliniarity
# soil on light
mctest(x=df_sem_complex[,10:12],y=df_sem_complex[,1])
mctest(x=df_sem_complex[,10:12],y=df_sem_complex[,2])
# soil on trees
mctest(x=df_sem_complex[,10:12],y=df_sem_complex[,5])
# soil on biodiv
mctest(x=df_sem_complex[,10:12],y=df_sem_complex[,6])
mctest(x=df_sem_complex[,10:12],y=df_sem_complex[,7])
# soil on herbs
mctest(x=df_sem_complex[,10:12],y=df_sem_complex[,8])
# light on herbs
mctest(x=df_sem_complex[,1:2],y=df_sem_complex[,8])
# light on biodiv
mctest(x=df_sem_complex[,1:2],y=df_sem_complex[,6])
mctest(x=df_sem_complex[,1:2],y=df_sem_complex[,7])
# fm on biodiv
mctest(x=df_sem_complex[,3:4],y=df_sem_complex[,6])
mctest(x=df_sem_complex[,3:4],y=df_sem_complex[,7])
# fm on herbs
mctest(x=df_sem_complex[,3:4],y=df_sem_complex[,8])
# fm on trees
mctest(x=df_sem_complex[,3:4],y=df_sem_complex[,5])
# fm on light
mctest(x=df_sem_complex[,3:4],y=df_sem_complex[,1])
mctest(x=df_sem_complex[,3:4],y=df_sem_complex[,2])


# cor mat sem_complex
df_sem10_4_cor<-data.frame(”LEC”=mean_LEC_corr/10,
                             ”ELV”=H_ell_weighted/10,
                             ForMI,SMI,
                             ”Tree Layer”=B2_fs/100,
                             ”Effective n. of sp.”=exp(shannon_herbs)/10,
                             ”Simpsons I”=Simpsons_herbs,
                             ”Forest specialists”=H_fs/100,
                             ”Tree density” = MTS_N/1000,
                             ”Sand”=(Fine_Sand+Coarse_Sand+Medium_Sand)/1000,
```

```
                          ”Silt”=(Fine_Silt+Medium_Silt+Coarse_Silt)/1000,
                          ”Clay”=Clay/1000)
get_cor_heatmap(rcorr(as.matrix(df_sem10_4_cor)))

# cross vailidation
df_cv<-df_sem10_4[1:52,]

sem_cv<-sem(model_cv, data=df_cv, ordered = c(”mts”))
summary(sem_cv,fit.measures=T,rsquare=T)


df_cv<-data.frame(”mean_LEC_corr”=mean_LEC_corr/10,
                  ”H_ell_weighted”=H_ell_weighted/10,
                  ForMI,SMI,
                  ”B2”=B2_fs/100,
                  ”shannon_herbs”=as.factor(exp(shannon_herbs)/10),
                  Simpsons_herbs,
                  ”H_fs”=H_fs/100,
                  ”MTS_N” = MTS_N/1000,
                  ”Sand”=(Fine_Sand+Coarse_Sand+Medium_Sand)/1000,
                  ”Silt”=(Fine_Silt+Medium_Silt+Coarse_Silt)/1000,
                  ”Clay”=Clay/1000)

df_cv<-df_cv[complete.cases(df_cv),]


train_control<-trainControl(method=”cv”, number=10)
grid<-expand.grid(.fL=c(0), .usekernel=c(FALSE))
model<-train(shannon_herbs ., data=df_cv, trControl=train_control,
             method=”glm”, tuneGrid=grid)

train_control<-trainControl(method=”repeatedcv”, number=10, repeats=3)
model<-train(shannon_herbs ., data=df_cv, trControl=train_control,
             method=”nb”)

train_control<-trainControl(method=”cv”, number=10, savePredictions = TRUE)
model<-train(shannon_herbs ., data=df_cv, trControl=train_control,
            method=”rpart”)

# full correlation plot
df_fullcor<-data.frame(”LEC”=mean_LEC_corr, ”ELV”=H_ell_weighted,
                       ForMI,SMI, ”Area Trees”=B2_fs,
                       ”EnoS”=exp(shannon_herbs),
                       ”Simpsons Index”=Simpsons_herbs,
                       ”MTS”=ordered(mainTreeSpecies),”MTS N” = MTS_N,
                       ”Sand”=(Fine_Sand+Coarse_Sand+Medium_Sand),
                       ”Silt”=(Fine_Silt+Medium_Silt+Coarse_Silt), Clay,
                       ”Area Forest spec”=H_fs)

get_cor_heatmap(rcorr(as.matrix(dplyr::select_if(df_fullcor, is.numeric))))

df_fullcor$MTS<-plyr::revalue(df_fullcor$MTS, c(”Bu”=”Fagus sylvatica”,
                                                ”Fi”=”Picea abies”,
                                                ”Ki”=”Pinus sylvestris”))

# complete hist

df_hist<-df_sem10_4
df_hist<-cbind(df_hist,MTS_N)

names(df_hist)<-c(”LEC”, ”ELV”, ”ForMI”, ”SMI”, ”Tree layer share”,
                  ”ENS in herb layer”, ”Simpson’s index in herb layer”,
                  ”Forest specialist in herb layer”, ”Main tree species”,
                  ”Sand”, ”Silt”, ”Clay”, ”Number of trees”)

df_hist<-within(df_hist, rm(”Main tree species”))

multi.hist(df_hist, main = ””, density=F, col=”#57984f”)

#### 4.4 Results ####
```

```
#### 4.5 Discussion ####

# Ellenberg light values found on FEP
df_floraweb_light<-read.xlsx(paste0(getwd(),
                "/species_forest_specialist_FLORAWEB_MANUALLY_ADDED_MISSING.xlsx"),1)
hist(as.numeric(df_floraweb_light$Lichtzahl[Typ=="Waldspezialist"]))
hist(as.numeric(df_floraweb_light$Lichtzahl[Typ=="kein Waldspezialist"]))


df_fw<-data.frame("elv"=c(as.numeric(df_floraweb_light$Lichtzahl[
  Typ=="Waldspezialist"]),
  as.numeric(df_floraweb_light$Lichtzahl[
    Typ=="kein Waldspezialist"])),
  "type"=c(rep("forest type",
               length(df_floraweb_light$Lichtzahl[
                 Typ=="Waldspezialist"])),
           rep("non-forest type",
               length(df_floraweb_light$Lichtzahl[
                 Typ=="kein Waldspezialist"]))))

mytable<-cbind(type=c("forest type", "non-forest type"),
                 mean=c(round(mean(as.numeric(df_floraweb_light$Lichtzahl[
                   Typ=="Waldspezialist"]),na.rm=T),2)
                   ,round(mean(as.numeric(df_floraweb_light$Lichtzahl[
                     Typ=="kein Waldspezialist"]),na.rm=T),2)),
                 indifferent=c(sum(is.na(df_floraweb_light$Lichtzahl[
                   Typ=="Waldspezialist"]))
                   ,sum(is.na(df_floraweb_light$Lichtzahl[
                     Typ=="kein Waldspezialist"]))))

ggplot(df_fw,aes(x=elv, fill=type))+
  geom_histogram(position = "dodge", binwidth = 0.5)+
  scale_x_continuous("Ellenberg light values", breaks=seq(1,9))+
  scale_y_continuous("Number of plant species on FEP")+
  annotation_custom(tableGrob(mytable), xmin=1, xmax=5, ymin=50, ymax=60)+
  scale_color_manual(values=c("#57984f", "#4253b4"))+
  scale_fill_manual(values=c("#57984f", "#4253b4"))


# formi

summary(Forest_EPs_pure$Inonat)
summary(Forest_EPs_pure$Iharv)
summary(Forest_EPs_pure$Idwcut)
var(Forest_EPs_pure$Inonat, na.rm=T)
var(Forest_EPs_pure$Iharv, na.rm=T)
var(Forest_EPs_pure$Idwcut, na.rm=T)

# soil

cor.test(Sand, (S_fs+S_nfs))
cor.test(Sand, (H_fs+H_nfs))
cor.test(Sand, B1_fs)
cor.test(Sand, B2_fs)
cor.test(Sand, exp(shannon_herbs))
cor.test(Sand, exp(shannon_shrubs))
cor.test(Forest_EPs_pure$Sand, pH)
cor.test(Forest_EPs_pure$Sand, Cmic_soil)

df_soil<-data.frame(pH, "Sand content"=Sand, "microbial biomass"=Cmic_soil,
                      "Tree smaller 10m share"=B1_fs,
                      "Trees taller 10m share"=B2_fs,
                      "Scrub share" = S_fs+S_nfs,
                      "Herb share"=H_fs+H_nfs, "enos herbs"=exp(shannon_herbs),
                      "enos scrubs"=exp(shannon_shrubs))
get_cor_heatmap(rcorr(as.matrix(df_soil)))

# soil reg

Forest_EPs_pure$Sand<-(Fine_Sand+Coarse_Sand+Medium_Sand)
```

```
Forest_EPs_pure$H_total <- H_fs+H_nfs
Forest_EPs_pure$S_total <- S_fs+S_nfs

df_soil_reg <- data.frame(”Sand”=(Fine_Sand+Coarse_Sand+Medium_Sand),
                          ”Trees taller 10m”=B2_fs,
                          ”Herb layer”=H_fs+H_nfs,
                          ”Trees under 10m”=B1_fs,
                          ”Shrub layer”=S_fs+S_nfs)

forest_long <- tidyr::gather(df_soil_reg, key = variable, value = value, -Sand)

ggplot(forest_long, aes(x = Sand, y = value, color = variable)) +
  geom_point() +
  geom_smooth(method = glm)+
  scale_x_continuous(”Sand content in g/kg”)+
  scale_y_continuous(”Share of respective layer in %”)+
  scale_color_manual(name = NULL, values = c(Trees.taller.10m = ”#57984f”,
                                             Herb.layer = ”black”,
                                             Trees.under.10m = ”#4253b4”,
                                             Shrub.layer = ”#a6a7ad”))+
  theme(legend.position = c(0.8, 0.8))


# light + ENS

df_light <- data.frame(”shannon_h”=exp(shannon_herbs),
                       ”shannon_s”=exp(shannon_shrubs),
                       H_ell_weighted, S_ell_weighted, mean_LEC_corr, H_fs,
                       B1_fs, ”S_total”=S_fs+H_fs, ”H_total”=H_fs+H_nfs, H_nfs,
                       S_fs, S_nfs, number_shrubs,
                       number_herbs, number_B1, number_woody_species_S_B1_B2)
get_cor_heatmap(rcorr(as.matrix(df_light)))

ggplot(data=df_light, aes(x=mean_LEC_corr))+
  geom_smooth(aes(y=exp(shannon_herbs)), method=”glm”, color=”blue”)+
  geom_smooth(aes(y=exp(shannon_shrubs)), method=”glm”, color=”red”)

# discussion forest specialist

df_fs_reg <- data.frame(”LEC”=mean_LEC_corr, ”forest type”=(H_fs/(H_fs+H_nfs)),
                        ”non-forest type”=(H_nfs/(H_fs+H_nfs)))

df_fs_long <- tidyr::gather(df_fs_reg, key = variable, value = value, -LEC)

ggplot(df_fs_long, aes(x=LEC,y = value, color = variable))+
  geom_point()+
  geom_smooth(method=”glm”)+
  scale_x_continuous(”LEC”)+
  scale_y_continuous(”Share of species in herb layer of respecitve type [%]”)+
  scale_color_manual(name = NULL, values = c(forest.type = ”#57984f”,
                                             non.forest.type = ”#4253b4”))+
  theme(legend.position = c(0.8, 0.5))
```

code/6_4_final_text.R

Appendix C

Table of data sets

Dataset ID	Title	Version
10580	EP all exploratories	2.3.4
14410	Vegetation Records for Forest EPs in 2010	5.1.4
14686	MinSoil 2011 Mineral Soil Texture	1.9.6
16466	Forest Management intensity index	1.3.4
17086	MinSoil 2011 Mineral Soil Bulk Density CN stocks	1.1.1
17486	Forest EP stand age	
17687	Forest EP stand structure and composition	1.4.5
17706	Forest EP - new forest type classification	
17746	Forest EP - SMI - Silvicultural management intensity index	
20366	Vegetation Records for Forest EPs, 2009 - 2016	1.4.5
22506	Light measurements in MIP forest plots	1.1.1
10574	forest soil fauna densities grouplevel 2008	5.1.4

Table C.1: Data sets used for my analysis from BExIS (2018)

Selbstständigkeitserklärung

Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen habe ich unter Angabe der Quellen kenntlich gemacht.

(I hereby declare that I have composed this document unassistedly and that I only used the sources and devices I declared. Passages taken verbatim or in meaning from other sources are identified as such and the sources are acknowledged and cited.)

Freiburg, 10th of July 2018