Universidad de Valladolid

Master Erasmus Mundus in
Mediterranean Forestry and Natural Resources
(MEDFOR)

Tree biomass and biodiversity relationship in a mixed Mediterranean forest in Spain

Student: Narangarav Dugarsuren

Co-advisors: Prof. Felipe Oviedo Bravo

Ing. Cristobal Ordonez Alonso

July, 2019

Co-funded by the Erasmus+ Programme of the European Union

INDEX

LIST OF FIGURES 4
LIST OF TABLES 4
RESUMEN 5
ABSTRACT 5
1. - INTRODUCTION 6
2. - OBJECTIVES 8
3. - MATERIAL AND METHODS 8
3.1. Study site 8
3.2. Biomass estimation 10
3.3. Tree species diversity estimation 11
3.3.1. Species richness indices 11
3.3.1.1. Simpson Diversity index (1-D) 11
3.3.1.2. Shannon index 11
3.3.1.3. Berker-Parker index (D) 12
3.3.1.4. Evenness index (E) 12
3.3.2. Species intermingling 12
3.3.2.1. Species segregation (S) index 12
3.3.2.2. Mingling (Mi) index 13
3.3.2.3. Spatial diversity status (MS) 14
3.3.3. Spatial structural indices 15
3.3.3.1. Uniform Angle Index (W) 16
3.3.3.2. Aggregation index R 16
3.3.3.3. Height Differentiation index (TH) 16
3.3.3.4. Vertical species profile (A) 17
3.4. Statistical analysis 17
4. - RESULTS 19
5. - DISCUSSION 23
6. - CONCLUSIONS 24
7. - AKNOWLEDGEMENTS 24
8. - REFERENCES 25
ANNEX 29

LIST OF FIGURES

Figure 1.Study area location of Marteloscope of Llano de San Marugan (Valladolid, Spain) 9
Figure 2. Workflow for the study of tree biomass and diversity relationship..... 9
Figure 3. Description and calculation of Mingling index (Mi), Spatial diversity status (MS) and Uniform Angle index (Wi) and corresponding likelihood values for structure unit of 4 neighbors around the reference *i* tree: a, b, c, d are the tree species types ; $\boldsymbol{\theta}$ are angles between adjacent neighbor trees; $\boldsymbol{\theta s}$ is a standard angle (which is equal to 360/4 for 4 neighbor trees) (Adapted from Gadow & Hui, 2002) 15

LIST OF TABLES

Table 1. Biomass allometric equations by species 10
Table 2. Categorization of indices 11
Table 3. Descriptions of variables for S index calculation..... 13
Table 4. Fitting models and their predictor variable reference 19
Table 5. Parameter estimates for biomass from selected models at community level ... 21
Table 6. Parameter estimates for biomass from selected models at species level 22

RESUMEN

La biomasa de árboles y su relación con la diversidad en bosques mixtos se ha convertido en uno de los temas de investigación más interesantes para los ecólogos en las últimas décadas debido a la importancia de los bosques mixtos para una mejor provisión de servicios ecosistémicos. La pregunta "¿Producen los bosques mixtos más a medida que aumenta la diversidad de árboles?" Ha sido objeto de muchos estudios que llevan a resultados no concluyentes. Este estudio se realizó para contrastar el resultado de estudios anteriores mediante la investigación de la biomasa de árboles y la relación de diversidad en un rodal mixto de bosque mediterráneo (Llano de San Marugan, España), tanto a nivel de masa como de árbol individual. Se analizaron diversos modelos que se ajustaron a partir de ecuaciones de regresión lineal y no lineal para determinar la relación entre la biomasa de los árboles y la diversidad. Se utilizaron 10 índices de diversidad que se pueden clasificar en 3 categorías: índices de riqueza de especies (Sm, Sn, D, E); Índices de composición / mezcla de especies (Mi, MS, S); Los índices estructurales verticales (W, A, TH) como variables predictoras de los modelos con el objeto de caracterizar diferentes estructuras de diversidad en el rodal. Nuestro resultado reveló que la relación entre la biomasa y la diversidad de árboles varía entre las especies. Una combinación de la relación negativa del índice D-Berker-Parker (abundancia de especies dominantes) y la relación positiva de TH (heterogeneidad de la altura) explica la variación de la biomasa a nivel rodal y para *Pinus pinea*. La biomasa de las especies de *Quercus* (*Quercus faginea* y *Quercus ilex*) se relaciona positivamente con la proporción de especies en área basimétrica (Gp); los índices de diversidad probados no mostraron ninguna relación con la biomasa de las especies del género *Quercus.*

Palabras clave: diversidad de especies arbóreas, índices de diversidad, riqueza específica, composición específica, estructura vertical del rodal.

ABSTRACT

Tree biomass and diversity relationship in mixed forest has become one of the attractive research subjects for ecologist in recent decades due to an importance of multicultural mixed forest for better provision of goods and services than monoculture. The questions "Does mixed forest produce more productive and the productivity increase as tree diversity increases?" have been subject of many researches that lead to two contrast results. This study was conducted to contrast the result of previous studies by investigating the tree biomass and diversity relation in Mediterreanean multicultural mixed stand, Llano de San Marugan, Spain, at stand and individual species level. A variety of models that developed from linear and nonlinear regression equations were employed to reveal tree biomass and diversity relation. 10 diversity indices that falls in 3 categories: species richness indices (Sm, Sn, D, E); species compositional/mingling indices (Mi, MS, S); vertical structural indices (W, A, TH) were used as predictor variables for the models to characterize different structure of diversity in the stand. Our result revealed that tree biomass and diversity relation varies among species. A combination of negative relation of D- Berker-Parker index (abundance of dominant species) and positive relation of TH (height heterogeneity) explains the variation of biomass at community level and for *Pinus pinea*. Biomass of *Quercus* species *(Quercus faginea* and *Quercus ilex)* was positively related with basal area proportion of species (Gp); the tested diversity indices didn't show any relation with biomass of *Quercus* species and *Juniperus thurifera* as concerned by metrics and models in this study.

Key words: tree biomass, tree diversity, mixed forest, diversity indices, species richness, species composition, stand vertical structure.

1. - INTRODUCTION

Mediterranean forests are characterized by a remarkable set of features that make them naturally and aesthetically attractive, but also quite fragile (Scarascia-Mugnozza *et al*., 2000). Mediterranean forest is a multi-functional, providing a wide range of goods and services for society ranging from products with high market value (fuelwood, cork, mushroom, pinecones etc) and non-market value ecosystem services (soil and landscape protection, water regulation, biodiversity conservation, carbon dioxide fixation, recreation, aesthetic view etc). The latter is more significant than their productive value, especially their significant role for carbon sequestration (del Río et al., 2017). One of notable characteristics of Mediterranean forests is its rich biodiversity, reflected by high genetic variability, exemplified by the large number of tree species in comparison to Nordic forests resulting from the survival of many conifer and broadleaf species during the glacial periods. Long-term exploitation (manipulation) of trees and forestland since ancient times is another feature of Mediterranean forest which results in the dispersion of species as *Pinus pinea*, *Castanea sativa*, and *Quercus suber* all over the Mediterranean basin (Scarascia-Mugnozza et al., 2000). Dry, hot, harsh climate along with long lasting and frequent droughts, pest and decease, increasing the risk of large-scale fires and severe water scarcity are main challenges for the Mediterranean forests which largely impact on forest health, growth and productivity. The role of mixed forest for promoting forest productivity while coping with these challenges has been increased in Mediterranean region in recent decades.

Multicultural mixed forest have been taken a great attention in recent decades due to its greater provision of goods and services, high ecological value in comparison to monoculture forest (Pretzsch & Schütze, 2014; Riofrío, *et al*., 2017). Mixed forest is defined as a forest unit of at least 0.5 ha that composes at least two tree species at any developmental stage, shares common resources (water, light, soil nutrients) and its structure and component species are altered over the time (Bravo-Oviedo et al., 2014). Main characterizations of mixed forests are described not only by better protection, preservation, maintaining and monitoring of biodiversity but also have high resistance capacity against both natural and anthropogenic disturbances such as climate change, storm, pest and decease, air pollution and its consequences. Economic importance of mixed forest is un-negligible because of its multi-use, multi-source than pure stands (Knoke *et al*., 2005).

The loss of biodiversity imposed by anthropogenic and climatic change has brought the importance of diversity under the control worldwide over the past 25 years (Hooper *et al*., 2012; J. Liang *et al*., 2016; Szwagrzyk *et al*., 2007) after the Earth Summit of world governments in Rio de Janeiro in 1992 (Evans, 2016). Since that time, scientists have been taken into account the importance of biodiversity on many ecosystem functioning and service such as productivity, stability, sustainability, sinking carbon dioxide, preserving soil fertility, controlling pest outbreaks, retaining water, and so on (Baskin, 1994). Among them, the importance of tree species diversity on biomass productivity has been studied based on the variety of genes, species, or functional traits of organisms in hundreds of types of ecological communities (Fraser *et al*., 2015; Jingjing Liang *et al*., 2016). A series of biodiversity-ecosystem functioning studies have revealed that biodiversity (including taxonomic, functional and phylogenetic diversity) promotes the functionality of ecosystems such as primary production, decomposition, nutrient cycling, trophic interactions and so on) and consequently supports a broad

range of ecosystem services (e.g. food production, climate regulation, pest control, pollination (Gamfeldt *et al*., 2013; Mori *et al*., 2017). However, contradictory results have been documented in the findings of previous researchers focused on relationship between species diversity and biomass: biomass decrease (Szwagrzyk *et al*., 2007) or doesn't change (Grace *et al*., 2016) with species diversity. In addition, numerous researches justified on loss of biodiversity ranks among the most pronounced changes to environment (Sala *et al*., 2000), reduction of diversity along with species composition changes alter fluxes of energy and essential services that ecosystem provide to human such as production of food, pest and disease control, water purification and so on (Daily, 1997). Biodiversity are largely and irreversibly being degraded and lost globally due to direct drivers; i.e. habitat disturbance, habitat fragmentation, land use change, over-exploitation and the spread of alien species and indirect drivers; i.e. climate change, population growth, economic growth and increasing demand for food, materials, water and energy (Iranah *et al*., 2018). The loss of biodiversity weakens species connections and impairs the ecosystems, leading to extinction of species and local populations, which will disrupt the capacity of ecosystem to contribute to human well-being and sustain future generations.

Tree diversity plays a fundamental role for forest diversity because it is often linked with major properties of forest ecosystem, leading to the possible enhancement of diversity of other forest assembles (Mori *et al*., 2017) and providing required resources and suitable habitat for other forest species (Ozcelik, 2009). Diversity is generally defined by the variety of organisms including micro-organisms, plants, and animals in different ecosystems, i.e. deserts, grassland, forests etc. The most commonly used representation of ecological diversity is species diversity, which is defined by the number of species and abundance of each species living within a certain area (Liu *et al*., 2018). The species coexisting in a certain area are interconnected and dependent on one another for survival, while doing so; they perform important ecosystem functions and offer different ecosystem services for human life and society: provisional service (products obtained from ecosystem: many different type of food, fresh water etc); regulating services (the benefits obtained from the regulation of ecosystem processes: air quality and pollination); cultural service (the non-material benefits that people obtain: spiritual enrichment, recreation and aesthetic experiences); supporting service needed to maintain other services (i.e photosynthesis and nutrient recycling). The provision of ecosystem for such goods and service depends basically on functions performed by living plants (Tilman *et al*., 1997).

Two main mechanisms explain the reasons that biodiversity influence on productivity: selection effects and complementary effects. Different plant species in a mixture have different physiologies, morphologies and life history traits might allow them to fully utilize limiting resources at different space and time than a monoculture of any species (Tilman 1997). For instance, some tree species have more ability to adapt and grow better in cooler and wetter environmental condition while others grow better in hotter and drier environment. If these species grow together in a mixture, these complementary characteristics of both species lead to greater productivity across the whole grown season than either species grows alone. Similarly, tree species that have different root morphologies occupy different soil profile which potentially allowing them to exploit soil resources from different soil depth. However it should be noted that these complementary occur solely when co-existing species exhibits various forms of niche differentiation that allow them to capture resources in different space or time (Cardinale

et al., 2007; David Tilman, 1999). Another mechanism that diversity effects on productivity is selection effect (sampling effect) which describes species specific effect on biomass: a greater productivity in more diverse communities is due to the most productive species which become dominant in the community due to competition. The likelihood of becoming a high productive species increases as diversity increases. Thus this causes in the increment of the total productivity of the community. Such considerations have led to the general perception of having higher productivity in an area where more plant species co-exist.

Forest is 3-dimensional system whose structure is a key element in ecosystem functioning and biological diversity by regulating resource related forest functioning (light, water, soil nutrients supply, capture, use), intra and inter specific interactions (Brockerhoff et al., 2017a; del Río et al., 2018), regeneration pattern, consequent self-thinning and past and present disturbance events (Bohn & Huth, 2017; Zhang *et al.*, 2018). Stand structural diversity leads to increase species richness and contributes to forest stability and integrity (Wang *et al*., 2016). Stand structural diversity combines the concepts of species richness (diversity), species composition (mixture), and spatial diversity (tree positioning) and size differentiation (Bravo & Guerra, 2002). Accordingly three distinct types of stand structural indices and methods have frequently been purposed in preceding literature for explaining the influence of stand structural diversity on productivity and functioning of forest stand: i) species richness - Simpson index (1949), Shannon index (1948), Berger-Parker index (Berger *et al.*, 1970) and Evenness index (Kohn, 1977); ii) species composition indices – Mingling index (Füldner, 1995), Spatial diversity status (Gadow & Hui, 2002) and Segregation index (Pielou, 1977); iii) tree distributional indices including horizontal and vertical patterns and size differentiation - Aggregation Index (Clark *et al*., 1954), Uniform Angle Index (Gadow *et al.*, 1998), Vertical Species Profile (Pretzch 1995b), Height differentiation index (Gadow 1993). Since forest structure is determined in 3 dimensions, it is appropriate to analyze the effect of tree diversity on biomass by the metrics that can fully address 3-dimensionality of mixed forest structures.

2. - OBJECTIVES

In this study, we addressed the question “Does stand diversity impact on biomass?” by examining the relationship between tree biomass and diversity indices (species richness, species compositional, horizontal and vertical structural indices) in the Mediterranean multi-species mixed forest, Llano de San Marugan, Valladolid, Spain. The specific aims were:

- To calculate biomass of individual trees by different compartments (stem, thin branches + needles, medium and thick branches) and in total
- To compute the tree diversity indices to represent the diversity of the stand
- To determine the relationship between tree biomass and diversity.

3. - MATERIAL AND METHODS

3.1. Study site

The study was conducted in a Mediterranean mixed forest stand, located in Llano de San Marugan, Valladolid, Castile and Leon, Spain. Valladolid has a continental

Mediterranean climate with cold winters and hot summers. The average annual temperature ranges between 11 and 12ºC. Fog is frequent in the long, cold winter, while summers are dry and hot with average temperature around 30ºC. Precipitation falls irregularly throughout the year with a minimum in the summer and a maximum in spring and autumn, with maximum of 400 mm. In a mixed forest stand, a marteloscope was installed in 2015 covering 1 hectare (ha). The marteloscope (a square of 100 by 100 m) was divided into 16 subplots (hereafter referred as quadrants), 25 x 25 m length, as shown in Figure 1. Within each quadrant, locations of all trees were recorded in 4.55728º W and 41.43948º N geographic coordinates and their species were identified and corresponding diameter at breast height (dbh, in cm) and, total height (in m). The diameter at breast height and total height were measured from trees whose diameters were greater than 5 cm. The study workflow is shown in Figure 2 and detailed explanations are given in following sub-sections.

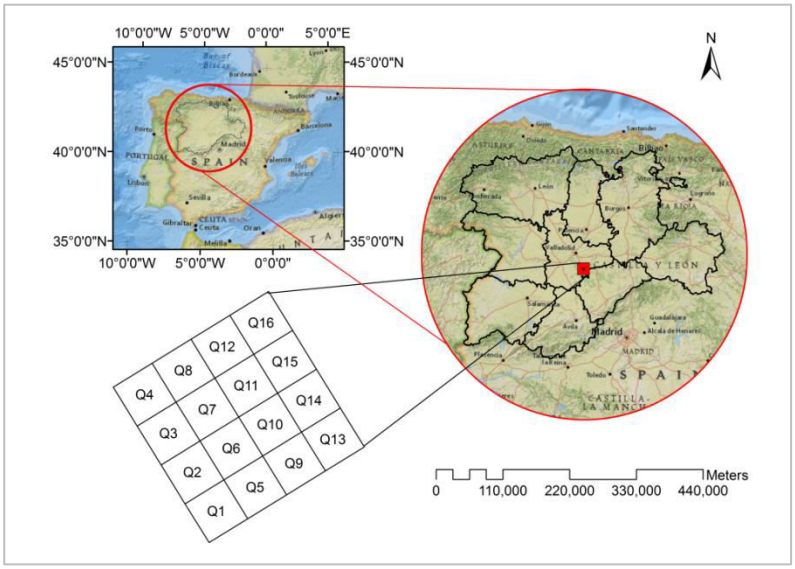


Figure 1. Study area location of Marteloscope of Llano de San Marugan (Valladolid, Spain)

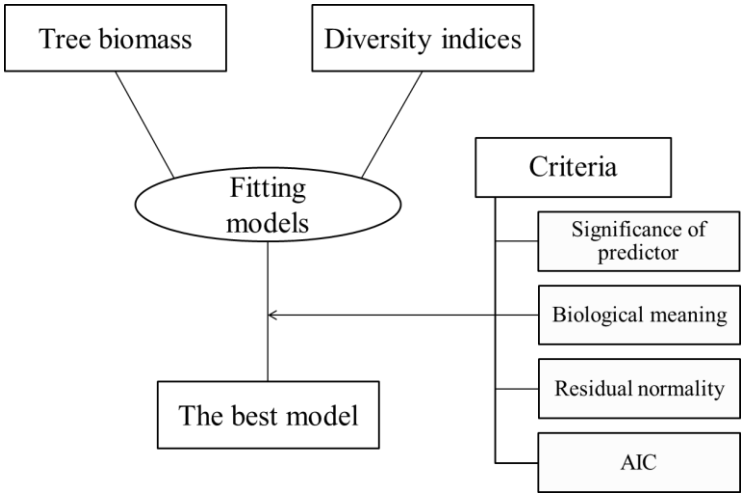


Figure 2. Workflow for the study of tree biomass and diversity relationship

3.2. Biomass estimation

Biomass in the different components of the tree e.g., stem and branches (thick, medium and thin+needles) and roots were calculated from dbh and height using existing relative allometric equations (Table 1) developed by (Ruiz-Peinado Gertrudix, Montero, & del Rio, 2012; Ruiz-Peinado, del Rio, & Montero, 2011). Tree component biomass values were computed for individual tree within each quadrat, and summed up to derive a summary of tree biomass for each quadrat. Total biomass obtained from sum of the biomass of all components.

Table 1. Biomass allometric equations by species

Species	Components	Model
Juniperus thuriferia L.	Stem	$W_s = 0.0132 \times d^2 \times h + 0.217 \times d \times h$
	Thick branches	$if \begin{cases} d \leq 22.5\ cm\ then\ Z = 0 \\ d \geq 22.5 cm\ then\ Z = 1 \end{cases}$ $W_{b7} = [0.107 \times (d - 22.5)^2] \times Z$
	Medium branches	$W_{b2-7} = 0.00792 \times d^2 \times h$
	Thin branches +needles	$W_{b2+n} = 0.273 \times d \times h$
	Roots	$W_r = 0.0767 \times d^2$
Pinus pinea	Stem	$W_s = 0.0224 \times d^{1.923} \times h^{1.0193}$
	Thick branches	$if \begin{cases} d \leq 22.5\ cm\ then\ Z = 0 \\ d \geq 22.5\ cm\ then\ Z = 1 \end{cases}$ $W_{b7} = [0.247 \times (d - 22.5)^2] \times Z$
	Medium branches	$W_{b2-7} = 0.0525 \times d^2$
	Thin branches +needles	$W_{b2+n} = 21.927 + 0.0707 \times d^2 - 2.827 \times h$
	Roots	$W_r = 0.117 \times d^2$
Quercus faginea	Stem	$W_s = 0.154 \times d^2$
	Thick branches	$W_{b7} = 0.0861 \times d^2$
	Medium branches	$W_{b2-7} = 0.127 \times d^2 - 0.00598 \times d^2 \times h$
	Thin branches + leaves	$W_{b2+1} = 0.0726 \times d^2 - 0.00275 \times d^2 \times h$
	Roots	$W_r = 0.169 \times d^2$
Quercus ilex	Stem	$W_s = 0.143 \times d^2$
	Thick branches	$if \begin{cases} d \leq 12.5 cm\ then\ Z = 0 \\ d \geq 12.5\ cm\ then\ Z = 1 \end{cases}$ $W_{b7} = [0.0684 \times (d - 12.5)^2 \times h] \times Z$
	Medium branches	$W_{b2-7} = 0.0898 \times d^2$
	Thin branches + leaves	$W_{b2+1} = 0.0824 \times d^2$
	Roots	$W_r = 0.254 \times d^2$

Source: (Ruiz-Peinado Gertrudix *et al*., 2012; Ruiz-Peinado *et al*., 2011).

3.3. Tree species diversity estimation

The diversity indices used in this study were classified into 3 categories (Table 2). The basic idea of a diversity index is to obtain a quantitative estimate of biological variability that can be used to compare biological entities, composed of discrete components, in space or in time (Morris *et al*., 2014).

Table 2. Categorization of indices

Richness & diversity	Species composition	Distribution pattern	
		Vertical	Horizontal
Simpson index (Sm)	Mingling (Mi)	Vertical Profile index (A)	Aggregation index (R)
Shannon index (Sn)	Spatial Diversity Status (MS)	Height Differentiation index (TH)	Uniform Angle Index (W
Evenness index (E)	Segregation index (S)		
Berker-Parker index (D)			

3.3.1. Species richness indices

Two different aspects are generally used to conceptualize the diversity in a community: species richness and evenness. Species richness represents the number of species or attributes present in a community which is the simplest and most commonly applied metric. The distribution of individuals over species is called evenness. Additionally, species or trait abundance is also important for diversity, and the proportional abundance of species can be incorporated into indices that represent diversity.

3.3.1.1. Simpson Diversity index (1-D)

The Simpson diversity index (Eq. 1) was introduced by Edward H. Simpson (Simpson, 1949) to measure species diversity in a community by taking into account the number of species present and the abundance of each species. The index represents the probability that two individuals that are randomly selected from a sample will belong to different species.

$$1 - D = 1 - \frac{\sum_{i=1}^{R} ni(ni-1)}{N(N-1)} \quad \textbf{Eq. 1}$$

where n_i is the number of individuals belonging to i-th type, *N* is total number of individuals in the dataset, *R* – richness (total number of species types in dataset).
It ranges 0 ≤ D ≤ 1. The value increases with species diversity. The higher the diversity, the greater the value of D.

3.3.1.2. Shannon index

Shannon index ($H^{\backprime}$) by Shannon and Weaver (Shannon, 1948) is distance independent index to characterize the species diversity in a given stand. It takes into account both abundance and evenness of the existing species (Eq. 2).

$$H^{\backprime} = -\sum_{i=1}^{S} p_i \ln p_i \quad \textbf{Eq. 2}$$

where S is the number of species, p_i – proportion of i-th species in the total number of individuals of all species and calculated from individuals of i-th species divided by total number of individuals present (n/N), $\ln p_i$ is natural logarithm of this proportion. Its value

ranges from 0 to $\ln(S)$. When all species in the dataset are equally common, all *pi* values equal 1/S, and the Shannon index hence takes the value In(S). When all abundance is concentrated in one species, and the other species are very rare (even if there are many of them), its value reduces to 0. The value is 0 when only one species in the dataset.

3.3.1.3. Berker-Parker index (D)

Berker-Parker index (Berger & Parker, 1970) is a measure of the numerical importance of the most abundant species in the population (Eq. 3). It has an analytical relationship with the geometric series of the species abundance model and represents the proportional abundance of only the most abundant species in the population (Morris *et al*., 2014).

$$D = \frac{N_{max}}{N} \quad \textbf{Eq. 3}$$

where N_{max} is the number of individuals is in the most abundant species and *N* is the total number of individuals in the sample. The reciprocal of the index, 1/D, is often used, so that an increase in the value of the index corresponds an increase in diversity and a reduction in dominance.

3.3.1.4. Evenness index (E)

Species evenness (E) (Pielou, 1975) refers to how species are close to each other in numbers (del Río et al., 2018). It represents the degree to which individuals disturb closely among species in terms of number. E is not calculated independently, but rather derived from compound diversity measures such as , D indices, as they inherently contain richness and evenness components. In Eq. 3, H' is the number derived from the Shannon diversity index and H'_{max} is the maximum possible value of H' (if every species was equally likely). *E* is supposed to be independent of a measure of species richness.

$$E = \frac{H'}{\ln S} \quad \textbf{Eq. 4}$$

where H' is a value of Shannon diversity index, $\ln S$ is natural logarithm of the number of species which equals to H'_{max}. Its value falls between 0 and 1 (1 demonstrates complete evenness). Low values indicate that one or a few species dominate, and high values indicate that relatively equal numbers of individuals belong to each species.

3.3.2. Species intermingling

The spatial relationships between two groups of individuals play important role for many components of a species' population biology. A numerous different types of tests indices have been designed to seek for an answer to the question whether two species are spatially segregated (individuals occur near the same species), associated (individuals occur near the other species), or neither.

3.3.2.1. Species segregation (S) index

Segregation index (S) developed by Pielou (1977) describes the degree of intermingling of two species groups based on nearest-neighbor method. S considers the ratio of the observed probability (p_{ij}) that reference tree *i* and its nearest-neighbor *j* belong to different species along with the same probability for completely randomly distributed or independent species attributes (del Río et al., 2018) (Eq. 5). There are 2 main procedures to calculate S index: 1) calculation between distances between reference

trees *i* to every tree in the plot which derived from Euclidean distance calculation. Once the distances were computed, trees were ranked from nearest to farthest to reference tree and the first n-th number of neighboring trees (which are user dependent) were selected, 2) computation of S index: which was computed based on the nearest-neighbor tree distances calculated in 1st step. S is originally designed for being applied to a two-species mixture (Biber & Weyerhaeuser, 1998). In Pielou's approach, a contingency table is constructed in form described in Table 3.

$$S = 1 - \frac{p_{ij}}{E(p_{ij})} = 1 - \frac{Observed\ number\ of\ mixed\ pairs}{Expected\ number\ of\ mixed\ pairs} \quad \textbf{Eq. 5}$$

P_{ij} and $E(p_{ij})$ can be solved by Eq. 6:

$$S = 1 - \frac{N * (b + c)}{mw + nv} \quad \textbf{Eq. 6}$$

where: *m* and *n* are the observed number of individual trees of species 1 and 2 respectively. *N* can easily be extracted from sum of m and n as described in the table. The *v* and *w* are the number of individual trees of species 1 and 2 that are found as the nearest-neighbors of a reference tree. These variables are clearly described in a contingency table (Table 3).

Table 3. Descriptions of variables for S index calculation

		Nearest-neighbor species (j)		
		species.1	species.2	Total (i)
Reference species (i)	species.1	a	b	m=a+b
	species.2	c	d	n=c+d
	Total (j)	v=a+c	w=b+d	N=m+n

If the nearest-neighbors are always the same species as the reference trees, then S=1 which implies that the reference tree is associated with itself. There is a segregation of reference species from others. If all neighbors are different species, S=-1 which indicates that the reference tree is associated with other species. There is association between 2 species. Independent distribution of species is indicated by value near to 0.

$$S = \begin{cases} 1, all\ nearest\ neighbors\ (j)\ are\ the\ same\ species\ with\ reference\ tree\ (i) \\ -1, all\ nearest\ neighbors\ are\ the\ different\ than\ reference\ tree\ (i) \end{cases} \quad 1 \geq S \geq -1$$

3.3.2.2. Mingling (Mi) index

Species spatial mingling (*Mi*) index is a measure of species diversity within a structure unit (neighbor trees plus reference tree) which describes the proportion of neighbor trees which don't belong to same species as the reference tree. The *Mi* by Füldner (1995) is defined as in Eq. 7 :

$$M_i = \frac{1}{n}\sum_{i}^{n} V_{ij} \quad \textbf{Eq. 7}$$

where n is the number of nearest neighbor trees considered, V_{ij} produces binary output which equals to 1 if the *j*-th neighbouring tree is not the same species as the i-th reference tree and $V_{ij} = 0$ otherwise.

$$V_{ij} = \begin{cases} 0, neighbour(j)\ belong\ to\ the\ sampe\ species\ as\ reference\ tree\ (i) \\ 1, neighbour(j)\ belong\ to\ different\ species\ from\ reference\ tree\ (i) \end{cases} \quad 0 \le M_i \le 1$$

A low degree of mingling indicates that trees of a particular species occur together with few or no trees of different species in the same area. High degree of mingling means that trees are surrounded by different species. Assume there are 4 neighbor trees to a reference tree, 5 different outputs are possible to derive as shown in Figure 3. The distribution of the *Mi* values, in conjunction with the species proportions within a given tree population, allows a detailed study of the spatial diversity within a forest. However, the number of different species in the structure unit was not taken into account, and this was a shortcoming of the *Mi* index. This shortcoming has fulfilled in spatial diversity status index.

3.3.2.3. Spatial diversity status (MS)

MS_i is improvement of mingling index. It considers not only the spatial mingling, but also the number of tree species. MS_i is determined by the relative species richness within the structure unit *i* and the degree of mingling of the reference tree and expressed by Eq. 8 (Gadow & Hui, 2002). The structural unit is defined by the neighborhoods that consisting a reference tree and its nearest neighbors (Zhang *et al.*, 2018).

$$MS_i = \frac{S_i}{n_{max}} \times M_i \qquad \textbf{Eq. 8}.$$

Where S_i is the number of tree species in the neighborhood of the reference tree i, including tree i, and n_{max} is the maximum number of species in this structure unit. M_i is the species mingling value. MS measures the tree species richness as well as an important species characteristic within a structure unit. Reference tree of a common species is more likely to have the neighbors of the same species, reflecting low MS value. R rare species have less probability to have same neighbor species, resulting in high value of MS. Thus, MS is considered as an index that sensitive to rare species.

Determinations of S_i and n_{max} for MS are given in Figure 3 where explanation are based on example images of Mi. For 4 nearest neighbors, the structural unit is considered to be 5 (4 neighbors plus reference tree). There are 4 different species (a, b, c, d) in the structure unit. So n_{max} is defined by 4. S_i can be calculated as a number of species within the structure unit as exemplified in Figure 3.

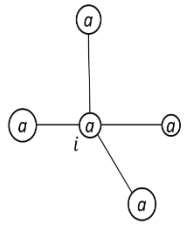
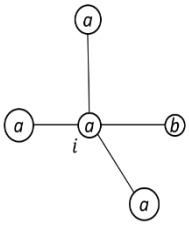
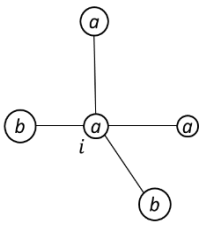
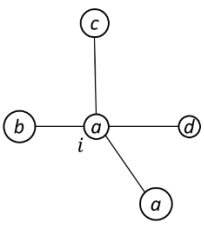
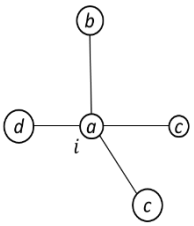
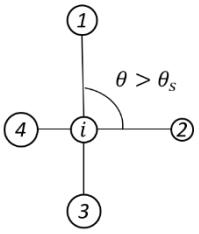
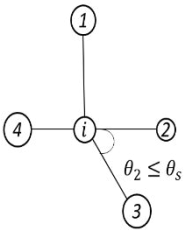
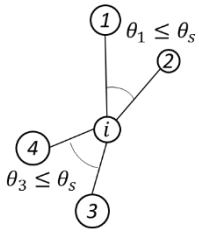
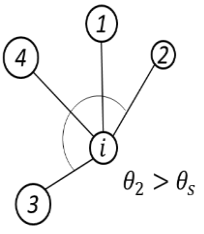
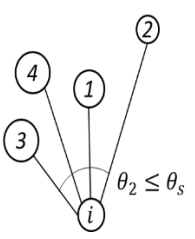
Mi index					
	$M_i = 0/4 = 0$	$M_i = 1/4 = 0.25$	$M_i = 2/4 = 0.5$	$M_i = 3/4 = 0.75$	$M_i = 04/4 = 1$
	None of the neighbor tree is different species than reference tree	One of the neighbor tree is different species than reference tree	Two of the neighbor tree are different species than reference tree	Three of the neighbor trees are different species than reference tree	All 4 neighbor trees are different species than reference tree
	No mixture	Low mixture	Medium mixture	High mixture	Complete mixture
MS index	S_i/n_{max} = 1/4	S_i = 2/4	S_i = 2/4	S_i = 4/4	S_i = 4/4
	Ther is 1 species (a) in the structure unit/max number of species in this structure unit is 4.	There are 2 species (a, b) in the structure unit/ max number of species in this structure unit is 4.	There are 2 species (a, b) in the structure unit/ max number of species in this structure unit is 4.	There are 4 species (a, b, c, d) in the structure unit/ max number of species in this structure unit is 4.	There are 4 species (a, b, c, d) in the structure unit/ max number of species in this structure unit is 4.
W index					
	$W_i = 0/4 = 0$	$W_i = 1/4 = 0.25$	$W_i = 2/4 = 0.5$	$W_i = 3/4 = 0.75$	$W_i = 04/4 = 1$
	None of the angles (θ) is smaller than standard angle (θ_s)	One of the angles (θ) is smaller than standard angle (θ_s)	Two of the angles (θ) are smaller than standard angle (θ_s)	Three of the angles (θ) are smaller than standard angle (θ_s)	All of the angles (θ) is smaller than standard angle (θ_s)
	Very regular	Regular	Random	Clumped	Very clumped

Figure 3. Description and calculation of Mingling index (Mi), Spatial diversity status (MS) and Uniform Angle index (Wi) and corresponding likelihood values for structure unit of 4 neighbors around the reference *i* tree: a, b, c, d are the tree species types ; $\boldsymbol{\theta}$ are angles between adjacent neighbor trees; $\boldsymbol{\theta_s}$ is a standard angle (which is equal to 360/4 for 4 neighbor trees) (Adapted from Gadow & Hui, 2002)

3.3.3. Spatial structural indices

The concept of spatial distribution consist vertical and horizontal spatial distributions which refer to the spatial arrangements (positioning) of different tree species along the

vertical or horizontal axis (del Río et al., 2018). The horizontal spatial distribution gives an idea of the variation of tree positioning (Bravo & Guerra, 2002). The indices that measure the horizontal spatial distribution quantifies the degree of regularity of the trees which are typically classified into regular, random, and clustered patterns and linked to processes of tree mortality, competitive interaction, regeneration and gap creation and so on. Vertical spatial distribution is most commonly described in terms of layers that refer to distinct classes or stratification of the canopy corresponding to height-related differentiation between trees.

3.3.3.1. Uniform Angle Index (W)

The Uniform Angle Index (W) formulates the degree of spatial dispersion of nearest neighbors around the reference tree based on angles between adjusting nearest neighbor trees defined as vectors from reference tree to each neighbors as shown in Figure 3 (Gadow *et al.*, 1998). W is determined as the proportion of the angles that are smaller than the standard angle α_0 (360/ n) and calculated as (Eq. 9):

$$W_i = \frac{1}{n}\sum_{j=1}^{n} v_j \quad \text{where } v_j = \begin{cases} 1, \alpha_j < \alpha_0 \\ 0, \text{otherwise} \end{cases} \qquad \textbf{Eq. 9}$$

where n is number of nearest neighbours

$$0 \le W \le 1; \quad \text{If} \begin{cases} W < 0.5, \ regular\ tree\ distributoin \\ 0.5 < W < 0.6, \ random\ distribution \\ W > 0.6, \ clumped \end{cases}$$

The value of W ranges from 0 to 1. The value of W increases from regular to clumped pattern (regular < random < clumped).

3.3.3.2. Aggregation index R

Aggregation index (R) by Clark & Evans (1954) is a single value index that is designed to describe aspects of variability of tree locations in forest stands (Eq. 10).

$$R = \frac{\bar{r}_{observed}}{E(r)} \quad \text{where } E(r) = \sqrt{\frac{A}{N}} \qquad \textbf{Eq. 10}$$

Where $\bar{r}_{observed}$ is an average distance to their nearest neighbours in a given forest stand while *E(r)* is an average nearest neighbor distance when trees completely random distributed, A is area of the plot, N is the total number of trees in the plot. The edge effect arising from the spatial limitations of experiment plots has minimized by applying the boundary correction factor by Donnelly (1978). Interpretation of R values is as follows:

$$0 \le R \le 2.149: \ R \begin{cases} < 1; \ then\ pattern\ shows\ clustering \\ \approx 1; then\ pattern\ is\ random\ distribution \\ > 1; then\ pattern\ is\ regular\ distribution \end{cases}$$

3.3.3.3. Height Differentiation index (TH)

Height differentiation index (TH) is size differentiation index, developed by Gadow (1993) which measures the variability in height between i-th reference tree to each neighboring trees (*j*-1…*n*) and describes vertical distribution of tree height (Eq. 11).

$$TH_{ij} = 1 - \frac{MIN(H_i, H_j)}{MAX(H_i, H_j)} \quad \textbf{Eq. 11}$$

where $H_i, \& H_j$ are the height of reference tree and neighbor tree respectively.

$0 \leq TH \leq 1$. If $\begin{cases} TH = 1, & neighbour\ trees\ have\ high\ differentiation\ in\ height \\ TH = 0, & neightbour\ trees\ have\ equal\ height \end{cases}$

3.3.3.4. Vertical species profile (A)

Vertical species profile (A) (Pretzsch, 1995) is outlined in Eq. 12. Calculation is based on the Shannon and Weaver (Shannon, 1948) diversity index. A considers both proportion of the species within a stand and the presence of each species in different height zones (Eq. 12). Height zones were determined as a same way as Pretzsch (2009).

$$A = -\sum_{i=1}^{S}\sum_{i=1}^{Z} p_{ij} \times \ln(p_{ij}) \quad \textbf{Eq. 12}$$

where S represents the number of species present, Z is the number of height zones (three in this case), p_{ij} is the proportion of a species in the height zone $p_{ij} = \frac{n_{ij}}{N}$, N is the total number of individuals, n_{ij} is the number of individuals of the species *i* in the zone *j*. Standardization of A can be done by dividing A value by the maximum value of the A index, i.e. $A_{max} = ln(S \times Z)$. Its value is greater than 0. For a pure stand with single layer, A equals to 0. Its value is increases as heterogeneous the vertical profile increases.

3.4. Statistical analysis

We used multiple linear regressions to evaluate the relationship between tree biomass and diversity indices in the stand, where total biomass (B) per tree was the dependent variable, and species richness, composition and species distribution indices were covariates of interest. Following previous researches that have developed a various regression models for estimating total-tree and tree compartment biomass, we utilized following three general forms of linear and non-linear regression equations (Eq. 13 to Eq. 15) for development of different forms of prediction models.

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_j + \varepsilon \quad \textbf{Eq. 13}$$

$$Y = \beta_0 {x_1}^{\beta_1} {x_2}^{\beta_2} \ldots {x_j}^{\beta_k} + \varepsilon \quad \textbf{Eq. 14}$$

$$Y = \beta_0 \pi (x_i x_j)^{\beta_k} \quad \textbf{Eq. 15}$$

Two different approaches of regression analysis were used to find the coefficients: the first approach was to apply, whenever possible, multiple linear regressions to the original equations; the second approach was to transform the above equations to the logarithmical form and then apply multiple linear regressions to the transformed equations. Eq. 13 used to develop multiple linear regressions that can be fitted by standard least squares estimation. Non-linear models (Eq. 14 and Eq. 15) were transformed into linear models (Eq. 16 and Eq. 17) by taking the logarithm of both sides of the equation. In this form, the equation parameters can easily be estimated by least squares procedures.

$$\ln Y = \ln \beta_0 + \beta_1 \ln x_i + \cdots \beta_2 \ln x_j + \varepsilon \quad \textbf{Eq. 16}$$

$$\ln Y = \ln \beta_0 + \beta_1 \ln(x_i * x_j) + \varepsilon \quad \textbf{Eq. 17}$$

where ln is the natural logarithm. ε is the random error term which is assumed to be normally distributed with mean zero and variance constant.

The models structures and their corresponding predictor variables are given in Table 4. 11 predictor variables: four species richness indices (Sm, Sn, D, E), three species composition indices (Mi, MS, S), four spatial distribution indices: A, TH, W and G were used as predictor variables for fitting models. Several different ways were implemented in variable selection process for the fitting models in order to avoid a problem of collinearity. First, all the variables were used as a single predictor variable for the models with single term. Second, basal area, species richness indices and species composition indices were utilized as state variables individually and each one of the spatial distribution indices were added into the multivariable models with two terms as secondary predictor variable. Third, we tested G, TH or A as a second, other individual indices as a third predictor variable for the multivariable models with three terms. Finally all possible combinations of indices are examined for multivariable models as well. In total, 537 alternative models were examined for each individual species and community level. For the community level analyze, basal area per quadrant G (m2/ha), for the species level analyze, species proportion of basal area G_p per quadrant were explored.

For the best models selection, four criteria were employed: i) significance of variables (in the ANOVA analysis, an effect is concerned to be significant when its coefficients have a probability less than or equal to the significant probability ($P < 0.05$), ii) biological meaning of parameters, iii) the normality of the residuals with Q-Q plots and iv) Akaike Information Criteria (AIC).

Table 4. Fitting models and their predictor variable reference

#	Fitted models	Predictor variable		
		x_1	x_2	x_3
Single predictor (33 alternative models)				
1.	$B = \beta_0 + \beta_1 x_1$	G, Sm, Sn, D, E, Mi, MS, S, A, TH, W		
2.	$\ln B = \beta_0 + \beta_1 \ln x_1$			
3.	$\ln B = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln({x_1}^2)$			
Multivariate models with 2 predictors (360 alternative models)				
4.	$B = \beta_0 + \beta_1 x_1 + \beta_2 x_2$	G, G_p	Mi, MS, S, A, TH, W	
5.	$B = \beta_0 + \beta_1 {x_1}^2 + \beta_2 {x_2}^2$	Sm		
6	$B = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$	Sn		
7	$B = \beta_0 + \beta_1 (x_1 * x_2)$	D		
8	$\ln B = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2$	E		
9	$\ln B = \beta_0 + \beta_1 \ln({x_1}^2) + \beta_2 \ln({x_2}^2)$	Mi	A, TH, W	
10	$\ln B = \beta_0 + \beta_1 \ln(x_1 * x_2)$	MS		
11	$\ln B = \beta_0 + \beta_1 \ln({x_1}^2 * x_2)$	S		
Multivariate models with 3 predictors (144 alternative models)				
12	$B = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$	G	TH	Sm, Sn, D, E, Mi, MS, S, A, W
13	$\ln B = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \beta_3 \ln x_3$		A	
14	$\ln B = \beta_0 + \beta_1 \ln(x_1 * x_2 * x_3)$	D, E, A, Sm, S, TH,	MS, S, TH, S, W,	

where B represents total biomass; x_1, x_2, x_3 represent the predictor variables; β_0, β_1, β_2, β_3 are the parameters of the models; For the species level analyze, basal area proportion of each species in each quadrant, for the whole community level analyze, basal area of each quadrant (m2/ha) were explored along with other indices.

4. - RESULTS

From the result of 537 investigated models, 11 significant models for community level are summarized in Table 5. Parameters of all models were statistically significant at ($p < 0.01$). TH alone or in combination with species richness or composition indices i.e Sm+TH, Sm*TH, Sn*TH, D+TH, E+TH, E*TH, Mi+TH, MS+TH, S+TH, G+TH+Sm, G+TH+MS was a main explanatory variable in all models except model 10 demonstrating the importance of vertical structural on tree biomass. According to parameter estimation, negative parameters were associated to species richness and compositional indices (Sm in model 1 and 4; MS in model 2 and 10; Mi in model 9; S in model 11 in logarithmic form) while positive parameters were correspond to TH in all the models excluding models 5 and 8 which implies that biomass increases as species richness or species composition decreases and height heterogeneity increases. From the selected 11 models, model 6 (Eq. 18) was found to be the best model for predicting community level tree biomass, showing negatively influence of D and positively influence of TH on tree biomass of the stand.

$$\mathrm{Ln}\, B = 6.594 - 1.349 \ln D + 0.841 \ln TH \qquad \textbf{Eq. 18}$$

Where B is total biomass per tree (kg), D is Berker-Parker index, TH is height differentiation index.

Table 6 showed that the parameter estimates of selected models by species based on models in Table 4**.** Similar trend that had observed at community level occur for *Pinus pinea*: TH alone or in combination with species richness or composition indices such as Sm+TH; Sm*TH; Sn+TH; Sn*TH; D+TH; D*TH; E*TH, Mi*TH was a main explanatory variable for all the models. The strength of the relationship ranges from 0.02 to 0.16 ($0.16 < R^2 < 0.2$). The highest statistically significant and the lowest AIC value belong to Eq. 19 (model 6, Table 6). The same form of model as community level analysis has been chosen as a best model for *Pinus pinea* as well. The reciprocal interactive effect of D and TH had the best prediction power on biomass. Gp has been shown to be the best explanatory variable for the biomass for both *Quercus faginea* (Eq. 20) and *Quercus ilex* (Eq. 21). None of the model was significant for *Juniperus thurifera* as concerned by models in this study.

$$\ln B = 7.249 - 0.935 \ln D + 0.988 \ln TH \qquad \textbf{Eq. 19}$$

$$\ln B = 4.278 + 0.156 \ln G_p \qquad \textbf{Eq. 20}$$

$$\ln B = 4.241 + 0.168 \ln G_p \qquad \textbf{Eq. 21}$$

Table 5. Parameter estimates for biomass from selected models at community level

Model	Explainatory variable											N	Adj.R2	AIC	RSE	F.value
	Intercept	*logG*	*logD*	*logTH*	*logSm*	*logMS*	*log(Sm*TH)*	*logE*	*logMi*	*logS*	*log(E*TH)*					
1	2.929*** (0.638)	0.812*** (0.217)		0.763*** (0.098)	-11.142*** (2.953)							437	0.20	1264.2	1.02	37.809***
2	4.426*** (0.594)	0.706*** (0.213)		0.990*** (0.097)		-2.028*** (0.465)						437	0.21	1259.6	1.02	39.746***
3	5.684*** (0.137)			0.883*** (0.096)								437	0.16	1283.9	1.05	84.958***
4	5.191*** (0.207)			0.797*** (0.099)	-9.264*** (2.954)							437	0.18	1276.1	1.04	48.260***
5	5.717*** (0.142)						0.881*** (0.097)					437	0.16	1285.8	1.05	82.659***
6	6.594*** (0.218)		-1.349*** (0.257)	0.841*** (0.093)								437	0.21	1258.9	1.02	58.877***
7	6.149*** (0.185)			0.894*** (0.095)				32.906*** (8.965)				437	0.19	1272.5	1.03	50.434***
8	5.701*** (0.138)										0.887*** (0.096)	437	0.16	1283.2	1.05	85.791***
9	6.466*** (0.217)			1.038*** (0.100)					-1.319*** (0.288)			437	0.20	1265.2	1.02	54.926***
10	6.288*** (0.197)			1.001*** (0.098)		-1.973*** (0.470)						437	0.19	1268.5	1.03	52.898***
11	5.963*** (0.146)			0.710*** (0.101)						-2.125*** (0.452)		437	0.20	1264.2	1.02	55.575***

****p<0.01*

Table 6. Parameter estimates for biomass from selected models at species level

Model	Explanotory variable															N	Adj.R2	AIC	RSE	F.value
	Intercept	*logTH*	*logSm*	*log(Sm*TH)*	*logSn*	*log(Sn*TH)*	*logD*	*log(D*TH)*	*log(E*TH)*	*log(Mi*TH+1)*	*logG*	*log(E+1)*	*log(E+1)^2*	*log(D+1)*	*log(D+1)^2*					
Pinus pinea																				
1	6.575*** (0.191)	0.937*** (0.163)														177	0.16	488.5	0.95	33.191***
2	6.158*** (0.249)	0.926*** (0.160)	-8.738** (3.391)													177	0.18	483.8	0.94	20.452***
3	6.596*** (0.199)			0.917*** (0.163)												177	0.15	489.9	0.96	31.494***
4	7.986*** (0.662)	0.923*** (0.161)			-1.230** (0.553)											177	0.17	485.5	0.94	19.444***
5	5.500*** (0.074)					0.774*** (0.162)										177	0.11	497.5	0.98	22.873***
6	7.249*** (0.289)	0.988*** (0.160)					-0.935*** (0.306)									177	0.19	481.2	0.93	22.060***
7	5.782*** (0.098)							0.532*** (0.145)								177	0.07	506.2	1.00	13.398***
8	6.594*** (0.193)								0.943*** (0.163)							177	0.16	488.1	0.95	33.613***
9	5.004*** (0.129)									2.864*** (0.556)						177	0.13	494.3	0.97	26.489***
Quercus faginea																				
1	4.278*** (0.170)										0.156** (0.072)					157	0.023	183.2	0.43	4.632**
Quercus ilex																				
1	*4.241**** *(0.126)*										*0.168**** *(0.045)*					*69*	*0.157*	74.3	*0.403*	*13.702****

**p < 0.05; ***p < 0.01

5. - DISCUSSION

Our finding revealed that the variation of tree biomass can be accounted by negative influence of D and positive influence of TH indicating that specific dominance in the stand influenced negatively and height heterogeneity influenced positively on biomass at community level. For the species level analysis, the variation of biomass of *Pinus pinea* can be explained by the same model as community level analysis which highlighted the negative impact of D and the positive impact of TH on biomass. This results is agreement with Bohn & Huth, (2017) who examined an influence of species diversity and forest structure on aboveground biomass over a broad range of forest stands and found out a positive relation between forest structural diversity and forest productivity. The same result was found in Riofrío *et al*., (2017). Size heterogeneity enables bigger trees to obtain greater amount of a certain resource and use them more efficiently than small trees (Brockerhoff et al., 2017b). In our stand, mixture of the *Pine, Quercus faginea, Quercus ilex and Juniperus thurifera* might create a different canopy strata; top layer occupied by *Pinus pinea*, enabling the light-demanding species – *Pine* - to capture more light and grow better than other species and become dominant in the stand.

The variation of biomass for *Quercus faginea* and *Quercus ilex* were predicted by basal area proportion of species (Gp). The examined diversity indices such as Sm, Sn, D, E, Mi, MS, S, W, A, TH were found not significant relation with biomass of *Quercus*. This might be explained by abundance of *Pinus pinea.* The abundance of *Pine* have an strong inhibitory effect on the abundance and richness of understory species through light, water, and soil nutrients and thus reversely influences on biomass productivity of understory species (Laughlin & Grace, 2006). Moreover, biomass of individual tree in a given stand is not only the reflection of diversity (species richness, composition and structure) but also various internal and external factors such as age, stand density, site productivity, competition at the tree level, climate, soil (texture, moisture content), geographical location, and length of grown season (Con *et al*., 2013; Poudel & Hailemariam, 2015) which might not be reflected by the available metrics or models that we are considered in this study.

One notable thing is that almost all the predictor variables of significant models (except model 1 in Table 6) for *Pinus pinea* were identical with the community level models in Table 5. This might be fact that community level analysis may be influenced by characteristic of *Pinus pinea* due to its dominance in the stand. This can be explained by "selection effect" which describes the impact of the most productive species on relationship between species richness and productivity. Positive relation between productivity (biomass accumulation) and tree diversity largely depends on presenting highly productivity species in multi-cultural communities (Tilman, 1999). However, in our case the proportion of the most productive, dominant species (*Pinus pinea*) showed a negative effect on biomass stand as represented by D in Eq. 18 & Eq. 19. As stress-tolerant and pioneer species, *Pinus* has the ability to become a dominant tree species in the mixed forest and accumulate biomass in a short time and it is considered as a strong competitor species with relatively high production due to its prolonged photosynthetic activity (coniferous evergreen tree) and high nutrient uptake through the rapid turnover of nutrients (Li, Su, Lang, Liu, & Ou, 2018). In old growth stand, *Pinus* with large diameter have a greater contribution to the stand biomass than small diameter trees (Baishya & Barik, 2011). In terms of complementarity of the species in this stand, *Pinus pinea* may facilitate development of *Quercus* by increasing seed protections, enhancing habitat condition which promote the recruitment of *Quercus* (Sheffer, 2012). Nevertheless, the facilitation of *Pine* for *Quercus* colonization depends on many factors such as stand density, development stage of *Quercus*, and environmental conditions. *Pine* forests with intermediate density enhance the site conditions for successful colonization of *Quercus* by reducing light intensity, creating partial shading and

improving the soil moisture status for *Quercus* seedlings. However, there is an opposition between suitable condition for recruitment and suitable condition for further growth and development of *Quercus* species that after seedlings and saplings stages (Puerta-Piñero, Gómez, & Valladares, 2007). In dense forest with poor environmental condition, although *Pinus* improved soil properties during a short period (a decade), it causes in decrement of soil moisture which may reduce recruitment of *Quercus* in their subcanopy. Low light interception levels and competition for water with *Pinus* reduces *Quercus* colonization in poor environmental condition. Maestre & Cortina, (2004) emphasized that *Pine* plantation in semi-arid environment don't facilitate the establishment of *Quercus* and causes in reduction of species richness and all plant cover. Therefore, based on these reviews, we can say that *Pinus pinea* might be a main contributor for biomass of the stand whereas its abundance has negatively impact on biomass of coexisting species i.e *Quercus faginea*, *Quercus ilex* and *Juniperus thurifera*.

6. - CONCLUSIONS

In this study, we examined the relationship between tree biomass and diversity in Llano de San Marugan, Valladolid, Spain at stand and individual species level by a various linear and non-linear models. After thoughtful examination of 537 models with 11 predictor tree diversity indices, we found out that there is a relation between tree biomass and diversity although this relation varies among the species. A stand level analyze revealed that tree biomass-diversity relation can be explained by an interaction of negative affect of abundance of dominant species and positive effect of tree' height heterogeneity which indicates that tree biomass increases as abundance of dominant species decreases and height heterogeneity increases. This result was identical for *Pinus pinea*. For *Quercus faginea* and *Quercus ilex,* only the species proportion of basal area (G_p) has a positive relation with biomass. The examined diversity indices were found not to have significant explanatory power for the explanation of biomass variation for *Quercus* species and *Juniperus thurifera*.

7. - AKNOWLEDGEMENTS

The author would like to greatly appreciate and acknowledge the following people for their help and support for success of this study:

- For supervisor Prof. Felipe Oviedo Bravo and co-supervisor Cristobal Ordonez Alonso for their enormous amount of contribution, guidance, critical comments, suggestion and helpful instructions from the initial to final stages which truly permits to successful completion of this work.
- For all the professors, staffs at University of Valladolid and University of Padova for sharing their valuable knowledge, skills and experiences which helped me a lot to strengthen my knowledge and experience in this science.
- For MEDfOR program for not only for the financial support but also letting us to build international network of friendship and living and experiencing in different countries with different culture.
- For all wonderful friends, labmates, postdoctoral students who are studying in the Department of Agriculture and Forestry at UVA for assisting and supporting me in many ways. Without their assistance and motivation, this work would not have been accomplished.

8. - REFERENCES

Baishya, R., & Barik, S. K. (2011). Estimation of tree biomass, carbon pool and net primary production of an old-growth Pinus kesiya Royle ex. Gordon forest in north-eastern India. *Annals of Forest Science*, *68*(4), 727–736. https://doi.org/10.1007/s13595-011-0089-8

Baskin, Y. (1994). Ecologists Dare to Ask: How Much Does Diversity Matter? *Science*, *264*(5156), 202–203. https://doi.org/10.1126/SCIENCE.264.5156.202

Berger, W. H., & Parker, F. L. (1970). Diversity of Planktonic Foraminifera in Deep-Sea Sediments. *Science*, *168*(3937), 1345–1347. https://doi.org/10.1126/science.168.3937.1345

Biber, P., & Weyerhaeuser, H. (1998). Numerical methods for characterizing structure and diversity applied to a natural tropical forest and an even aged teak stand. *… and Socio-Economic Analysis …*. Retrieved from http://www.wwk.forst.tu-muenchen.de/info/publications/OnlinePublications/482.pdf

Bohn, F. J., & Huth, A. (2017). The importance of forest structure to biodiversity–productivity relationships. *Royal Society Open Science*, *4*(1), 160521. https://doi.org/10.1098/rsos.160521

Bravo-Oviedo, Andres, Pretzsch, H., Ammer, C., Andenmatten, E., Barbati, A., Barreiro, S., … Zlatanov, T. (2014). European Mixed Forests: definition and research perspectives. *Forest Systems*, *23*(3), 518. https://doi.org/10.5424/fs/2014233-06256

Bravo, F., & Guerra, B. (2002). Forest Structure and Diameter Growth in Maritime Pine in a Mediterranean Area (pp. 123–134). Springer, Dordrecht. https://doi.org/10.1007/978-94-015-9886-6_10

Brockerhoff, E. G., Barbaro, L., Castagneyrol, B., Forrester, D. I., Gardiner, B., González-Olabarria, J. R., … Jactel, H. (2017a). Forest biodiversity, ecosystem functioning and the provision of ecosystem services. *Biodiversity and Conservation*, *26*(13), 3005–3035. https://doi.org/10.1007/s10531-017-1453-2

Brockerhoff, E. G., Barbaro, L., Castagneyrol, B., Forrester, D. I., Gardiner, B., González-Olabarria, J. R., … Jactel, H. (2017b). Forest biodiversity, ecosystem functioning and the provision of ecosystem services. *Biodiversity and Conservation*, *26*(13), 3005–3035. https://doi.org/10.1007/s10531-017-1453-2

Cardinale, B. J., Wright, J. P., Cadotte, M. W., Carroll, I. T., Hector, A., Srivastava, D. S., … Weis, J. J. (2007). Impacts of plant diversity on biomass production increase through time because of species complementarity. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 18123–18128. https://doi.org/10.1073/pnas.0709069104

Clark, P. J., & Evans, F. C. (1954). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology*, *35*(4), 445–453. https://doi.org/10.2307/1931034

Con, T. Van, Thang, N. T., Ha, D. T. T., Khiem, C. C., Quy, T. H., Lam, V. T., … Sato, T. (2013). Relationship between aboveground biomass and measures of structure and species diversity in tropical forests of Vietnam. *Forest Ecology and Management*, *310*, 213–218. https://doi.org/10.1016/J.FORECO.2013.08.034

Daily, G. C. (1997). *Nature's services : societal dependence on natural ecosystems*. Island Press.

del Río, M., Barbeito, I., Bravo-Oviedo, A., Calama, R., Cañellas, I., Herrero, C., … Bravo, F. (2017). Mediterranean Pine Forests: Management Effects on Carbon

Stocks (pp. 301–327). Springer, Cham. https://doi.org/10.1007/978-3-319-28250-3_15

del Río, M., Pretzsch, H., Alberdi, I., Bielak, K., Bravo, F., Brunner, A., … Bravo-Oviedo, A. (2018). Characterization of Mixed Forests BT - Dynamics, Silviculture and Management of Mixed Forests. In Andrés Bravo-Oviedo, H. Pretzsch, & M. del Río (Eds.) (pp. 27–71). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-91953-9_2

Evans, E. W. (2016). Biodiversity, ecosystem functioning, and classical biological control. *Applied Entomology and Zoology*. Springer. https://doi.org/10.1007/s13355-016-0401-z

Fraser, L. H., Pither, J., Jentsch, A., Sternberg, M., Zobel, M., Askarizadeh, D., … Zupo, T. (2015). Plant ecology. Worldwide evidence of a unimodal relationship between productivity and plant species richness. *Science (New York, N.Y.)*, *349*(6245), 302–305. https://doi.org/10.1126/science.aab3916

Füldner, K. (1995). *Strukturbeschreibung von Buchen-Edellaubholz-Mischwäldern. PhD, [Describing forest structures in mixed beech-ash-maple-sycamore stands*. University of Göttingen, Cuvillier Verlag Göttingen,.

Gadow, K.v., Hui, G.Y., Albert, M. (1998). Das Winkelmaß–ein Strukturparameter zur Beschreibung der Individualverteilung in Waldbesta¨nden [The uniform angle index—a structural parameter for describing tree distribution in forest stands]. *Centralbl. Ges. Forstwesen*, (115), 1–10.

Gadow, K.v. (1993). Zur Bestandesbeschreibung in der Forsteinrichtung. *Forst Und Holz*, *48*, 601–606.

Gadow Kv; Hui GY. (2002). No Title. In *Characterizing forest spatial structure and diversity*. Lund: Proceedings of the Sustainable Forestry in Southern Sweden (SUFOR) conference “Sustainable Forestry in Temperate Regions”,.

Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., … Bengtsson, J. (2013). Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms2328

Grace, J. B., Anderson, T. M., Seabloom, E. W., Borer, E. T., Adler, P. B., Harpole, W. S., … Smith, M. D. (2016). Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature*, *529*(7586), 390–393. https://doi.org/10.1038/nature16524

Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E. K., Hungate, B. A., Matulich, K. L., … O’Connor, M. I. (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, *486*(7401), 105–108. https://doi.org/10.1038/nature11118

Iranah, P., Lal, P., Wolde, B. T., & Burli, P. (2018). Valuing visitor access to forested areas and exploring willingness to pay for forest conservation and restoration finance: The case of small island developing state of Mauritius. *Journal of Environmental Management*, *223*, 868–877. https://doi.org/10.1016/j.jenvman.2018.07.008

Knoke, T., Stimm, B., Ammer, C., & Moog, M. (2005). Mixed forests reconsidered: A forest economics contribution on an ecological concept. *Forest Ecology and Management*, *213*(1–3), 102–116. https://doi.org/10.1016/j.foreco.2005.03.043

Kohn, A. J. (1977). Pielou, E. C. 1975. Ecological diversity. John Wiley & Sons, New York, viii + 165 p. $14.95. *Limnology and Oceanography*, *22*(1), 174–174. https://doi.org/10.4319/LO.1977.22.1.0174B

Laughlin, D. C., & Grace, J. B. (2006). A multivariate model of plant species richness in forested systems: old-growth montane forests with a long history of fire. *Oikos*, *114*(1), 60–70. https://doi.org/10.1111/j.0030-1299.2006.14424.x

Li, S., Su, J., Lang, X., Liu, W., & Ou, G. (2018). Positive relationship between species richness and aboveground biomass across forest strata in a primary Pinus kesiya forest. *Scientific Reports*, *8*(1), 2227. https://doi.org/10.1038/s41598-018-20165-y

Liang, J., Crowther, T. W., Picard, N., Wiser, S., Zhou, M., Alberti, G., … Reich, P. B. (2016). Positive biodiversity-productivity relationship predominant in global forests. *Science*, *354*(6309), aaf8957–aaf8957. https://doi.org/10.1126/science.aaf8957

Liu, C. L. C., Kuchma, O., & Krutovsky, K. V. (2018). Mixed-species versus monocultures in plantation forestry: Development, benefits, ecosystem services and perspectives for the future. *Global Ecology and Conservation*, *15*, e00419. https://doi.org/10.1016/J.GECCO.2018.E00419

Maestre, F. T., & Cortina, J. (2004). Are Pinus halepensis plantations useful as a restoration tool in semiarid Mediterranean areas? *Forest Ecology and Management*, *198*(1–3), 303–317. https://doi.org/10.1016/J.FORECO.2004.05.040

Mori, A. S., Lertzman, K. P., & Gustafsson, L. (2017). REVIEW: FOREST BIODIVERSITY AND ECOSYSTEM SERVICES Biodiversity and ecosystem services in forest ecosystems: a research agenda for applied forest ecology. *Journal of Applied Ecology*, *54*, 12–27. https://doi.org/10.1111/1365-2664.12669

Morris, E. K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T. S., … Rillig, M. C. (2014). Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution*, *4*(18), 3514–3524. https://doi.org/10.1002/ece3.1155

Ozcelik, R. (2009). Tree species diversity of natural mixed stands in eastern Black sea and western Mediterranean region of Turkey. *Journal of Environmental Biology*, *30*(5), 761–766.

Pielou, E. C. (1975). *Ecological diversity*. New York: Wiley.

Pielou, E. C. (1977). *Mathematical ecology*. New York: John Wiley & Sons, Ltd.

Poudel, K. P., & Hailemariam, T. (2015). *METHODS FOR ESTIMATING ABOVEGROUND BIOMASS AND ITS COMPONENTS FOR FIVE PACIFIC NORTHWEST TREE SPECIES*. Retrieved from https://www.fs.fed.us/pnw/pubs/pnw_gtr931/pnw_gtr931_004.pdf

Pretzsch, H. (2009). Forest Dynamics, Growth, and Yield. In *Forest Dynamics, Growth and Yield: From Measurement to Model* (pp. 1–39). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-88307-4_1

Pretzsch, H., & Schütze, G. (2014). Size-structure dynamics of mixed versus pure forest stands. *Forest Systems*, *23*(3), 560. https://doi.org/10.5424/fs/2014233-06112

Pretzsch, & H. (1995). Zum Einfluss des Baumvertielungsmusters auf den Bestandszuwachs. *Allg Forst-Jagd*, *166*, 190–201. Retrieved from https://ci.nii.ac.jp/naid/10029731191/#cit

Puerta-Piñero, C., Gómez, J. M., & Valladares, F. (2007). Irradiance and oak seedling survival and growth in a heterogeneous environment. *Forest Ecology and Management*, *242*(2–3), 462–469. https://doi.org/10.1016/J.FORECO.2007.01.079

Riofrío, J., del Río, M., Pretzsch, H., & Bravo, F. (2017). Changes in structural heterogeneity and stand productivity by mixing Scots pine and Maritime pine. *Forest Ecology and Management*, *405*, 219–228. https://doi.org/10.1016/J.FORECO.2017.09.036

Ruiz-Peinado Gertrudix, R., Montero, G., & del Rio, M. (2012). Biomass models to estimate carbon stocks for hardwood tree species. *Forest Systems; Vol 21, No 1 (2012)DO - 10.5424/Fs/2112211-02193* . Retrieved from http://revistas.inia.es/index.php/fs/article/view/2193

Ruiz-Peinado, R., del Rio, M., & Montero, G. (2011). New models for estimating the carbon sink capacity of Spanish softwood species. *Forest Systems; Vol 20, No 1 (2011)DO - 10.5424/Fs/2011201-11643* . Retrieved from http://revistas.inia.es/index.php/fs/article/view/1880

Sala, O. E., Chapin, F. S., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., … Wall, D. H. (2000). Global biodiversity scenarios for the year 2100. *Science (New York, N.Y.)*, *287*(5459), 1770–1774. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10710299

Scarascia-Mugnozza, G., Oswald, H., Piussi, P., & Radoglou, K. (2000). Forests of the Mediterranean region: gaps in knowledge and research needs. *Forest Ecology and Management*, *132*(1), 97–109. https://doi.org/10.1016/S0378-1127(00)00383-2

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sheffer, E. (2012). A review of the development of Mediterranean pine–oak ecosystems after land abandonment and afforestation: are they novel ecosystems? *Annals of Forest Science*, *69*(4), 429–443. https://doi.org/10.1007/s13595-011-0181-0

Simpson, E. H. (1949). Measurement of Diversity. *Nature*, *163*(4148), 688–688. https://doi.org/10.1038/163688a0

Szwagrzyk, J., & Gazda, A. (2007). Above-ground standing biomass and tree species diversity in natural stands of Central Europe. *Journal of Vegetation Science*, *18*(4), 555–562. https://doi.org/10.1111/j.1654-1103.2007.tb02569.x

Tilman, D, Lehman, C. L., & Thomson, K. T. (1997). Plant diversity and ecosystem productivity: theoretical considerations. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(5), 1857–1861. https://doi.org/10.1073/pnas.94.5.1857

Tilman, David. (1999). THE ECOLOGICAL CONSEQUENCES OF CHANGES IN BIODIVERSITY: A SEARCH FOR GENERAL PRINCIPLES. *Ecology*, *80*(5), 1455–1474. https://doi.org/10.1890/0012-9658(1999)080[1455:TECOCI]2.0.CO;2

Wang, H., Zhang, G., Hui, G., Li, Y., Hu, Y., & Zhao, Z. (2016). The influence of sampling unit size and spatial arrangement patterns on neighborhood-based spatial structure analyses of forest stands. *Forest Systems*, *25*(1). https://doi.org/10.5424/fs/2016251-07968

Zhang, L., Hui, G., Hu, Y., & Zhao, Z. (2018). Spatial structural characteristics of forests dominated by Pinus tabulaeformis Carr. *PLOS ONE*, *13*(4), 1–13. https://doi.org/10.1371/journal.pone.0194710

ANNEX

Annex 1. Descriptive statistics of stand variables (437 trees in the one hectare plot)

Variables	Mean	St.Dev	Min	Pctl(25)	Pctl(75)	Max
B	137.4	196.7	2.9	26.2	128.4	1.067.5
d	16.6	11.7	3.0	8.6	23.1	50.5
h	5.8	2.5	2.0	4.0	7.6	13.0
V	0.1	0.2	0.001	0.01	0.1	1.2
Gi (m2/ha)	14.3	3.1	8.8	12.5	16.8	19.4
Sm	1.0	0.02	0.9	1.0	1.0	1.0
Sn	3.3	0.4	2.4	3.2	3.5	3.9
E	1.0	0.01	1.0	1.0	1.0	1.0
D	2.1	0.4	1.2	1.9	2.3	2.6
Mi	0.6	0.3	0.0	0.5	0.8	1.0
MS	0.3	0.1	0.0	0.2	0.4	0.6
W	0.5	0.2	0.0	0.2	0.8	1.0
S	0.3	0.2	0.1	0.2	0.3	0.9
A	2.5	0.8	0.0	1.9	3.1	3.7
TH	0.3	0.1	0.0	0.2	0.4	0.7

B – total biomass (kg); d – breast height diameter (m); h – total height (m), V – volume (m^3), Sm - simson index; Sn – shannon index; E – evenness index; G- basal area (m2/ha), D – barker Parker index, Mi – mingling index; MS – spatial diversity status; W – uniform angle index; S – segregation index; A – vertical profile index; TH – height differentiation index.

Annex 2. Tree attributes by species

#	Species	N	d (cm)	h (m)	v (m3)	g (m2)	B (kg)	Sm	Sn	D	E	Mi	MS	S	W	A	TH
1	*Pinus pinea*	177	27.512	8.163	0.256	0.070	290.110	0.955	3.215	1.985	0.987	0.549	0.251	0.194	0.489	1.900	0.366
2	*Quercus faginea*	157	9.604	4.183	0.019	0.008	38.757	0.963	3.390	2.141	0.987	0.562	0.260	0.239	0.497	2.965	0.263
3	*Quercus ilex*	69	9.100	4.122	0.013	0.007	27.122	0.969	3.548	2.130	0.987	0.558	0.270	0.612	0.446	3.137	0.226
4	*Juniperus thurifera*	34	7.626	4.331	0.271	0.005	21.764	0.956	3.202	2.189	0.981	0.765	0.312	0.241	0.507	2.683	0.251

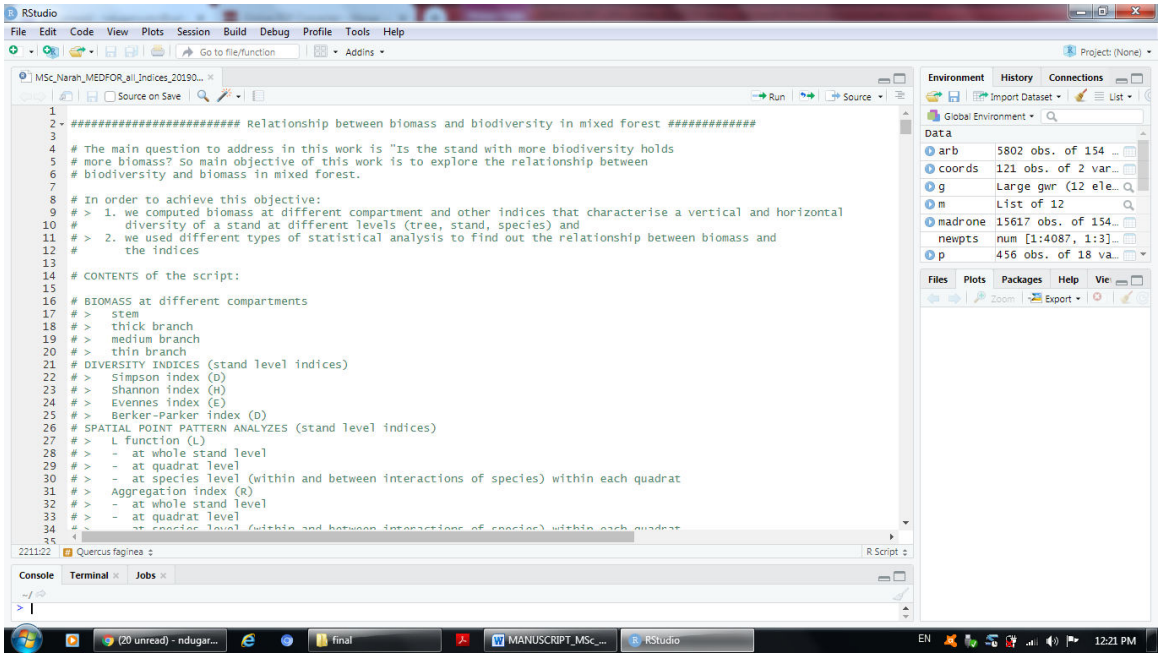
Average values are presented for the tree species variables: B- above-ground biomass (kg); d –tree diameter at breast height (cm); h- total height (m); v – volume (m^3); g – basal area per tree (m^2), Mi - Mingling index; MS – Spatial diversity status; A – vertical profile Index; W – Uniform angle index; S – segregation index; TH – height differentiate index; G – basal area per quadrant (m^2/ha); Sm – simpson index; Sn - shannon index; D- Berker- Parker index; E – evennes index; R – aggregation index, respectively.

Annex 3. Main attributes of tree by quadrant.

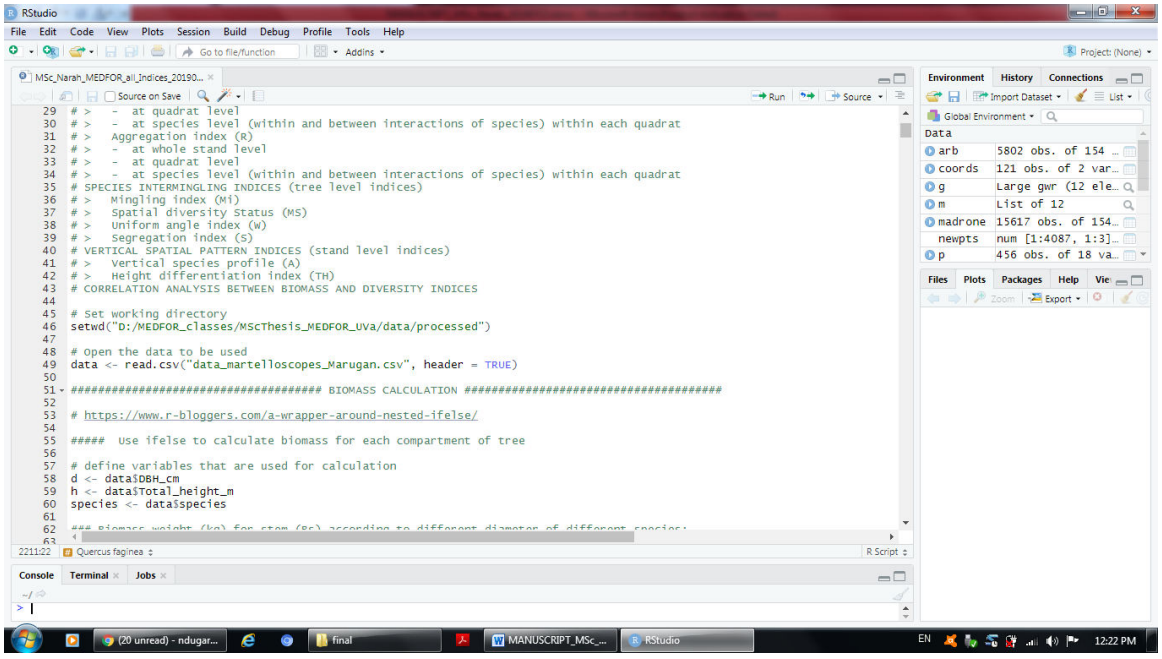
Quadrant number	Number of trees	Tree variables											Stand variables					
		B	*d*	*h*	*v*	*g*	*Mi*	*MS*	*A*	*W*	*S*	*TH*	*G*	*Sm*	*Sn*	*D*	*E*	*R*
1	26	110.4	14.1	5.4	0.18	0.027	0.55	0.25	2.29	0.47	0.21	0.29	11.06	0.96	3.18	2.60	0.98	1.02
2	31	127.5	16.0	5.7	0.14	0.030	0.65	0.27	2.56	0.52	0.21	0.29	14.73	0.97	3.40	1.94	0.99	1.27
3	21	152.0	16.5	6.0	0.19	0.034	0.73	0.35	2.25	0.48	0.22	0.36	11.36	0.95	2.97	2.33	0.97	1.00
4	34	132.8	16.4	5.6	0.18	0.032	0.71	0.34	2.57	0.54	0.22	0.30	16.81	0.97	3.41	2.29	0.98	0.92
5	12	272.9	25.8	7.5	0.38	0.065	0.50	0.18	1.59	0.52	0.17	0.41	12.55	0.91	2.47	1.33	0.99	0.99
6	29	130.3	16.5	5.6	0.25	0.031	0.59	0.26	2.62	0.45	0.18	0.33	14.27	0.96	3.33	1.93	0.99	1.15
7	28	140.6	17.5	5.9	0.11	0.033	0.65	0.29	2.41	0.54	0.24	0.33	14.80	0.96	3.29	2.15	0.99	1.00
8	53	83.4	13.1	5.0	0.07	0.020	0.53	0.24	3.35	0.53	0.39	0.21	16.63	0.98	3.94	2.21	0.99	1.06
9	26	93.5	12.9	5.0	0.17	0.021	0.64	0.29	2.52	0.49	0.21	0.28	8.76	0.96	3.20	2.17	0.98	1.12
10	13	353.7	28.6	8.0	0.21	0.081	0.31	0.12	1.56	0.38	0.18	0.36	16.89	0.92	2.54	1.18	0.99	1.08
11	23	173.9	17.8	6.0	0.04	0.038	0.60	0.30	2.42	0.38	0.29	0.33	14.14	0.95	3.08	2.30	0.98	0.89
12	36	66.7	12.6	5.0	0.09	0.016	0.39	0.18	2.81	0.53	0.53	0.22	9.46	0.97	3.52	1.89	0.98	0.97
13	11	365.9	29.9	8.3	0.36	0.085	0.55	0.22	1.54	0.45	0.16	0.37	14.93	0.91	2.38	1.38	0.99	1.21
14	25	175.8	20.8	7.1	0.18	0.044	0.41	0.16	2.13	0.49	0.29	0.31	17.48	0.96	3.18	1.47	0.99	1.03
15	31	105.5	14.8	5.3	0.01	0.025	0.60	0.29	2.71	0.36	0.34	0.33	12.46	0.97	3.40	2.58	0.99	1.02
16	38	131.3	17.1	6.2	0.02	0.032	0.62	0.32	2.61	0.51	0.28	0.30	19.44	0.97	3.60	2.11	0.99	1.22

Average values are presented for the tree stand variables: B- above-ground biomass (kg); d – tree diameter at breast height (cm); h- total height (m); v – volume (m^3); g – Basal area per tree (m^2), Mi - mingling index; MS – spatial diversity status; A – vertical profile Index; W – uniform angle index; S – segregation index; TH – height differentiate index; G – basal area per quadrant (m^2/ha); Sm – simpson index; Sn - shannon index; D- Berker-Parker index; E – evennes index; R – aggregation index, respectively.

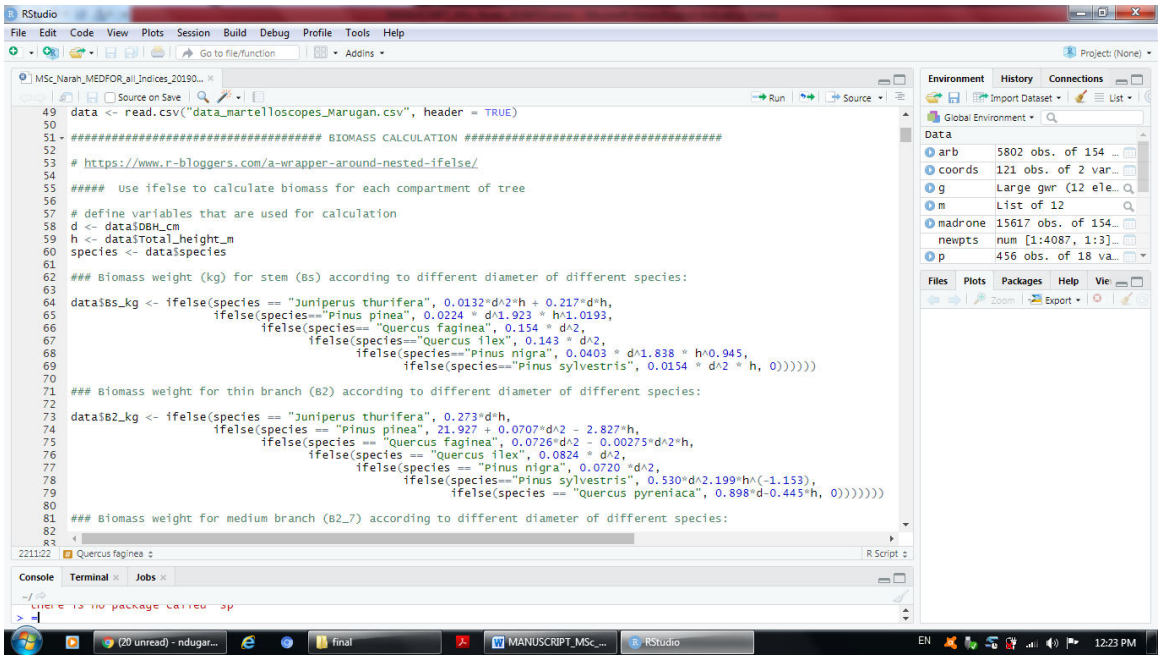
Annex 4. R scripts



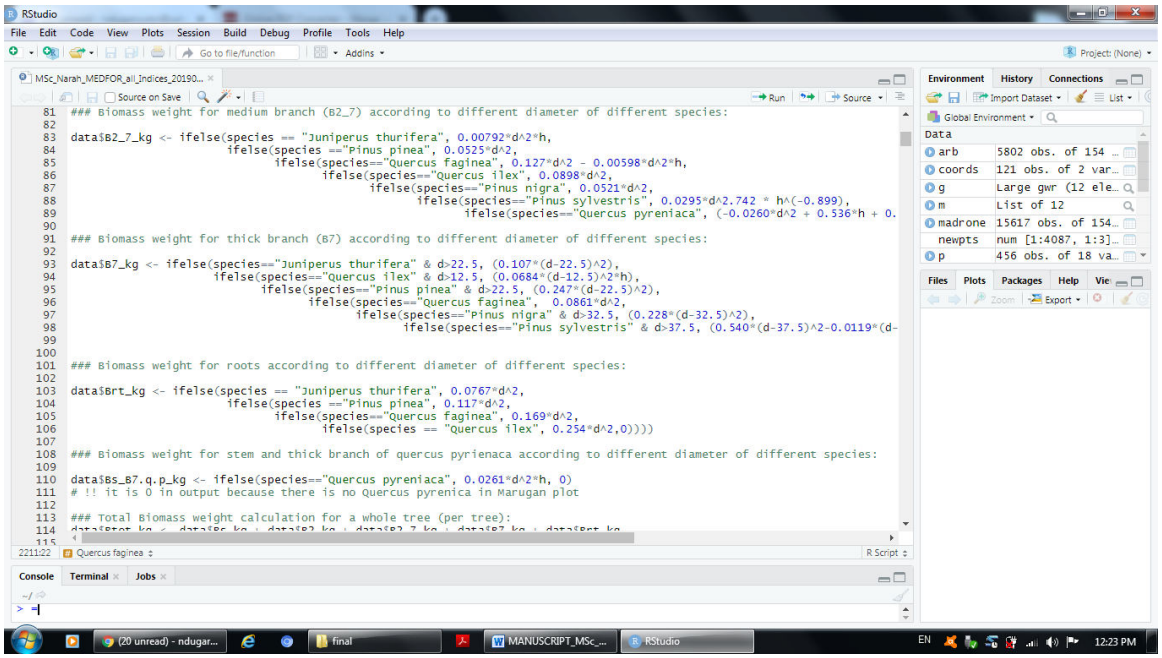
```
########################## Relationship between biomass and biodiversity in mixed forest #############

# The main question to address in this work is "Is the stand with more biodiversity holds
# more biomass? So main objective of this work is to explore the relationship between
# biodiversity and biomass in mixed forest.

# In order to achieve this objective:
# >  1. we computed biomass at different compartment and other indices that characterise a vertical and horizontal
#       diversity of a stand at different levels (tree, stand, species) and
# >  2. we used different types of statistical analysis to find out the relationship between biomass and
#       the indices

# CONTENTS of the script:

# BIOMASS at different compartments
# >   stem
# >   thick branch
# >   medium branch
# >   thin branch
# DIVERSITY INDICES (stand level indices)
# >   Simpson index (D)
# >   Shannon index (H)
# >   Evennes index (E)
# >   Berker-Parker index (D)
# SPATIAL POINT PATTERN ANALYZES (stand level indices)
# >   L function (L)
# >   -  at whole stand level
# >   -  at quadrat level
# >   -  at species level (within and between interactions of species) within each quadrat
# >   Aggregation index (R)
# >   -  at whole stand level
# >   -  at quadrat level
# >   -  at species level (within and between interactions of species) within each quadrat
```



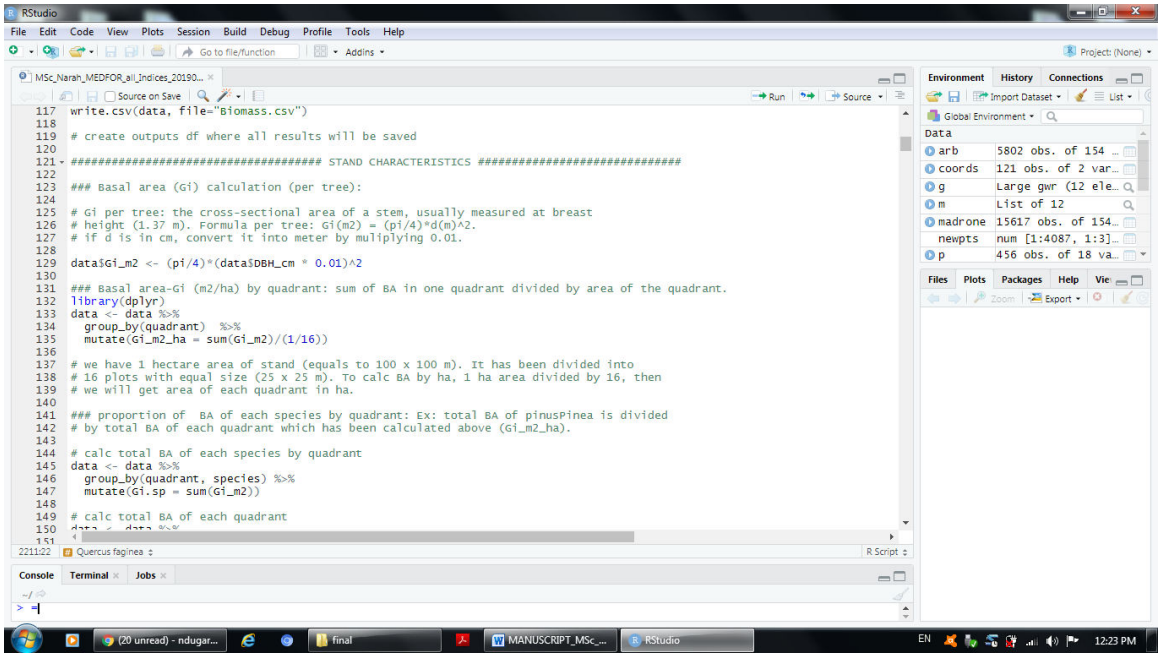
```
# >   -  at quadrat level
# >   -  at species level (within and between interactions of species) within each quadrat
# >   Aggregation index (R)
# >   -  at whole stand level
# >   -  at quadrat level
# >   -  at species level (within and between interactions of species) within each quadrat
# SPECIES INTERMINGLING INDICES (tree level indices)
# >   Mingling index (Mi)
# >   Spatial diversity Status (MS)
# >   Uniform angle index (W)
# >   Segregation index (S)
# VERTICAL SPATIAL PATTERN INDICES (stand level indices)
# >   Vertical species profile (A)
# >   Height differentiation index (TH)
# CORRELATION ANALYSIS BETWEEN BIOMASS AND DIVERSITY INDICES

# Set working directory
setwd("D:/MEDFOR_classes/MScThesis_MEDFOR_UVa/data/processed")

# Open the data to be used
data <- read.csv("data_martelloscopes_Marugan.csv", header = TRUE)

################################### BIOMASS CALCULATION ###################################

# https://www.r-bloggers.com/a-wrapper-around-nested-ifelse/

#####  Use ifelse to calculate biomass for each compartment of tree

# define variables that are used for calculation
d <- data$DBH_cm
h <- data$Total_height_m
species <- data$species
```



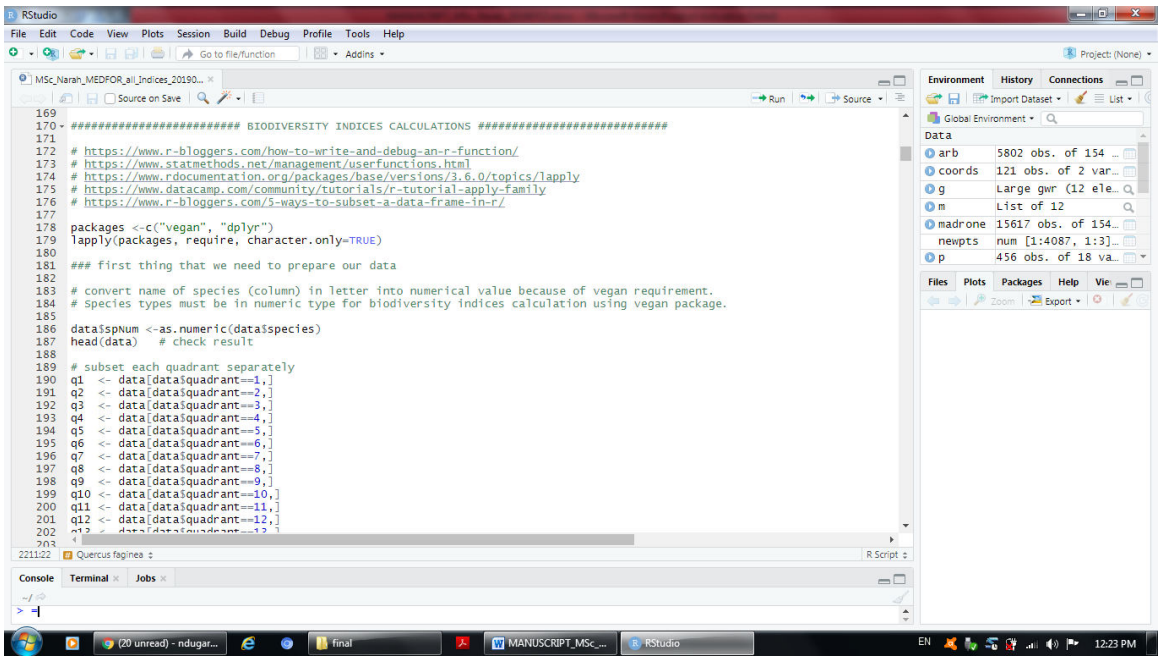
```
data <- read.csv("data_martelloscopes_Marugan.csv", header = TRUE)

################################### BIOMASS CALCULATION ###################################

# https://www.r-bloggers.com/a-wrapper-around-nested-ifelse/

#####  Use ifelse to calculate biomass for each compartment of tree

# define variables that are used for calculation
d <- data$DBH_cm
h <- data$Total_height_m
species <- data$species

### Biomass weight (kg) for stem (Bs) according to different diameter of different species:

data$Bs_kg <- ifelse(species == "Juniperus thurifera", 0.0132*d^2*h + 0.217*d*h,
                     ifelse(species=="Pinus pinea", 0.0224 * d^1.923 * h^1.0193,
                            ifelse(species== "Quercus faginea", 0.154 * d^2,
                                   ifelse(species=="Quercus ilex", 0.143 * d^2,
                                          ifelse(species=="Pinus nigra", 0.0403 * d^1.838 * h^0.945,
                                                 ifelse(species=="Pinus sylvestris", 0.0154 * d^2 * h, 0))))))

### Biomass weight for thin branch (B2) according to different diameter of different species:

data$B2_kg <- ifelse(species == "Juniperus thurifera", 0.273*d*h,
                     ifelse(species == "Pinus pinea", 21.927 + 0.0707*d^2 - 2.827*h,
                            ifelse(species == "Quercus faginea", 0.0726*d^2 - 0.00275*d^2*h,
                                   ifelse(species == "Quercus ilex", 0.0824 * d^2,
                                          ifelse(species == "Pinus nigra", 0.0720 *d^2,
                                                 ifelse(species=="Pinus sylvestris", 0.530*d^2.199*h^(-1.153),
                                                        ifelse(species == "Quercus pyreniaca", 0.898*d-0.445*h, 0)))))))

### Biomass weight for medium branch (B2_7) according to different diameter of different species:
```



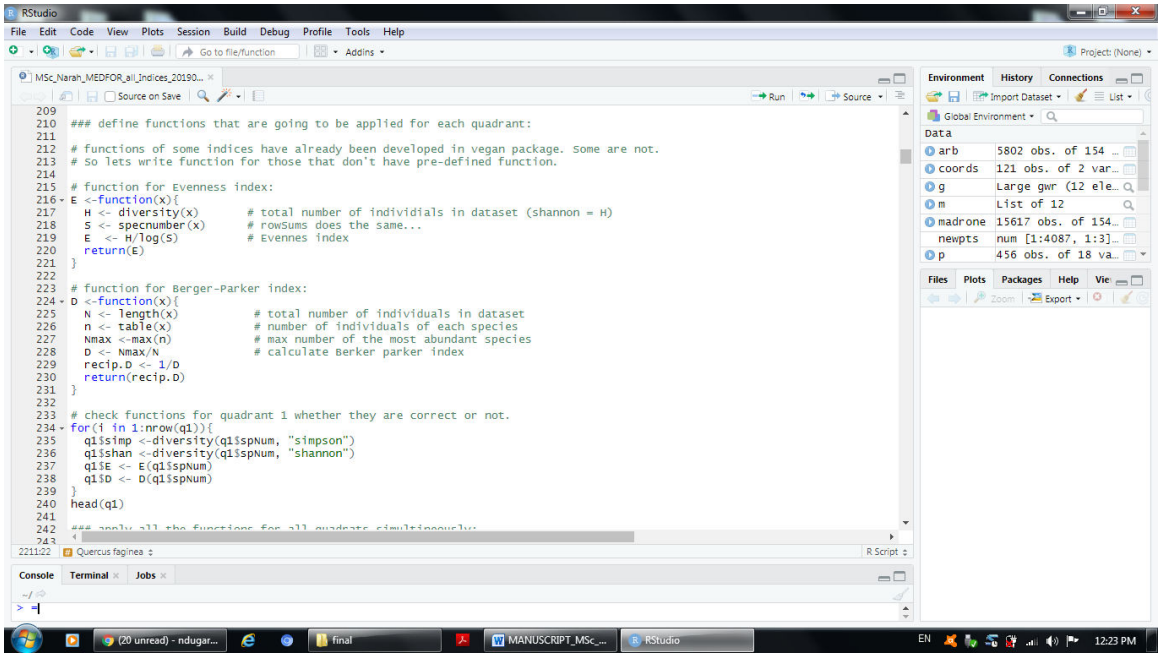
```
### Biomass weight for medium branch (B2_7) according to different diameter of different species:

data$B2_7_kg <- ifelse(species == "Juniperus thurifera", 0.00792*d^2*h,
                       ifelse(species =="Pinus pinea", 0.0525*d^2,
                              ifelse(species=="Quercus faginea", 0.127*d^2 - 0.00598*d^2*h,
                                     ifelse(species=="Quercus ilex", 0.0898*d^2,
                                            ifelse(species=="Pinus nigra", 0.0521*d^2,
                                                   ifelse(species=="Pinus sylvestris", 0.0295*d^2.742 * h^(-0.899),
                                                          ifelse(species=="Quercus pyreniaca", (-0.0260*d^2 + 0.536*h + 0.

### Biomass weight for thick branch (B7) according to different diameter of different species:

data$B7_kg <- ifelse(species=="Juniperus thurifera" & d>22.5, (0.107*(d-22.5)^2),
                     ifelse(species=="Quercus ilex" & d>12.5, (0.0684*(d-12.5)^2*h),
                            ifelse(species=="Pinus pinea" & d>22.5, (0.247*(d-22.5)^2),
                                   ifelse(species=="Quercus faginea",  0.0861*d^2,
                                          ifelse(species=="Pinus nigra" & d>32.5, (0.228*(d-32.5)^2),
                                                 ifelse(species=="Pinus sylvestris" & d>37.5, (0.540*(d-37.5)^2-0.0119*(d-


### Biomass weight for roots according to different diameter of different species:

data$Brt_kg <- ifelse(species == "Juniperus thurifera", 0.0767*d^2,
                      ifelse(species =="Pinus pinea", 0.117*d^2,
                             ifelse(species=="Quercus faginea", 0.169*d^2,
                                    ifelse(species == "Quercus ilex", 0.254*d^2,0))))

### Biomass weight for stem and thick branch of quercus pyrienaca according to different diameter of different species:

data$Bs_B7.q.p_kg <- ifelse(species=="Quercus pyreniaca", 0.0261*d^2*h, 0)
# !! it is 0 in output because there is no Quercus pyrenica in Marugan plot

### Total Biomass weight calculation for a whole tree (per tree):
```



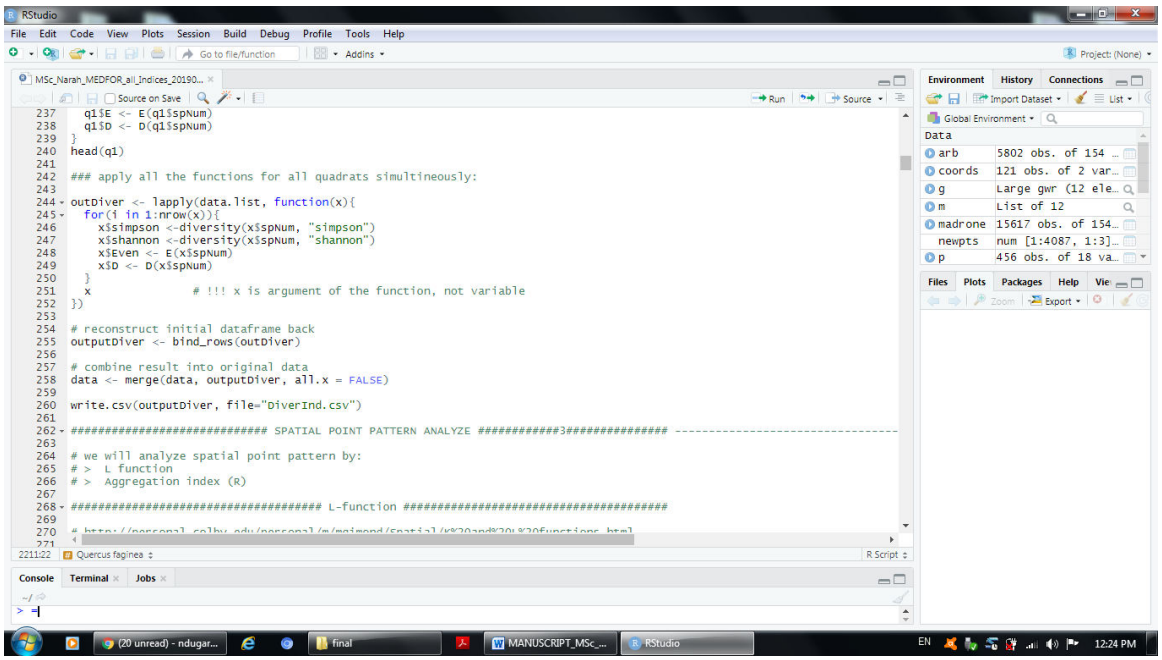
```
write.csv(data, file="Biomass.csv")

# create outputs df where all results will be saved

#################################### STAND CHARACTERISTICS ############################

### Basal area (Gi) calculation (per tree):

# Gi per tree: the cross-sectional area of a stem, usually measured at breast
# height (1.37 m). Formula per tree: Gi(m2) = (pi/4)*d(m)^2.
# if d is in cm, convert it into meter by multiplying 0.01.

data$Gi_m2 <- (pi/4)*(data$DBH_cm * 0.01)^2

### Basal area-Gi (m2/ha) by quadrant: sum of BA in one quadrant divided by area of the quadrant.
library(dplyr)
data <- data %>%
  group_by(quadrant)  %>%
  mutate(Gi_m2_ha = sum(Gi_m2)/(1/16))

# we have 1 hectare area of stand (equals to 100 x 100 m). It has been divided into
# 16 plots with equal size (25 x 25 m). To calc BA by ha, 1 ha area divided by 16, then
# we will get area of each quadrant in ha.

### proportion of  BA of each species by quadrant: Ex: total BA of pinusPinea is divided
# by total BA of each quadrant which has been calculated above (Gi_m2_ha).

# calc total BA of each species by quadrant
data <- data %>%
  group_by(quadrant, species) %>%
  mutate(Gi.sp = sum(Gi_m2))

# calc total BA of each quadrant
```



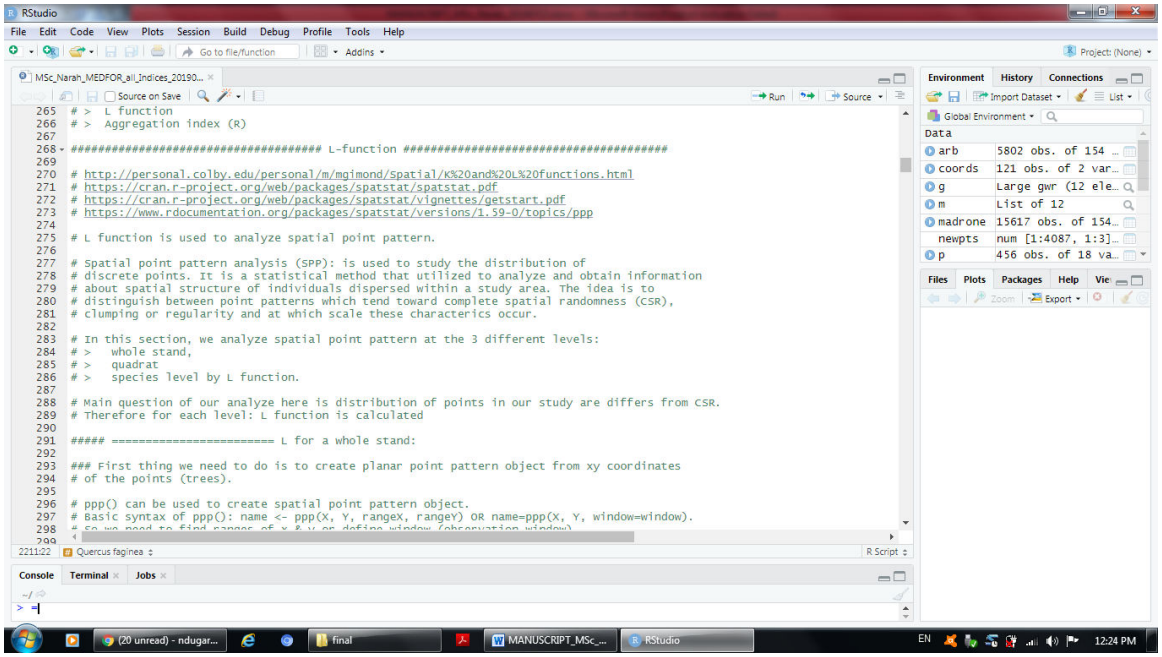
```
######################### BIODIVERSITY INDICES CALCULATIONS ############################

# https://www.r-bloggers.com/how-to-write-and-debug-an-r-function/
# https://www.statmethods.net/management/userfunctions.html
# https://www.rdocumentation.org/packages/base/versions/3.6.0/topics/lapply
# https://www.datacamp.com/community/tutorials/r-tutorial-apply-family
# https://www.r-bloggers.com/5-ways-to-subset-a-data-frame-in-r/

packages <-c("vegan", "dplyr")
lapply(packages, require, character.only=TRUE)

### first thing that we need to prepare our data

# convert name of species (column) in letter into numerical value because of vegan requirement.
# Species types must be in numeric type for biodiversity indices calculation using vegan package.

data$spNum <-as.numeric(data$species)
head(data)   # check result

# subset each quadrant separately
q1  <- data[data$quadrant==1,]
q2  <- data[data$quadrant==2,]
q3  <- data[data$quadrant==3,]
q4  <- data[data$quadrant==4,]
q5  <- data[data$quadrant==5,]
q6  <- data[data$quadrant==6,]
q7  <- data[data$quadrant==7,]
q8  <- data[data$quadrant==8,]
q9  <- data[data$quadrant==9,]
q10 <- data[data$quadrant==10,]
q11 <- data[data$quadrant==11,]
q12 <- data[data$quadrant==12,]
```

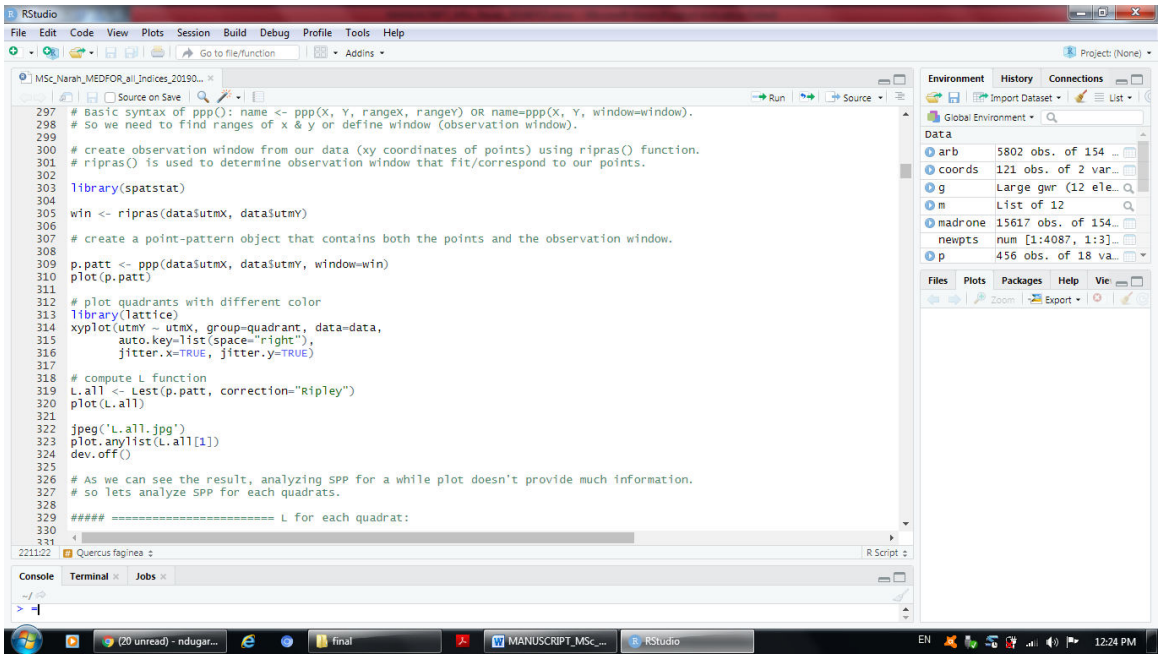


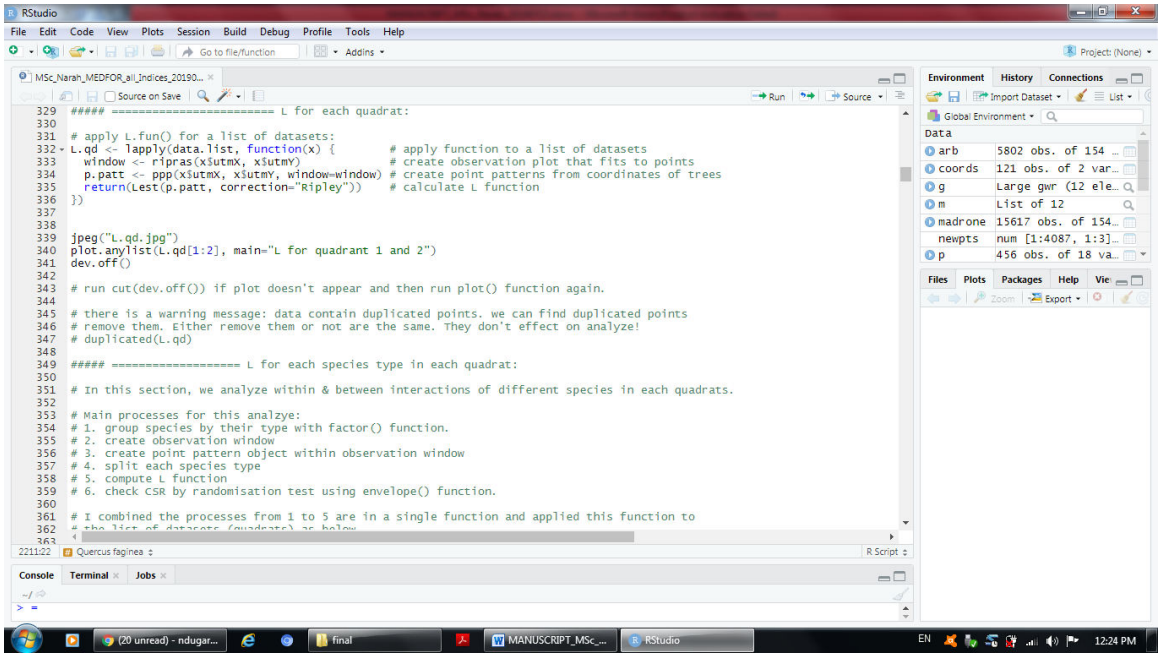
```
### define functions that are going to be applied for each quadrant:

# functions of some indices have already been developed in vegan package. Some are not.
# So lets write function for those that don't have pre-defined function.

# function for Evenness index:
E <-function(x){
  H <- diversity(x)       # total number of individials in dataset (shannon = H)
  S <- specnumber(x)      # rowSums does the same...
  E  <- H/log(S)          # Evennes index
  return(E)
}

# function for Berger-Parker index:
D <-function(x){
  N <- length(x)           # total number of individuals in dataset
  n <- table(x)            # number of individuals of each species
  Nmax <-max(n)            # max number of the most abundant species
  D <- Nmax/N              # calculate Berker parker index
  recip.D <- 1/D
  return(recip.D)
}

# check functions for quadrant 1 whether they are correct or not.
for(i in 1:nrow(q1)){
  q1$simp <-diversity(q1$spNum, "simpson")
  q1$shan <-diversity(q1$spNum, "shannon")
  q1$E <- E(q1$spNum)
  q1$D <- D(q1$spNum)
}
head(q1)

### apply all the functions for all quadrats simultineously:
```



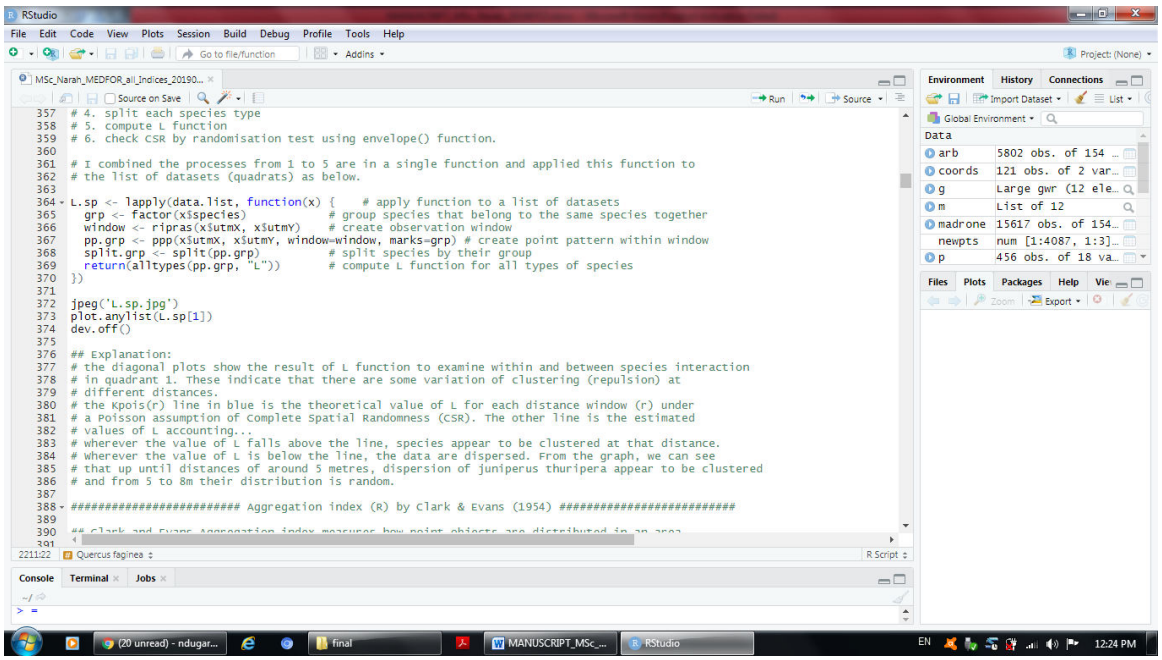
```
  q1$E <- E(q1$spNum)
  q1$D <- D(q1$spNum)
}
head(q1)

### apply all the functions for all quadrats simultineously:

outDiver <- lapply(data.list, function(x){
  for(i in 1:nrow(x)){
    x$simpson <-diversity(x$spNum, "simpson")
    x$shannon <-diversity(x$spNum, "shannon")
    x$Even <- E(x$spNum)
    x$D <- D(x$spNum)
  }
  x               # !!! x is argument of the function, not variable
})

# reconstruct initial dataframe back
outputDiver <- bind_rows(outDiver)

# combine result into original data
data <- merge(data, outputDiver, all.x = FALSE)

write.csv(outputDiver, file="DiverInd.csv")

############################## SPATIAL POINT PATTERN ANALYZE ############3############## --------------------------------

# we will analyze spatial point pattern by:
# >  L function
# >  Aggregation index (R)

################################### L-function ######################################
```


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
265 # > L function
266 # > Aggregation index (R)
268 ##### L-function #####
270 # http://personal.colby.edu/personal/m/mgimond/Spatial/K%20and%20L%20functions.html
271 # https://cran.r-project.org/web/packages/spatstat/spatstat.pdf
272 # https://cran.r-project.org/web/packages/spatstat/vignettes/getstart.pdf
273 # https://www.rdocumentation.org/packages/spatstat/versions/1.59-0/topics/ppp
275 # L function is used to analyze spatial point pattern.
277 # Spatial point pattern analysis (SPP): is used to study the distribution of
278 # discrete points. It is a statistical method that utilized to analyze and obtain information
279 # about spatial structure of individuals dispersed within a study area. The idea is to
280 # distinguish between point patterns which tend toward complete spatial randomness (CSR),
281 # clumping or regularity and at which scale these characterics occur.
283 # In this section, we analyze spatial point pattern at the 3 different levels:
284 # > whole stand,
285 # > quadrat
286 # > species level by L function.
288 # Main question of our analyze here is distribution of points in our study are differs from CSR.
289 # Therefore for each level: L function is calculated
291 ##### ========== L for a whole stand:
293 ### First thing we need to do is to create planar point pattern object from xy coordinates
294 # of the points (trees).
296 # ppp() can be used to create spatial point pattern object.
297 # Basic syntax of ppp(): name <- ppp(X, Y, rangeX, rangeY) OR name=ppp(X, Y, window=window).
298 # So we need to find ranges of x & y or define window (observation window)
211:22 Quercus faginea
R Script
Console Terminal Jobs
> =|
Environment History Connections
Import Dataset List
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
(20 unread) - ndugar...
final
MANUSCRIPT_MSc_...
RStudio
EN
12:24 PM

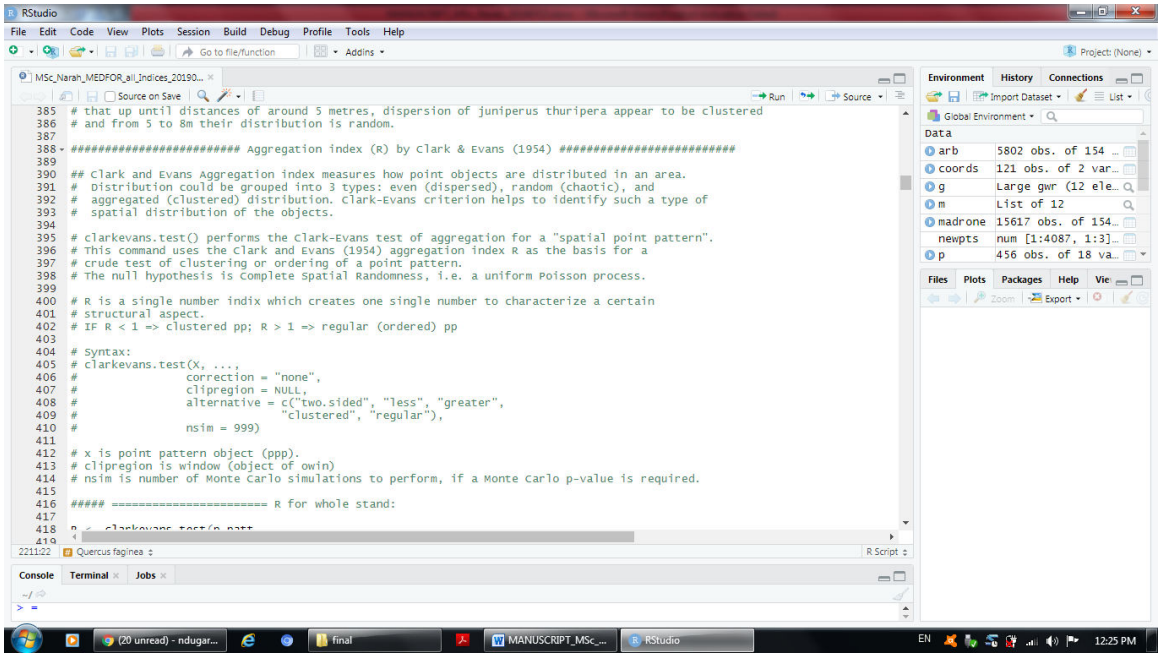

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
297 # Basic syntax of ppp(): name <- ppp(X, Y, rangeX, rangeY) OR name=ppp(X, Y, window=window).
298 # So we need to find ranges of x & y or define window (observation window).
300 # create observation window from our data (xy coordinates of points) using ripras() function.
301 # ripras() is used to determine observation window that fit/correspond to our points.
303 library(spatstat)
305 win <- ripras(data$utmX, data$utmY)
307 # create a point-pattern object that contains both the points and the observation window.
309 p.patt <- ppp(data$utmX, data$utmY, window=win)
310 plot(p.patt)
312 # plot quadrants with different color
313 library(lattice)
314 xyplot(utmY ~ utmX, group=quadrant, data=data,
315 auto.key=list(space="right"),
316 jitter.x=TRUE, jitter.y=TRUE)
318 # compute L function
319 L.all <- Lest(p.patt, correction="Ripley")
320 plot(L.all)
322 jpeg('L.all.jpg')
323 plot.anylist(L.all[1])
324 dev.off()
326 # As we can see the result, analyzing SPP for a while plot doesn't provide much information.
327 # so lets analyze SPP for each quadrats.
329 ##### ========== L for each quadrat:
211:22 Quercus faginea
R Script
Console Terminal Jobs
> =|
Environment History Connections
Import Dataset List
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
(20 unread) - ndugar...
final
MANUSCRIPT_MSc_...
RStudio
EN
12:24 PM

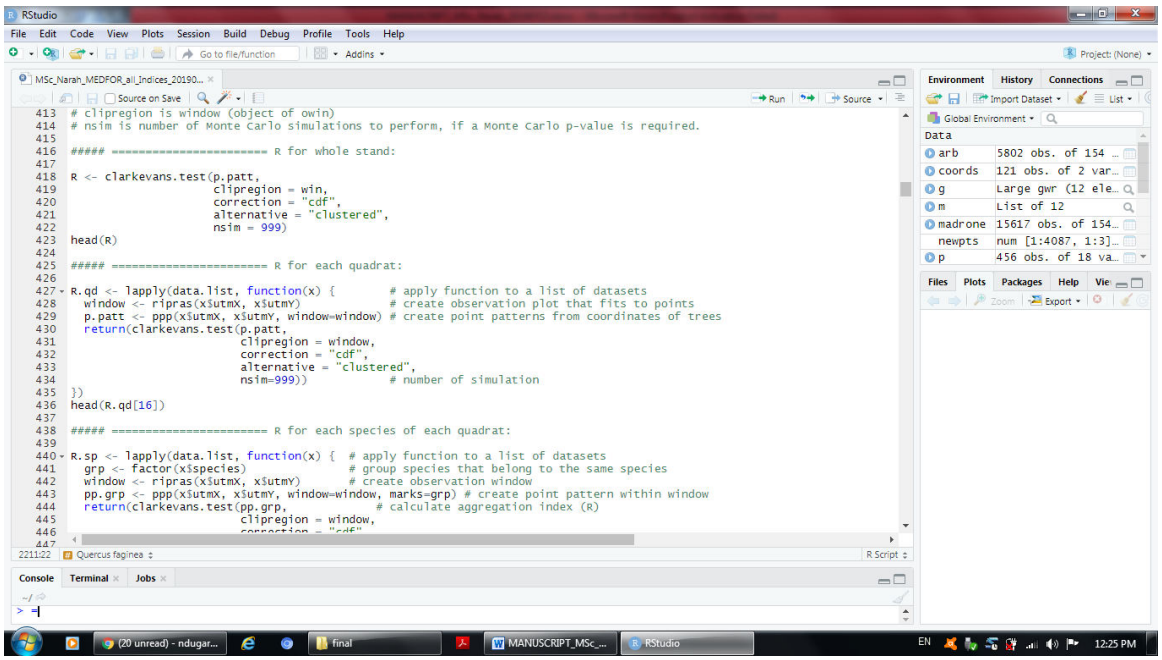


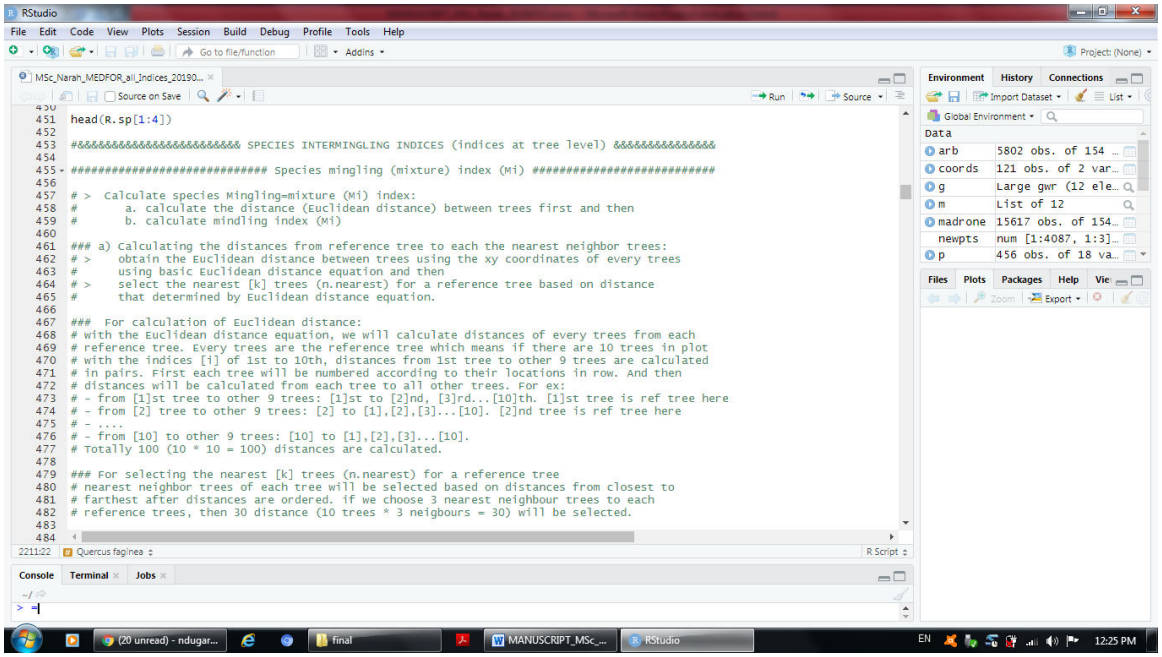
```
##### ========================= L for each quadrat:

# apply L.fun() for a list of datasets:
L.qd <- lapply(data.list, function(x) {        # apply function to a list of datasets
  window <- ripras(x$utmx, x$utmY)             # create observation plot that fits to points
  p.patt <- ppp(x$utmX, x$utmY, window=window) # create point patterns from coordinates of trees
  return(Lest(p.patt, correction="Ripley"))    # calculate L function
})


jpeg("L.qd.jpg")
plot.anylist(L.qd[1:2], main="L for quadrant 1 and 2")
dev.off()

# run cut(dev.off()) if plot doesn't appear and then run plot() function again.

# there is a warning message: data contain duplicated points. we can find duplicated points
# remove them. Either remove them or not are the same. They don't effect on analyze!
# duplicated(L.qd)

##### ==================== L for each species type in each quadrat:

# In this section, we analyze within & between interactions of different species in each quadrats.

# Main processes for this analzye:
# 1. group species by their type with factor() function.
# 2. create observation window
# 3. create point pattern object within observation window
# 4. split each species type
# 5. compute L function
# 6. check CSR by randomisation test using envelope() function.

# I combined the processes from 1 to 5 are in a single function and applied this function to
# the list of datasets (quadrats) as below.
```

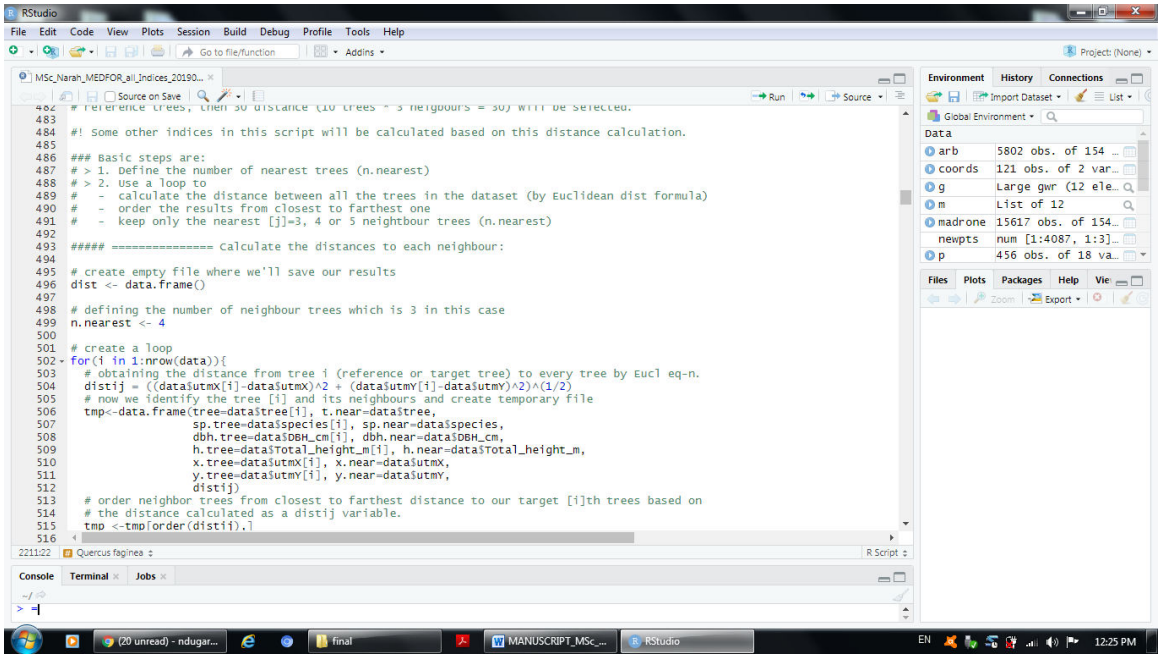


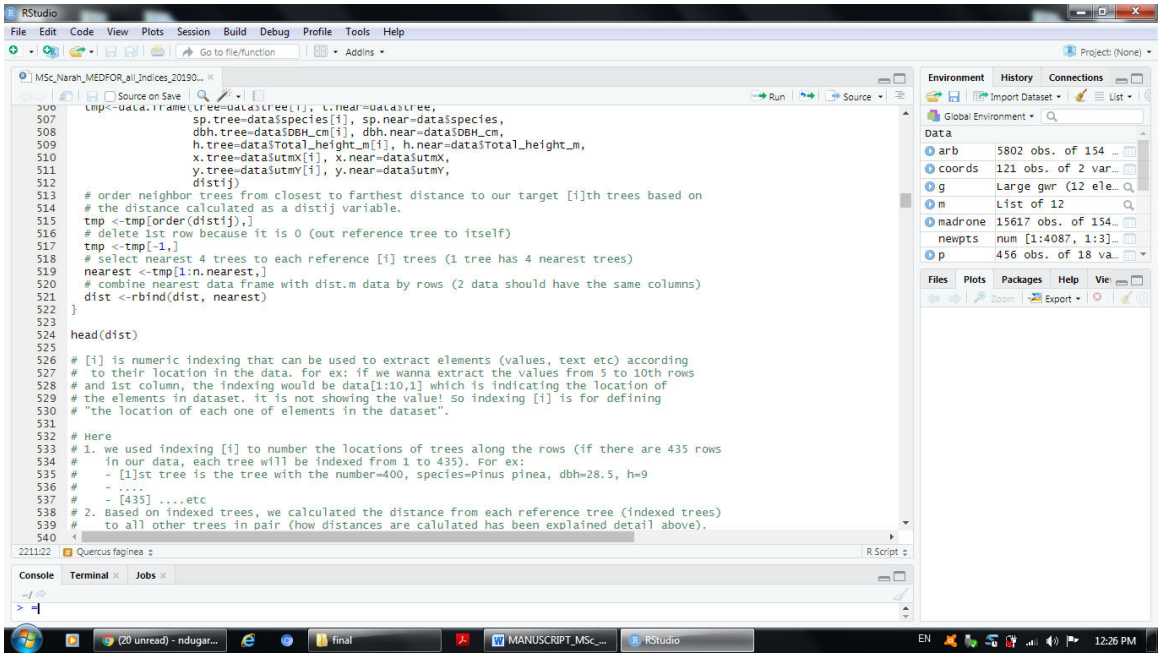
```
# 4. split each species type
# 5. compute L function
# 6. check CSR by randomisation test using envelope() function.

# I combined the processes from 1 to 5 are in a single function and applied this function to
# the list of datasets (quadrats) as below.

L.sp <- lapply(data.list, function(x) {    # apply function to a list of datasets
  grp <- factor(x$species)                 # group species that belong to the same species together
  window <- ripras(x$utmx, x$utmY)         # create observation window
  pp.grp <- ppp(x$utmx, x$utmY, window=window, marks=grp) # create point pattern within window
  split.grp <- split(pp.grp)               # split species by their group
  return(alltypes(pp.grp, "L"))            # compute L function for all types of species
})

jpeg('L.sp.jpg')
plot.anylist(L.sp[1])
dev.off()

## Explanation:
# the diagonal plots show the result of L function to examine within and between species interaction
# in quadrant 1. These indicate that there are some variation of clustering (repulsion) at
# different distances.
# the Kpois(r) line in blue is the theoretical value of L for each distance window (r) under
# a Poisson assumption of complete Spatial Randomness (CSR). The other line is the estimated
# values of L accounting...
# wherever the value of L falls above the line, species appear to be clustered at that distance.
# Wherever the value of L is below the line, the data are dispersed. From the graph, we can see
# that up until distances of around 5 metres, dispersion of juniperus thuripera appear to be clustered
# and from 5 to 8m their distribution is random.

######################### Aggregation index (R) by Clark & Evans (1954) ##########################
```

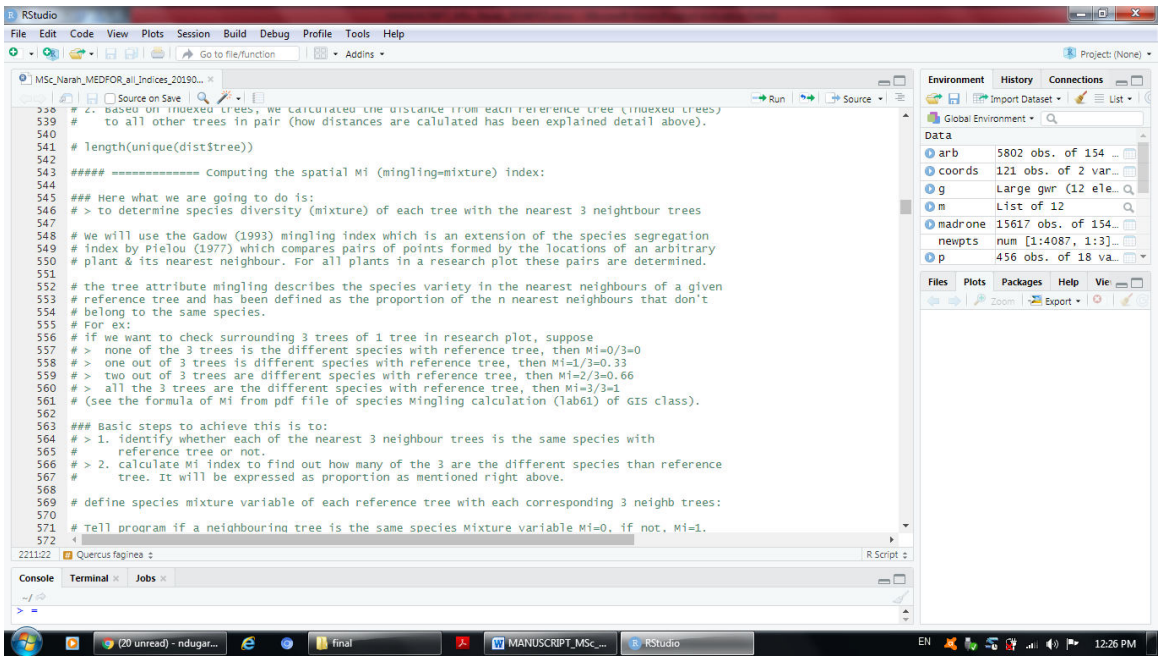

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
Source on Save
Run
Source
385 # that up until distances of around 5 metres, dispersion of juniperus thuripera appear to be clustered
386 # and from 5 to 8m their distribution is random.
387
388 ######################### Aggregation index (R) by Clark & Evans (1954) #########################
389
390 ## Clark and Evans Aggregation index measures how point objects are distributed in an area.
391 # Distribution could be grouped into 3 types: even (dispersed), random (chaotic), and
392 # aggregated (clustered) distribution. Clark-Evans criterion helps to identify such a type of
393 # spatial distribution of the objects.
394
395 # clarkevans.test() performs the Clark-Evans test of aggregation for a "spatial point pattern".
396 # This command uses the Clark and Evans (1954) aggregation index R as the basis for a
397 # crude test of clustering or ordering of a point pattern.
398 # The null hypothesis is Complete Spatial Randomness, i.e. a uniform Poisson process.
399
400 # R is a single number indix which creates one single number to characterize a certain
401 # structural aspect.
402 # IF R < 1 => clustered pp; R > 1 => regular (ordered) pp
403
404 # Syntax:
405 # clarkevans.test(X, ...,
406 # correction = "none",
407 # clipregion = NULL,
408 # alternative = c("two.sided", "less", "greater",
409 # "clustered", "regular"),
410 # nsim = 999)
411
412 # x is point pattern object (ppp).
413 # clipregion is window (object of owin)
414 # nsim is number of Monte Carlo simulations to perform, if a Monte Carlo p-value is required.
415
416 ##### ======================== R for whole stand:
417
418 R <- clarkevans.test(p.patt
419
211:22 Quercus faginea
R Script
Console Terminal Jobs
> =
Environment History Connections
Import Dataset List
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
(20 unread) - ndugar...
final
MANUSCRIPT_MSc_...
RStudio
EN
12:25 PM

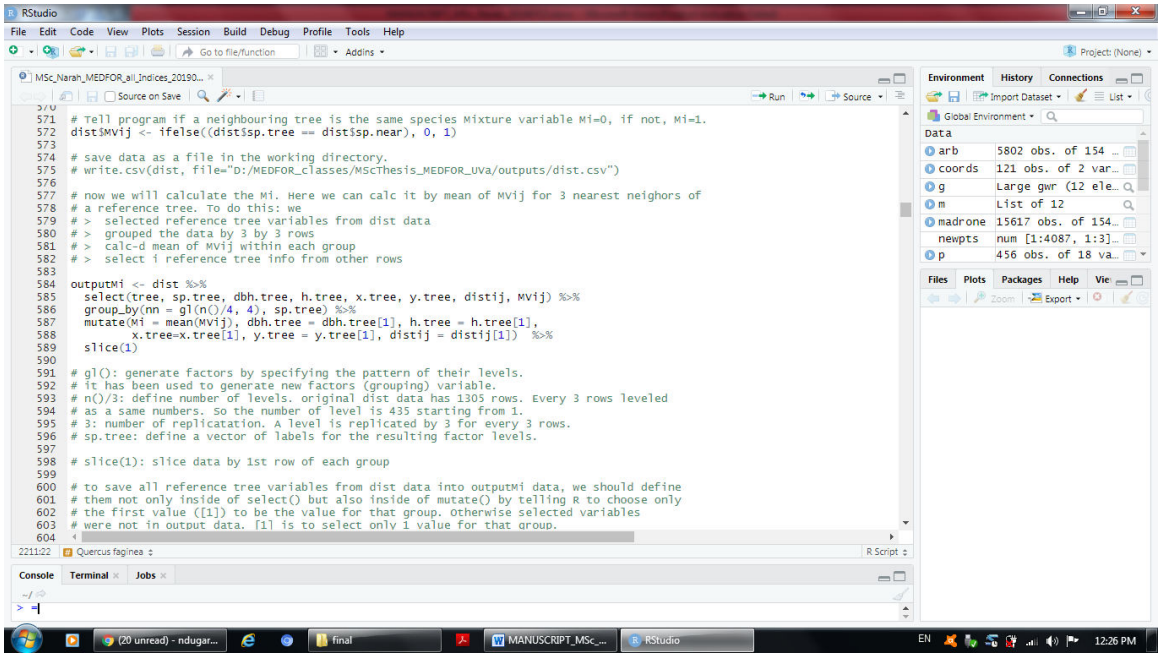

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
Source on Save
Run
Source
413 # clipregion is window (object of owin)
414 # nsim is number of Monte Carlo simulations to perform, if a Monte Carlo p-value is required.
415
416 ##### ======================== R for whole stand:
417
418 R <- clarkevans.test(p.patt,
419 clipregion = win,
420 correction = "cdf",
421 alternative = "clustered",
422 nsim = 999)
423 head(R)
424
425 ##### ======================== R for each quadrat:
426
427 R.qd <- lapply(data.list, function(x) { # apply function to a list of datasets
428 window <- ripras(x$utmX, x$utmY) # create observation plot that fits to points
429 p.patt <- ppp(x$utmX, x$utmY, window=window) # create point patterns from coordinates of trees
430 return(clarkevans.test(p.patt,
431 clipregion = window,
432 correction = "cdf",
433 alternative = "clustered",
434 nsim=999)) # number of simulation
435 })
436 head(R.qd[16])
437
438 ##### ======================== R for each species of each quadrat:
439
440 R.sp <- lapply(data.list, function(x) { # apply function to a list of datasets
441 grp <- factor(x$species) # group species that belong to the same species
442 window <- ripras(x$utmX, x$utmY) # create observation window
443 pp.grp <- ppp(x$utmX, x$utmY, window=window, marks=grp) # create point pattern within window
444 return(clarkevans.test(pp.grp, # calculate aggregation index (R)
445 clipregion = window,
446
447
211:22 Quercus faginea
R Script
Console Terminal Jobs
> =
Environment History Connections
Import Dataset List
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
(20 unread) - ndugar...
final
MANUSCRIPT_MSc_...
RStudio
EN
12:25 PM

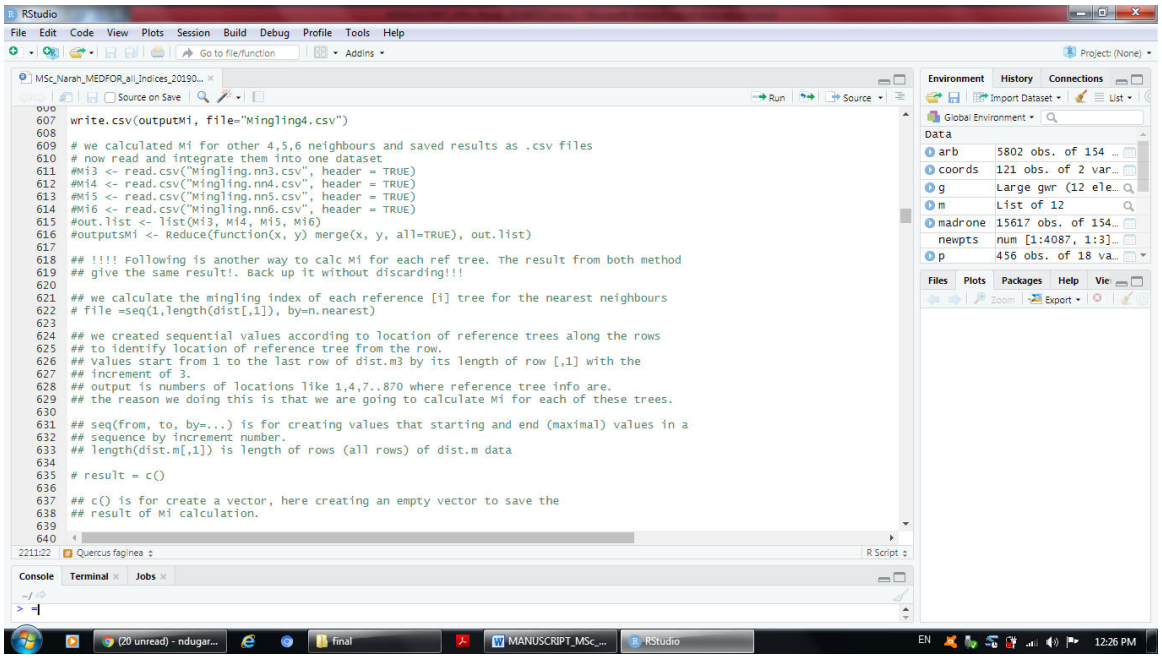

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
head(R.sp[1:4])
########################## SPECIES INTERMINGLING INDICES (indices at tree level) ##################
############################ Species mingling (mixture) index (Mi) ##########################
> Calculate species Mingling=mixture (Mi) index:
a. calculate the distance (Euclidean distance) between trees first and then
b. calculate mindling index (Mi)
a) Calculating the distances from reference tree to each the nearest neighbor trees:
> obtain the Euclidean distance between trees using the xy coordinates of every trees
using basic Euclidean distance equation and then
> select the nearest [k] trees (n.nearest) for a reference tree based on distance
that determined by Euclidean distance equation.
For calculation of Euclidean distance:
with the Euclidean distance equation, we will calculate distances of every trees from each
reference tree. Every trees are the reference tree which means if there are 10 trees in plot
with the indices [i] of 1st to 10th, distances from 1st tree to other 9 trees are calculated
in pairs. First each tree will be numbered according to their locations in row. And then
distances will be calculated from each tree to all other trees. For ex:
- from [1]st tree to other 9 trees: [1]st to [2]nd, [3]rd...[10]th. [1]st tree is ref tree here
- from [2] tree to other 9 trees: [2] to [1],[2],[3]...[10]. [2]nd tree is ref tree here
- ....
- from [10] to other 9 trees: [10] to [1],[2],[3]...[10].
Totally 100 (10 * 10 = 100) distances are calculated.
For selecting the nearest [k] trees (n.nearest) for a reference tree
nearest neighbor trees of each tree will be selected based on distances from closest to
farthest after distances are ordered. if we choose 3 nearest neighbour trees to each
reference trees, then 30 distance (10 trees * 3 neigbours = 30) will be selected.
Environment History Connections
Import Dataset
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
Console Terminal Jobs
Quercus faginea
R Script
(20 unread) - ndugar...
final
MANUSCRIPT_MSc_...
RStudio
EN
12:25 PM

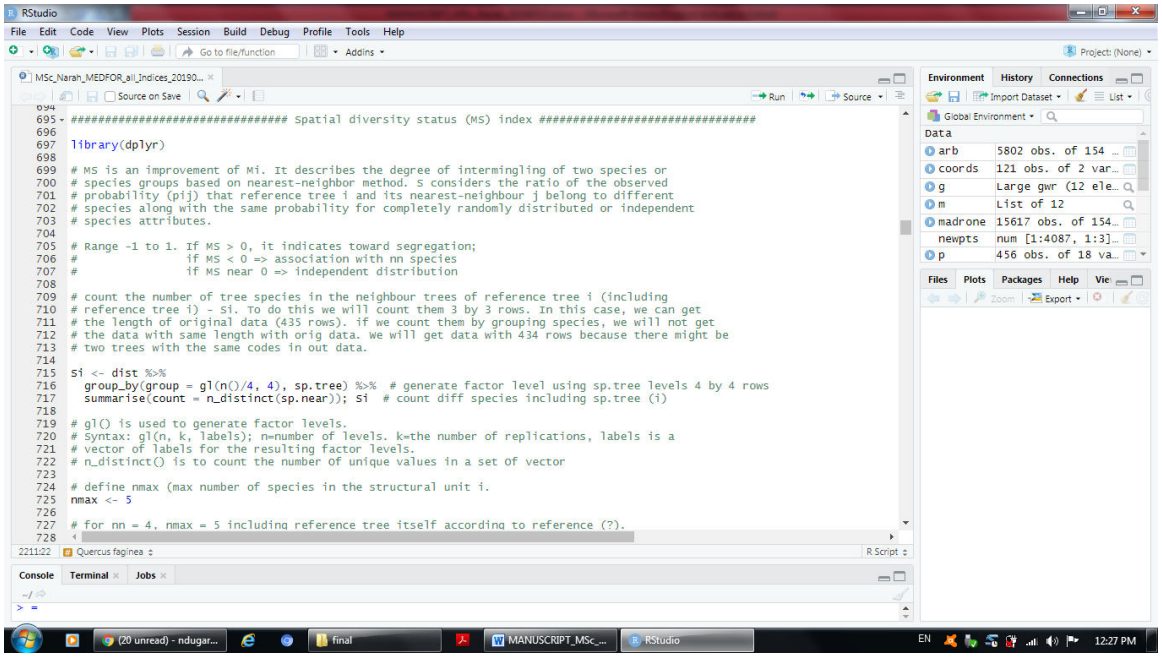

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
#! Some other indices in this script will be calculated based on this distance calculation.
Basic steps are:
> 1. Define the number of nearest trees (n.nearest)
> 2. Use a loop to
- calculate the distance between all the trees in the dataset (by Euclidean dist formula)
- order the results from closest to farthest one
- keep only the nearest [j]=3, 4 or 5 neightbour trees (n.nearest)
================ Calculate the distances to each neighbour:
# create empty file where we'll save our results
dist <- data.frame()
# defining the number of neighbour trees which is 3 in this case
n.nearest <- 4
# create a loop
for(i in 1:nrow(data)){
# obtaining the distance from tree i (reference or target tree) to every tree by Eucl eq-n.
distij = ((data$utmX[i]-data$utmX)^2 + (data$utmY[i]-data$utmY)^2)^(1/2)
# now we identify the tree [i] and its neighbours and create temporary file
tmp<-data.frame(tree=data$tree[i], t.near=data$tree,
sp.tree=data$species[i], sp.near=data$species,
dbh.tree=data$DBH_cm[i], dbh.near=data$DBH_cm,
h.tree=data$Total_height_m[i], h.near=data$Total_height_m,
x.tree=data$utmX[i], x.near=data$utmX,
y.tree=data$utmY[i], y.near=data$utmY,
distij)
# order neighbor trees from closest to farthest distance to our target [i]th trees based on
# the distance calculated as a distij variable.
tmp <-tmp[order(distij),]
Environment History Connections
Import Dataset
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
Console Terminal Jobs
Quercus faginea
R Script
(20 unread) - ndugar...
final
MANUSCRIPT_MSc_...
RStudio
EN
12:25 PM

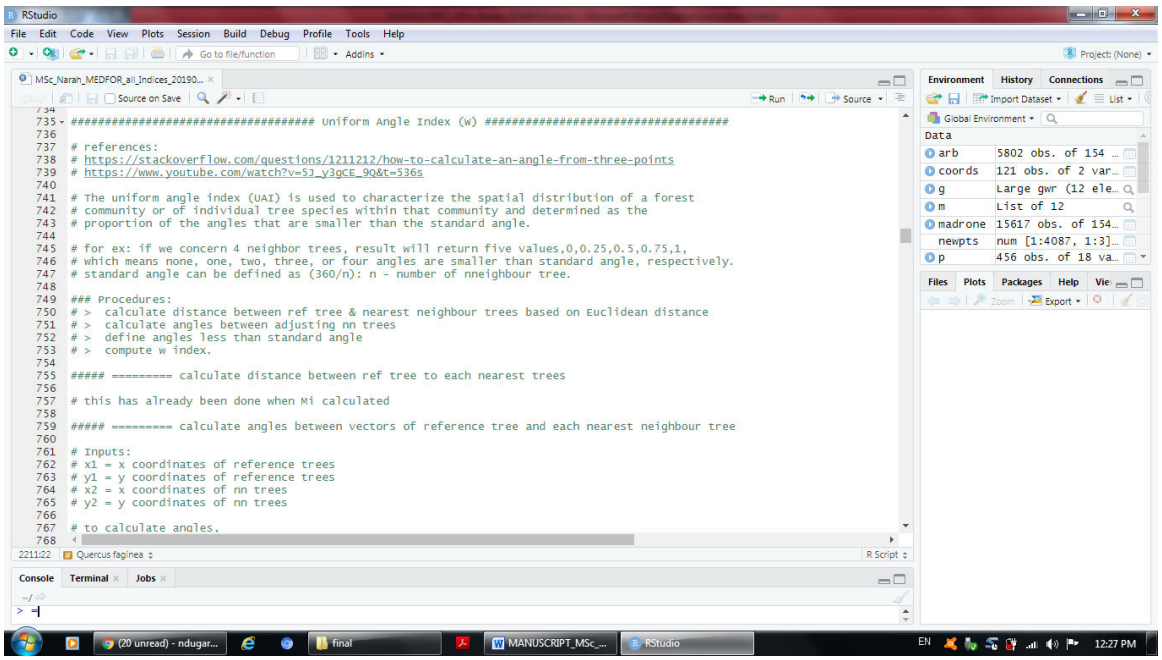

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
Source on Save
Run Source
Environment History Connections
Import Dataset List
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
tmp<-data.frame(tree=data$tree[i], t.near=data$tree,
sp.tree=data$species[i], sp.near=data$species,
dbh.tree=data$DBH_cm[i], dbh.near=data$DBH_cm,
h.tree=data$Total_height_m[i], h.near=data$Total_height_m,
x.tree=data$utmX[i], x.near=data$utmX,
y.tree=data$utmY[i], y.near=data$utmY,
distij)
# order neighbor trees from closest to farthest distance to our target [i]th trees based on
# the distance calculated as a distij variable.
tmp <-tmp[order(distij),]
# delete 1st row because it is 0 (out reference tree to itself)
tmp <-tmp[-1,]
# select nearest 4 trees to each reference [i] trees (1 tree has 4 nearest trees)
nearest <-tmp[1:n.nearest,]
# combine nearest data frame with dist.m data by rows (2 data should have the same columns)
dist <-rbind(dist, nearest)
}
head(dist)
# [i] is numeric indexing that can be used to extract elements (values, text etc) according
# to their location in the data. for ex: if we wanna extract the values from 5 to 10th rows
# and 1st column, the indexing would be data[1:10,1] which is indicating the location of
# the elements in dataset. it is not showing the value! So indexing [i] is for defining
# "the location of each one of elements in the dataset".
# Here
# 1. we used indexing [i] to number the locations of trees along the rows (if there are 435 rows
# in our data, each tree will be indexed from 1 to 435). For ex:
# - [1]st tree is the tree with the number=400, species=Pinus pinea, dbh=28.5, h=9
# - ....
# - [435] ....etc
# 2. Based on indexed trees, we calculated the distance from each reference tree (indexed trees)
# to all other trees in pair (how distances are calulated has been explained detail above).
2211:22 Quercus faginea
R Script
Console Terminal Jobs
> =
(20 unread) - ndugar...
final
MANUSCRIPT_MSc_...
RStudio
EN
12:26 PM

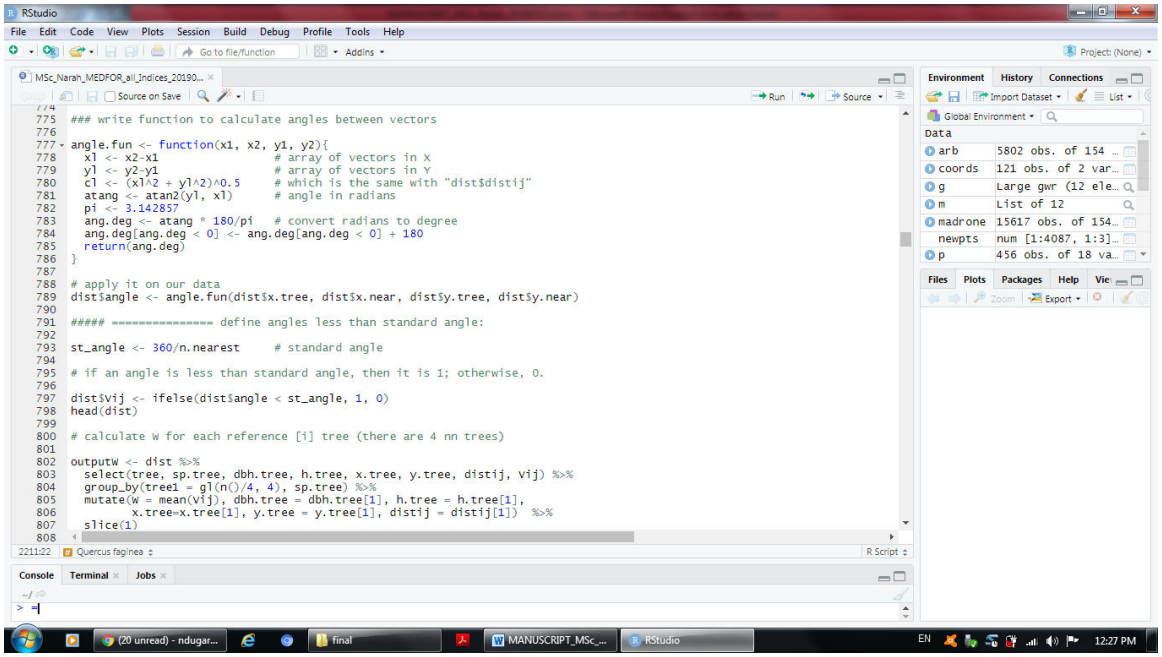

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
Source on Save
Run Source
Environment History Connections
Import Dataset List
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
2. Based on indexed trees, we calculated the distance from each reference tree (indexed trees)
to all other trees in pair (how distances are calulated has been explained detail above).
length(unique(dist$tree))
============= Computing the spatial Mi (mingling=mixture) index:
Here what we are going to do is:
> to determine species diversity (mixture) of each tree with the nearest 3 neightbour trees
We will use the Gadow (1993) mingling index which is an extension of the species segregation
index by Pielou (1977) which compares pairs of points formed by the locations of an arbitrary
plant & its nearest neighbour. For all plants in a research plot these pairs are determined.
the tree attribute mingling describes the species variety in the nearest neighbours of a given
reference tree and has been defined as the proportion of the n nearest neighbours that don't
belong to the same species.
For ex:
if we want to check surrounding 3 trees of 1 tree in research plot, suppose
> none of the 3 trees is the different species with reference tree, then Mi=0/3=0
> one out of 3 trees is different species with reference tree, then Mi=1/3=0.33
> two out of 3 trees are different species with reference tree, then Mi=2/3=0.66
> all the 3 trees are the different species with reference tree, then Mi=3/3=1
(see the formula of Mi from pdf file of species Mingling calculation (lab61) of GIS class).
Basic steps to achieve this is to:
> 1. identify whether each of the nearest 3 neighbour trees is the same species with
reference tree or not.
> 2. calculate Mi index to find out how many of the 3 are the different species than reference
tree. It will be expressed as proportion as mentioned right above.
define species mixture variable of each reference tree with each corresponding 3 neighb trees:
Tell program if a neighbouring tree is the same species Mixture variable Mi=0, if not, Mi=1.
2211:22 Quercus faginea
R Script
Console Terminal Jobs
> =
(20 unread) - ndugar...
final
MANUSCRIPT_MSc_...
RStudio
EN
12:26 PM


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
Source on Save
Run
Source
Environment History Connections
Import Dataset
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
Tell program if a neighbouring tree is the same species Mixture variable Mi=0, if not, Mi=1.
dist$MVij <- ifelse((dist$sp.tree == dist$sp.near), 0, 1)
# save data as a file in the working directory.
# write.csv(dist, file="D:/MEDFOR_classes/MScThesis_MEDFOR_UVa/outputs/dist.csv")
# now we will calculate the Mi. Here we can calc it by mean of MVij for 3 nearest neighors of
# a reference tree. To do this: we
# > selected reference tree variables from dist data
# > grouped the data by 3 by 3 rows
# > calc-d mean of MVij within each group
# > select i reference tree info from other rows
outputMi <- dist %>%
select(tree, sp.tree, dbh.tree, h.tree, x.tree, y.tree, distij, MVij) %>%
group_by(nn = gl(n()/4, 4), sp.tree) %>%
mutate(Mi = mean(MVij), dbh.tree = dbh.tree[1], h.tree = h.tree[1],
x.tree=x.tree[1], y.tree = y.tree[1], distij = distij[1]) %>%
slice(1)
# gl(): generate factors by specifying the pattern of their levels.
# it has been used to generate new factors (grouping) variable.
# n()/3: define number of levels. original dist data has 1305 rows. Every 3 rows leveled
# as a same numbers. So the number of level is 435 starting from 1.
# 3: number of replicatation. A level is replicated by 3 for every 3 rows.
# sp.tree: define a vector of labels for the resulting factor levels.
# slice(1): slice data by 1st row of each group
# to save all reference tree variables from dist data into outputMi data, we should define
# them not only inside of select() but also inside of mutate() by telling R to choose only
# the first value ([1]) to be the value for that group. Otherwise selected variables
# were not in output data. [1] is to select only 1 value for that group.
2211:22 Quercus faginea
R Script
Console Terminal Jobs
> =|
(20 unread) - ndugar...
final
MANUSCRIPT_MSc...
RStudio
EN
12:26 PM

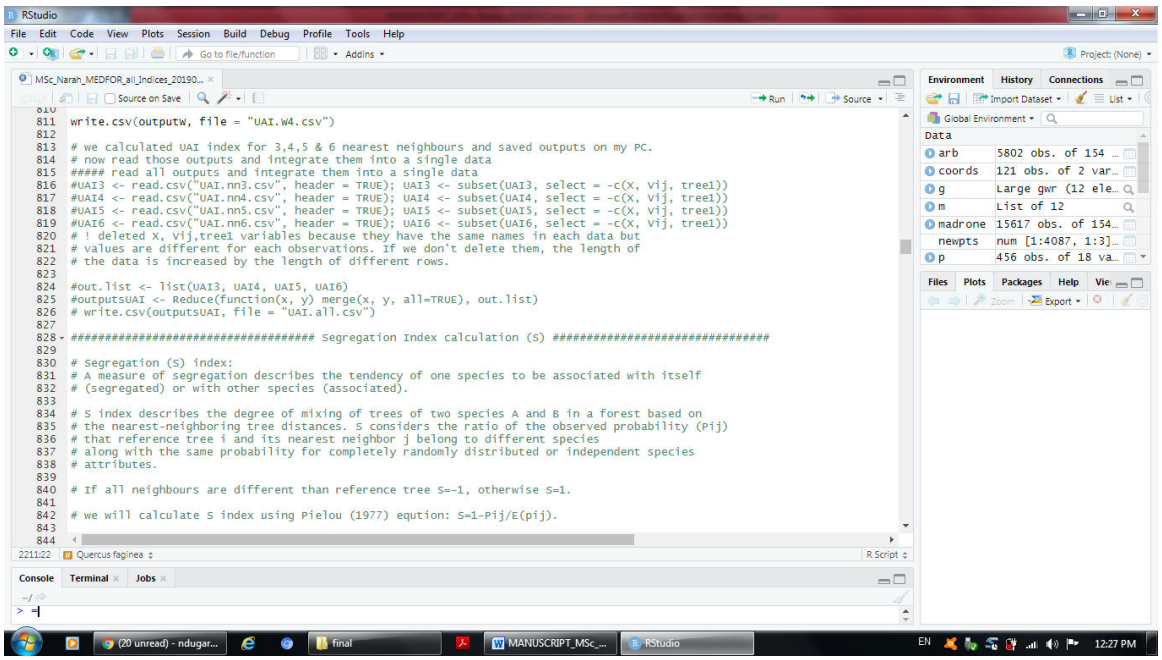

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
Source on Save
Run
Source
Environment History Connections
Import Dataset
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
write.csv(outputMi, file="Mingling4.csv")
# we calculated Mi for other 4,5,6 neighbours and saved results as .csv files
# now read and integrate them into one dataset
#Mi3 <- read.csv("Mingling.nn3.csv", header = TRUE)
#Mi4 <- read.csv("Mingling.nn4.csv", header = TRUE)
#Mi5 <- read.csv("Mingling.nn5.csv", header = TRUE)
#Mi6 <- read.csv("Mingling.nn6.csv", header = TRUE)
#out.list <- list(Mi3, Mi4, Mi5, Mi6)
#outputsMi <- Reduce(function(x, y) merge(x, y, all=TRUE), out.list)
!!!! Following is another way to calc Mi for each ref tree. The result from both method
give the same result!. Back up it without discarding!!!
we calculate the mingling index of each reference [i] tree for the nearest neighbours
file =seq(1,length(dist[,1]), by=n.nearest)
we created sequential values according to location of reference trees along the rows
to identify location of reference tree from the row.
Values start from 1 to the last row of dist.m3 by its length of row [,1] with the
increment of 3.
output is numbers of locations like 1,4,7..870 where reference tree info are.
the reason we doing this is that we are going to calculate Mi for each of these trees.
seq(from, to, by=....) is for creating values that starting and end (maximal) values in a
sequence by increment number.
length(dist.m[,1]) is length of rows (all rows) of dist.m data
result = c()
c() is for create a vector, here creating an empty vector to save the
result of Mi calculation.
2211:22 Quercus faginea
R Script
Console Terminal Jobs
> =|
(20 unread) - ndugar...
final
MANUSCRIPT_MSc...
RStudio
EN
12:26 PM


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
Source on Save
Run Source
Environment History Connections
Import Dataset List
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
695 ################################ Spatial diversity status (MS) index ################################
697 library(dplyr)
699 # MS is an improvement of Mi. It describes the degree of intermingling of two species or
700 # species groups based on nearest-neighbor method. s considers the ratio of the observed
701 # probability (pij) that reference tree i and its nearest-neighbour j belong to different
702 # species along with the same probability for completely randomly distributed or independent
703 # species attributes.
705 # Range -1 to 1. If MS > 0, it indicates toward segregation;
706 # If MS < 0 => association with nn species
707 # If MS near 0 => independent distribution
709 # count the number of tree species in the neighbour trees of reference tree i (including
710 # reference tree i) - Si. To do this we will count them 3 by 3 rows. In this case, we can get
711 # the length of original data (435 rows). if we count them by grouping species, we will not get
712 # the data with same length with orig data. We will get data with 434 rows because there might be
713 # two trees with the same codes in out data.
715 Si <- dist %>%
716 group_by(group = gl(n()/4, 4), sp.tree) %>% # generate factor level using sp.tree levels 4 by 4 rows
717 summarise(count = n_distinct(sp.near)); Si # count diff species including sp.tree (i)
719 # gl() is used to generate factor levels.
720 # Syntax: gl(n, k, labels); n=number of levels. k=the number of replications, labels is a
721 # vector of labels for the resulting factor levels.
722 # n_distinct() is to count the number of unique values in a set of vector
724 # define nmax (max number of species in the structural unit i.
725 nmax <- 5
727 # for nn = 4, nmax = 5 including reference tree itself according to reference (?).
2211:22 Quercus faginea
R Script
Console Terminal Jobs
> =
(20 unread) - ndugar... final MANUSCRIPT_MSc_... RStudio
EN 12:27 PM

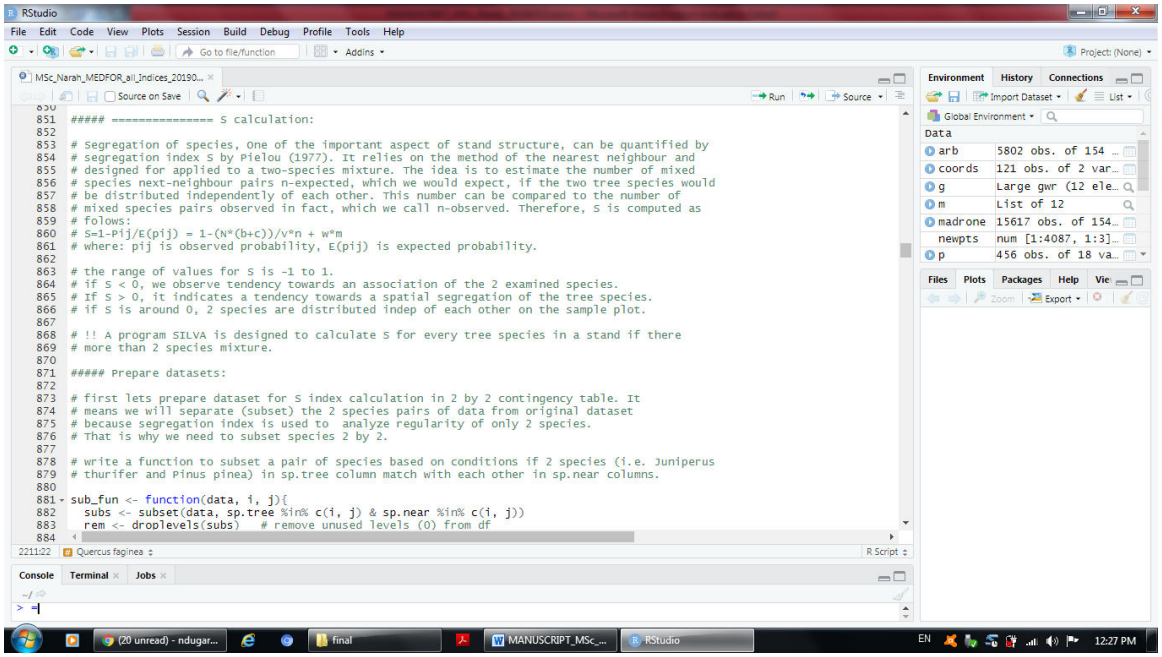

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
Source on Save
Run Source
Environment History Connections
Import Dataset List
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
735 ################################ Uniform Angle Index (W) ################################
737 # references:
738 # https://stackoverflow.com/questions/1211212/how-to-calculate-an-angle-from-three-points
739 # https://www.youtube.com/watch?v=5J_y3gCE_9Q&t=536s
741 # The uniform angle index (UAI) is used to characterize the spatial distribution of a forest
742 # community or of individual tree species within that community and determined as the
743 # proportion of the angles that are smaller than the standard angle.
745 # for ex: if we concern 4 neighbor trees, result will return five values,0,0.25,0.5,0.75,1,
746 # which means none, one, two, three, or four angles are smaller than standard angle, respectively.
747 # standard angle can be defined as (360/n): n - number of nneighbour tree.
749 ### Procedures:
750 # > calculate distance between ref tree & nearest neighbour trees based on Euclidean distance
751 # > calculate angles between adjusting nn trees
752 # > define angles less than standard angle
753 # > compute w index.
755 ##### ========= calculate distance between ref tree to each nearest trees
757 # this has already been done when Mi calculated
759 ##### ========= calculate angles between vectors of reference tree and each nearest neighbour tree
761 # Inputs:
762 # x1 = x coordinates of reference trees
763 # y1 = y coordinates of reference trees
764 # x2 = x coordinates of nn trees
765 # y2 = y coordinates of nn trees
767 # to calculate angles,
2211:22 Quercus faginea
R Script
Console Terminal Jobs
> =
(20 unread) - ndugar... final MANUSCRIPT_MSc_... RStudio
EN 12:27 PM



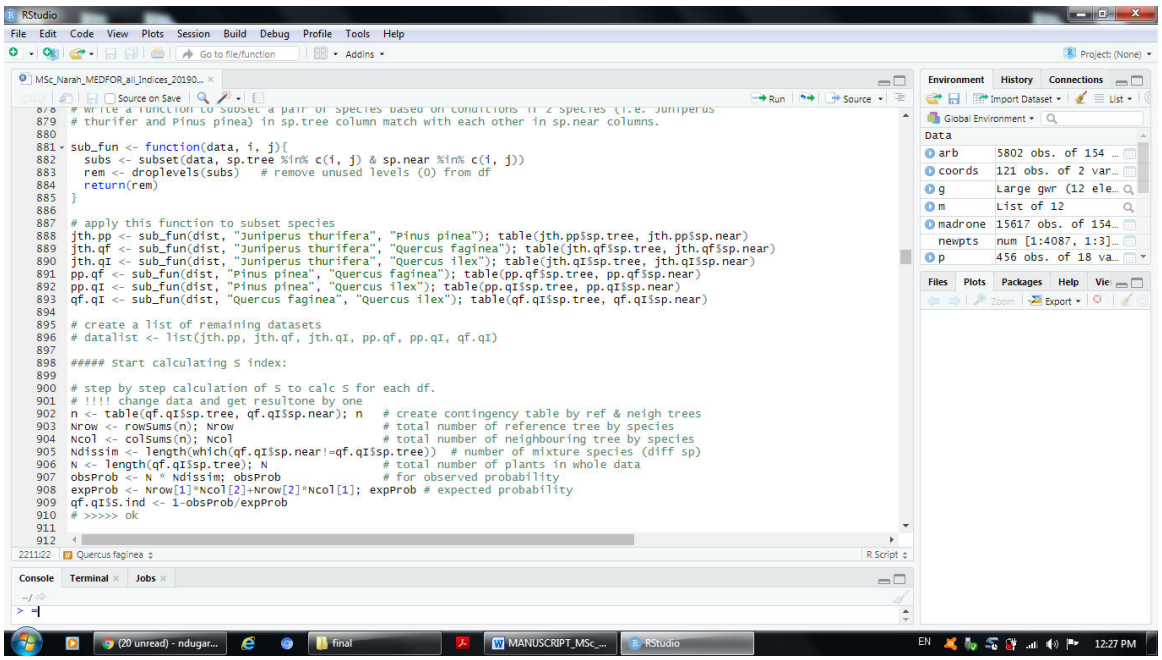
```
### write function to calculate angles between vectors

angle.fun <- function(x1, x2, y1, y2){
  xl <- x2-x1                 # array of vectors in X
  yl <- y2-y1                 # array of vectors in Y
  cl <- (xl^2 + yl^2)^0.5     # which is the same with "dist$distij"
  atang <- atan2(yl, xl)      # angle in radians
  pi <- 3.142857
  ang.deg <- atang * 180/pi   # convert radians to degree
  ang.deg[ang.deg < 0] <- ang.deg[ang.deg < 0] + 180
  return(ang.deg)
}

# apply it on our data
dist$angle <- angle.fun(dist$x.tree, dist$x.near, dist$y.tree, dist$y.near)

##### ================ define angles less than standard angle:

st_angle <- 360/n.nearest     # standard angle

# if an angle is less than standard angle, then it is 1; otherwise, 0.

dist$vij <- ifelse(dist$angle < st_angle, 1, 0)
head(dist)

# calculate W for each reference [i] tree (there are 4 nn trees)

outputW <- dist %>%
  select(tree, sp.tree, dbh.tree, h.tree, x.tree, y.tree, distij, vij) %>%
  group_by(tree1 = gl(n()/4, 4), sp.tree) %>%
  mutate(W = mean(vij), dbh.tree = dbh.tree[1], h.tree = h.tree[1],
         x.tree=x.tree[1], y.tree = y.tree[1], distij = distij[1])  %>%
  slice(1)
```



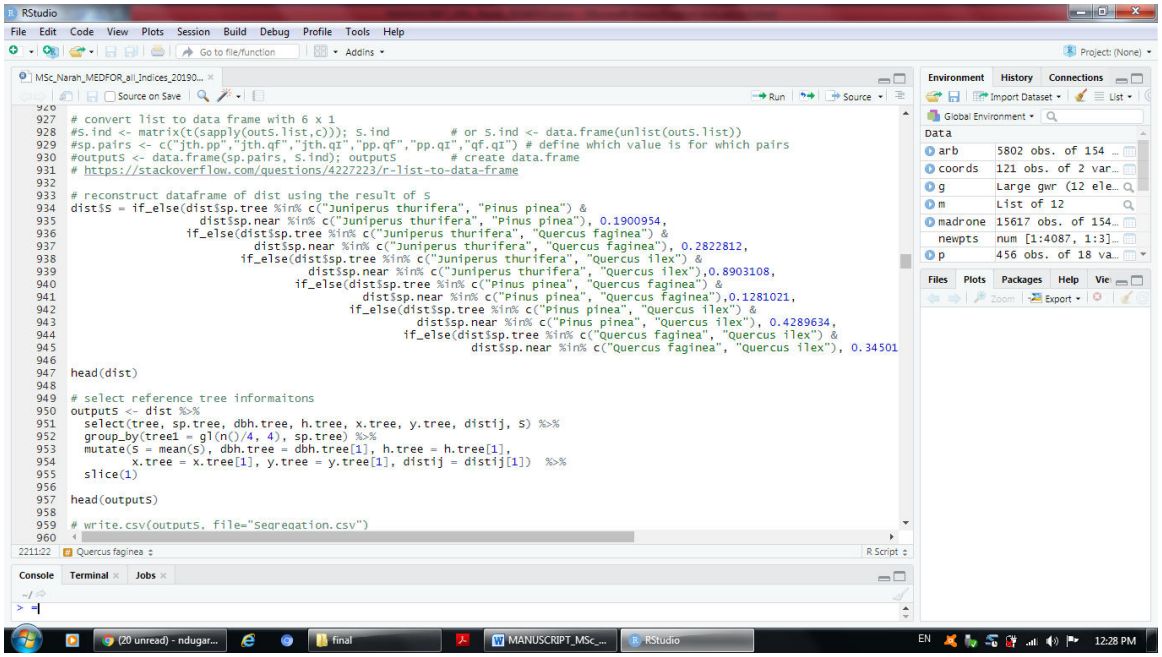
```
write.csv(outputW, file = "UAI.W4.csv")

# we calculated UAI index for 3,4,5 & 6 nearest neighbours and saved outputs on my PC.
# now read those outputs and integrate them into a single data
##### read all outputs and integrate them into a single data
#UAI3 <- read.csv("UAI.nn3.csv", header = TRUE); UAI3 <- subset(UAI3, select = -c(X, vij, tree1))
#UAI4 <- read.csv("UAI.nn4.csv", header = TRUE); UAI4 <- subset(UAI4, select = -c(X, vij, tree1))
#UAI5 <- read.csv("UAI.nn5.csv", header = TRUE); UAI5 <- subset(UAI5, select = -c(X, vij, tree1))
#UAI6 <- read.csv("UAI.nn6.csv", header = TRUE); UAI6 <- subset(UAI6, select = -c(X, vij, tree1))
# I deleted X, vij,tree1 variables because they have the same names in each data but
# values are different for each observations. If we don't delete them, the length of
# the data is increased by the length of different rows.

#out.list <- list(UAI3, UAI4, UAI5, UAI6)
#outputsUAI <- Reduce(function(x, y) merge(x, y, all=TRUE), out.list)
# write.csv(outputsUAI, file = "UAI.all.csv")

################################### Segregation Index calculation (S) ##############################

# Segregation (S) index:
# A measure of segregation describes the tendency of one species to be associated with itself
# (segregated) or with other species (associated).

# S index describes the degree of mixing of trees of two species A and B in a forest based on
# the nearest-neighboring tree distances. S considers the ratio of the observed probability (Pij)
# that reference tree i and its nearest neighbor j belong to different species
# along with the same probability for completely randomly distributed or independent species
# attributes.

# If all neighbours are different than reference tree S=-1, otherwise S=1.

# we will calculate S index using Pielou (1977) eqution: S=1-Pij/E(pij).
```



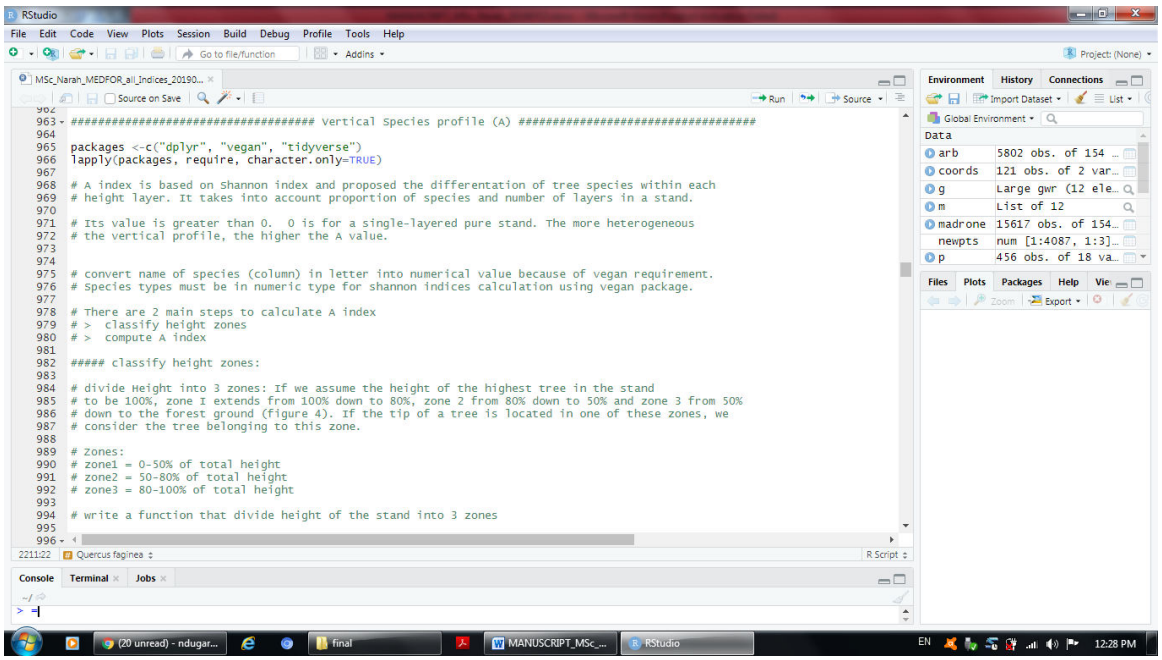
```
##### ============== S calculation:

# Segregation of species, One of the important aspect of stand structure, can be quantified by
# segregation index S by Pielou (1977). It relies on the method of the nearest neighbour and
# designed for applied to a two-species mixture. The idea is to estimate the number of mixed
# species next-neighbour pairs n-expected, which we would expect, if the two tree species would
# be distributed independently of each other. This number can be compared to the number of
# mixed species pairs observed in fact, which we call n-observed. Therefore, S is computed as
# folows:
# S=1-Pij/E(pij) = 1-(N*(b+c))/v*n + w*m
# where: pij is observed probability, E(pij) is expected probability.

# the range of values for S is -1 to 1.
# if S < 0, we observe tendency towards an association of the 2 examined species.
# If S > 0, it indicates a tendency towards a spatial segregation of the tree species.
# if S is around 0, 2 species are distributed indep of each other on the sample plot.

# !! A program SILVA is designed to calculate S for every tree species in a stand if there
# more than 2 species mixture.

##### Prepare datasets:

# first lets prepare dataset for S index calculation in 2 by 2 contingency table. It
# means we will separate (subset) the 2 species pairs of data from original dataset
# because segregation index is used to  analyze regularity of only 2 species.
# That is why we need to subset species 2 by 2.

# write a function to subset a pair of species based on conditions if 2 species (i.e. Juniperus
# thurifer and Pinus pinea) in sp.tree column match with each other in sp.near columns.

sub_fun <- function(data, i, j){
  subs <- subset(data, sp.tree %in% c(i, j) & sp.near %in% c(i, j))
  rem <- droplevels(subs)   # remove unused levels (0) from df
```



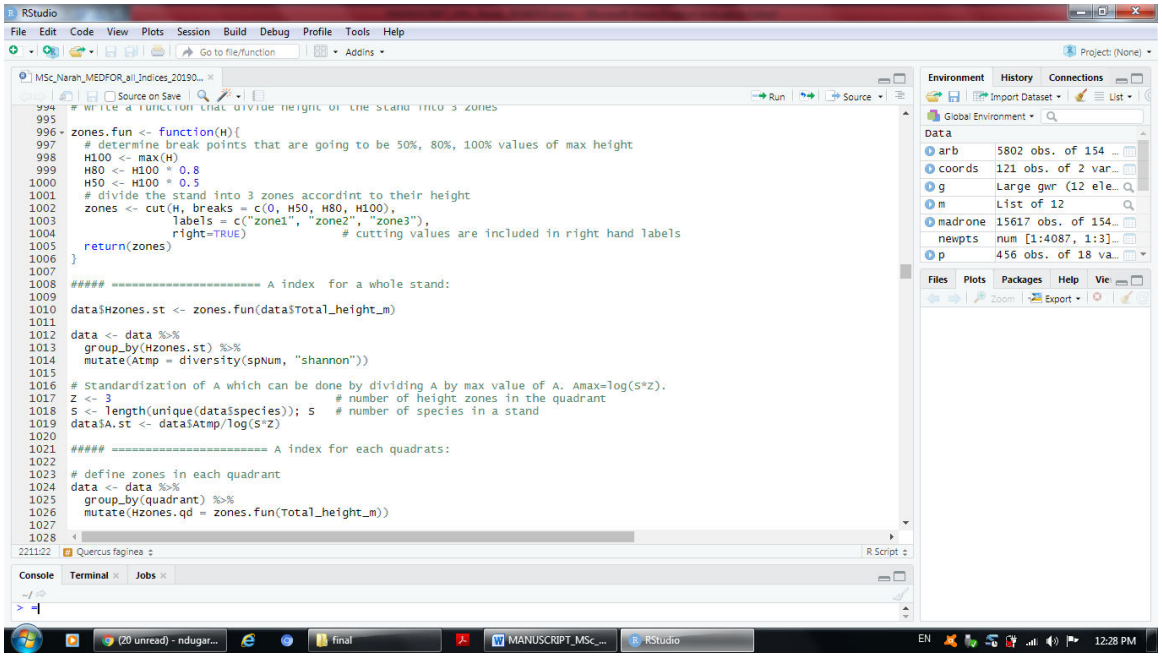
```
# thurifer and Pinus pinea) in sp.tree column match with each other in sp.near columns.

sub_fun <- function(data, i, j){
  subs <- subset(data, sp.tree %in% c(i, j) & sp.near %in% c(i, j))
  rem <- droplevels(subs)   # remove unused levels (0) from df
  return(rem)
}

# apply this function to subset species
jth.pp <- sub_fun(dist, "Juniperus thurifera", "Pinus pinea"); table(jth.pp$sp.tree, jth.pp$sp.near)
jth.qf <- sub_fun(dist, "Juniperus thurifera", "Quercus faginea"); table(jth.qf$sp.tree, jth.qf$sp.near)
jth.qI <- sub_fun(dist, "Juniperus thurifera", "Quercus ilex"); table(jth.qI$sp.tree, jth.qI$sp.near)
pp.qf <- sub_fun(dist, "Pinus pinea", "Quercus faginea"); table(pp.qf$sp.tree, pp.qf$sp.near)
pp.qI <- sub_fun(dist, "Pinus pinea", "Quercus ilex"); table(pp.qI$sp.tree, pp.qI$sp.near)
qf.qI <- sub_fun(dist, "Quercus faginea", "Quercus ilex"); table(qf.qI$sp.tree, qf.qI$sp.near)

# create a list of remaining datasets
# datalist <- list(jth.pp, jth.qf, jth.qI, pp.qf, pp.qI, qf.qI)

##### Start calculating S index:

# step by step calculation of S to calc S for each df.
# !!!! change data and get resultone by one
n <- table(qf.qI$sp.tree, qf.qI$sp.near); n   # create contingency table by ref & neigh trees
Nrow <- rowSums(n); Nrow                      # total number of reference tree by species
Ncol <- colSums(n); Ncol                      # total number of neighbouring tree by species
Ndissim <- length(which(qf.qI$sp.near!=qf.qI$sp.tree))  # number of mixture species (diff sp)
N <- length(qf.qI$sp.tree); N                 # total number of plants in whole data
obsProb <- N * Ndissim; obsProb               # for observed probability
expProb <- Nrow[1]*Ncol[2]+Nrow[2]*Ncol[1]; expProb # expected probability
qf.qI$S.ind <- 1-obsProb/expProb
# >>>>> ok
```



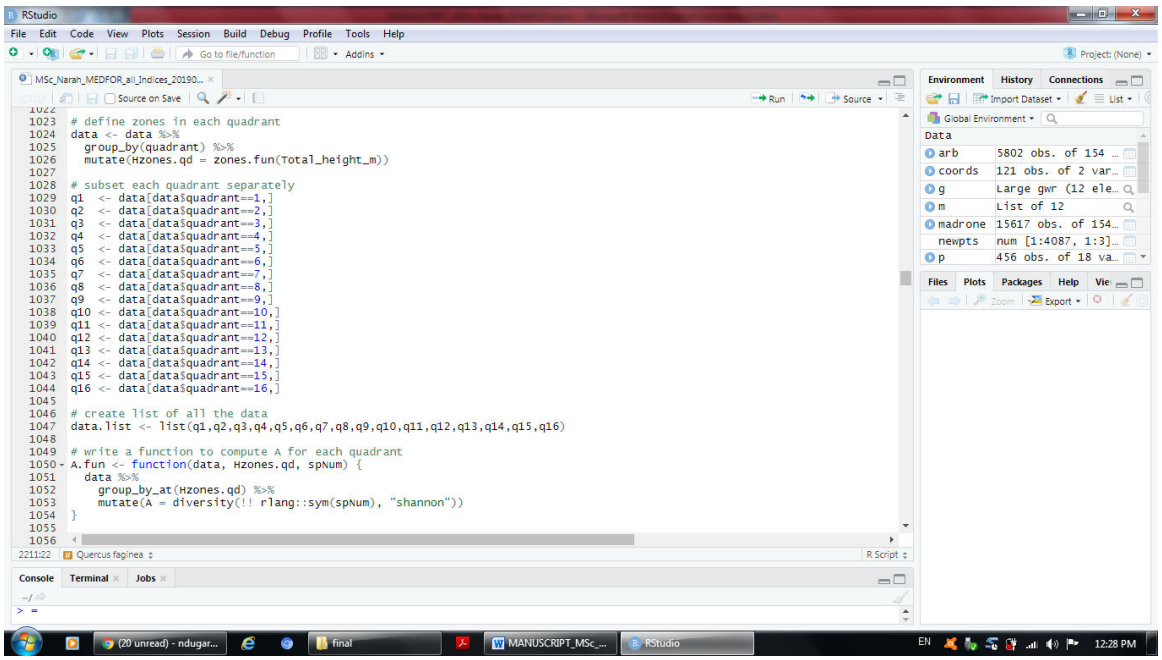
```
# convert list to data frame with 6 x 1
#S.ind <- matrix(t(sapply(outS.list,c))); S.ind          # or S.ind <- data.frame(unlist(outS.list))
#sp.pairs <- c("jth.pp","jth.qf","jth.qi","pp.qf","pp.qi","qf.qi") # define which value is for which pairs
#outputS <- data.frame(sp.pairs, S.ind); outputS         # create data.frame
# https://stackoverflow.com/questions/4227223/r-list-to-data-frame

# reconstruct dataframe of dist using the result of S
dist$S = if_else(dist$sp.tree %in% c("Juniperus thurifera", "Pinus pinea") &
                   dist$sp.near %in% c("Juniperus thurifera", "Pinus pinea"), 0.1900954,
                 if_else(dist$sp.tree %in% c("Juniperus thurifera", "Quercus faginea") &
                           dist$sp.near %in% c("Juniperus thurifera", "Quercus faginea"), 0.2822812,
                         if_else(dist$sp.tree %in% c("Juniperus thurifera", "Quercus ilex") &
                                   dist$sp.near %in% c("Juniperus thurifera", "Quercus ilex"),0.8903108,
                                 if_else(dist$sp.tree %in% c("Pinus pinea", "Quercus faginea") &
                                           dist$sp.near %in% c("Pinus pinea", "Quercus faginea"),0.1281021,
                                         if_else(dist$sp.tree %in% c("Pinus pinea", "Quercus ilex") &
                                                   dist$sp.near %in% c("Pinus pinea", "Quercus ilex"), 0.4289634,
                                                 if_else(dist$sp.tree %in% c("Quercus faginea", "Quercus ilex") &
                                                           dist$sp.near %in% c("Quercus faginea", "Quercus ilex"), 0.34501

head(dist)

# select reference tree informaitons
outputS <- dist %>%
  select(tree, sp.tree, dbh.tree, h.tree, x.tree, y.tree, distij, S) %>%
  group_by(tree1 = gl(n()/4, 4), sp.tree) %>%
  mutate(S = mean(S), dbh.tree = dbh.tree[1], h.tree = h.tree[1],
         x.tree = x.tree[1], y.tree = y.tree[1], distij = distij[1])  %>%
  slice(1)

head(outputS)

# write.csv(outputS, file="Segregation.csv")
```



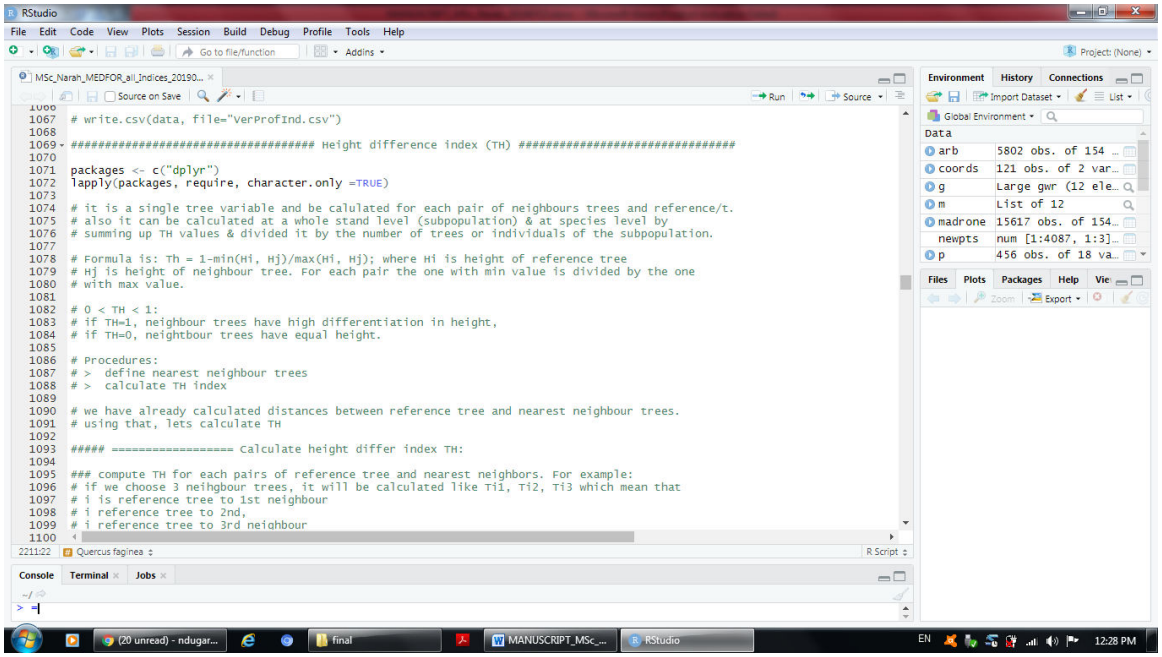
```
##################################### Vertical Species profile (A) ###################################

packages <-c("dplyr", "vegan", "tidyverse")
lapply(packages, require, character.only=TRUE)

# A index is based on Shannon index and proposed the differentation of tree species within each
# height layer. It takes into account proportion of species and number of layers in a stand.

# Its value is greater than 0.  0 is for a single-layered pure stand. The more heterogeneous
# the vertical profile, the higher the A value.


# convert name of species (column) in letter into numerical value because of vegan requirement.
# Species types must be in numeric type for shannon indices calculation using vegan package.

# There are 2 main steps to calculate A index
# >  classify height zones
# >  compute A index

##### classify height zones:

# divide Height into 3 zones: If we assume the height of the highest tree in the stand
# to be 100%, zone I extends from 100% down to 80%, zone 2 from 80% down to 50% and zone 3 from 50%
# down to the forest ground (figure 4). If the tip of a tree is located in one of these zones, we
# consider the tree belonging to this zone.

# zones:
# zone1 = 0-50% of total height
# zone2 = 50-80% of total height
# zone3 = 80-100% of total height

# write a function that divide height of the stand into 3 zones
```



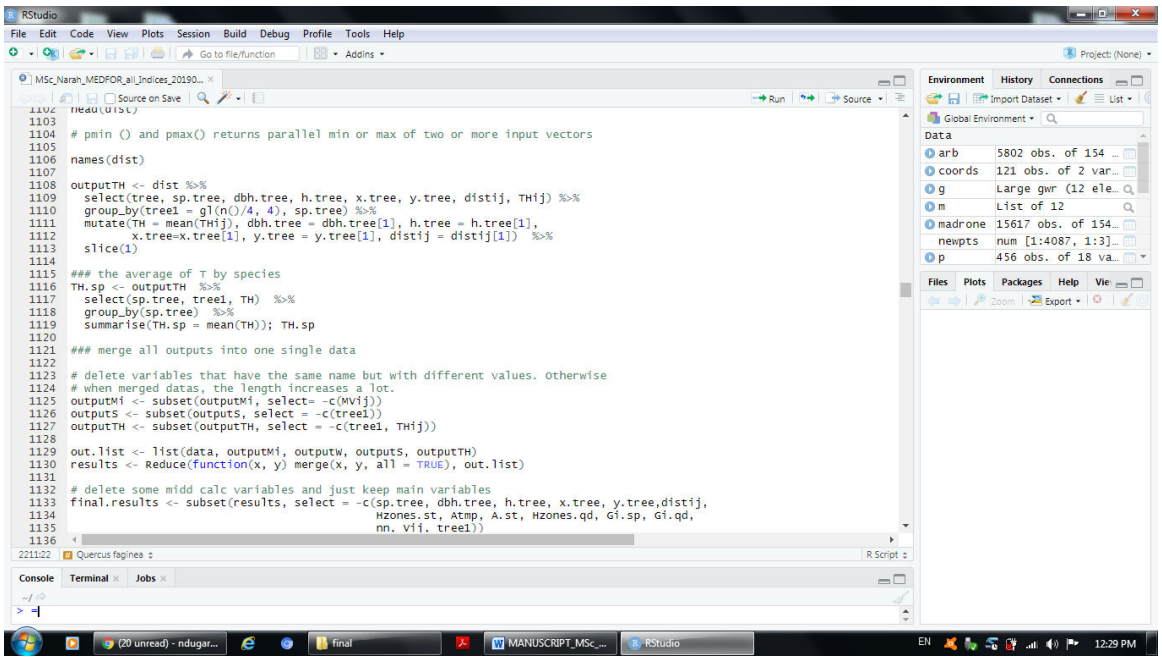
```
# write a function that divide height of the stand into 3 zones

zones.fun <- function(H){
  # determine break points that are going to be 50%, 80%, 100% values of max height
  H100 <- max(H)
  H80 <- H100 * 0.8
  H50 <- H100 * 0.5
  # divide the stand into 3 zones accordint to their height
  zones <- cut(H, breaks = c(0, H50, H80, H100),
               labels = c("zone1", "zone2", "zone3"),
               right=TRUE)              # cutting values are included in right hand labels
  return(zones)
}

##### ====================== A index  for a whole stand:

data$Hzones.st <- zones.fun(data$Total_height_m)

data <- data %>%
  group_by(Hzones.st) %>%
  mutate(Atmp = diversity(spNum, "shannon"))

# Standardization of A which can be done by dividing A by max value of A. Amax=log(S*Z).
Z <- 3                              # number of height zones in the quadrant
S <- length(unique(data$species)); S   # number of species in a stand
data$A.st <- data$Atmp/log(S*Z)

##### ====================== A index for each quadrats:

# define zones in each quadrant
data <- data %>%
  group_by(quadrant) %>%
  mutate(Hzones.qd = zones.fun(Total_height_m))
```



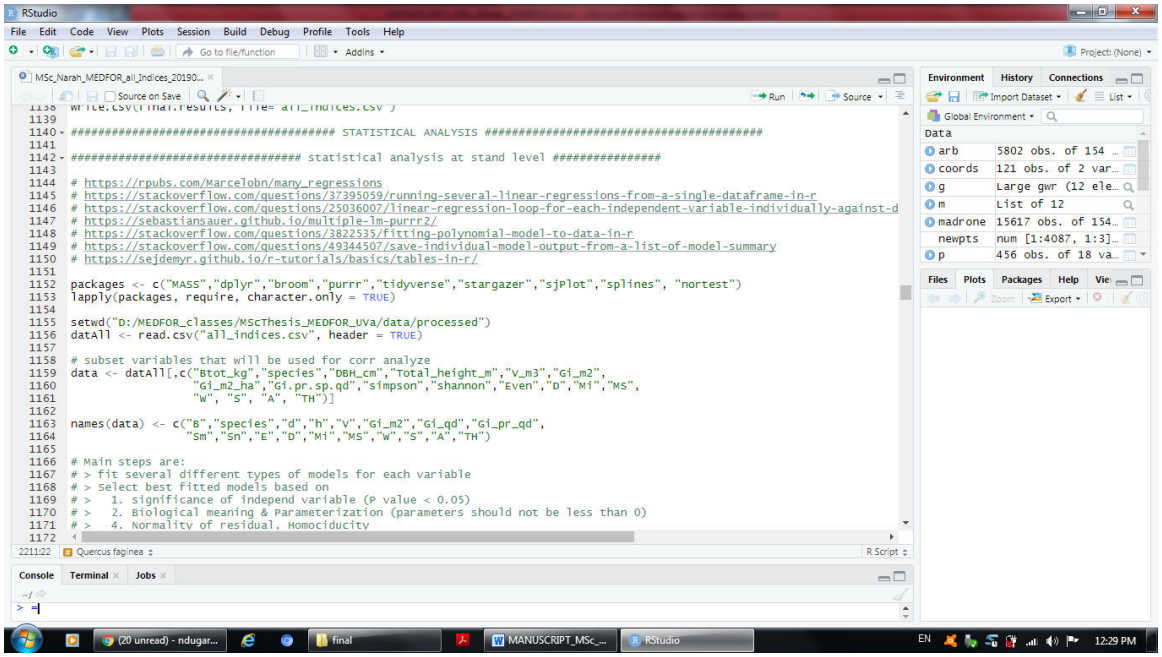
```
# define zones in each quadrant
data <- data %>%
  group_by(quadrant) %>%
  mutate(Hzones.qd = zones.fun(Total_height_m))

# subset each quadrant separately
q1  <- data[data$quadrant==1,]
q2  <- data[data$quadrant==2,]
q3  <- data[data$quadrant==3,]
q4  <- data[data$quadrant==4,]
q5  <- data[data$quadrant==5,]
q6  <- data[data$quadrant==6,]
q7  <- data[data$quadrant==7,]
q8  <- data[data$quadrant==8,]
q9  <- data[data$quadrant==9,]
q10 <- data[data$quadrant==10,]
q11 <- data[data$quadrant==11,]
q12 <- data[data$quadrant==12,]
q13 <- data[data$quadrant==13,]
q14 <- data[data$quadrant==14,]
q15 <- data[data$quadrant==15,]
q16 <- data[data$quadrant==16,]

# create list of all the data
data.list <- list(q1,q2,q3,q4,q5,q6,q7,q8,q9,q10,q11,q12,q13,q14,q15,q16)

# write a function to compute A for each quadrant
A.fun <- function(data, Hzones.qd, spNum) {
  data %>%
    group_by_at(Hzones.qd) %>%
    mutate(A = diversity(!! rlang::sym(spNum), "shannon"))
}
```



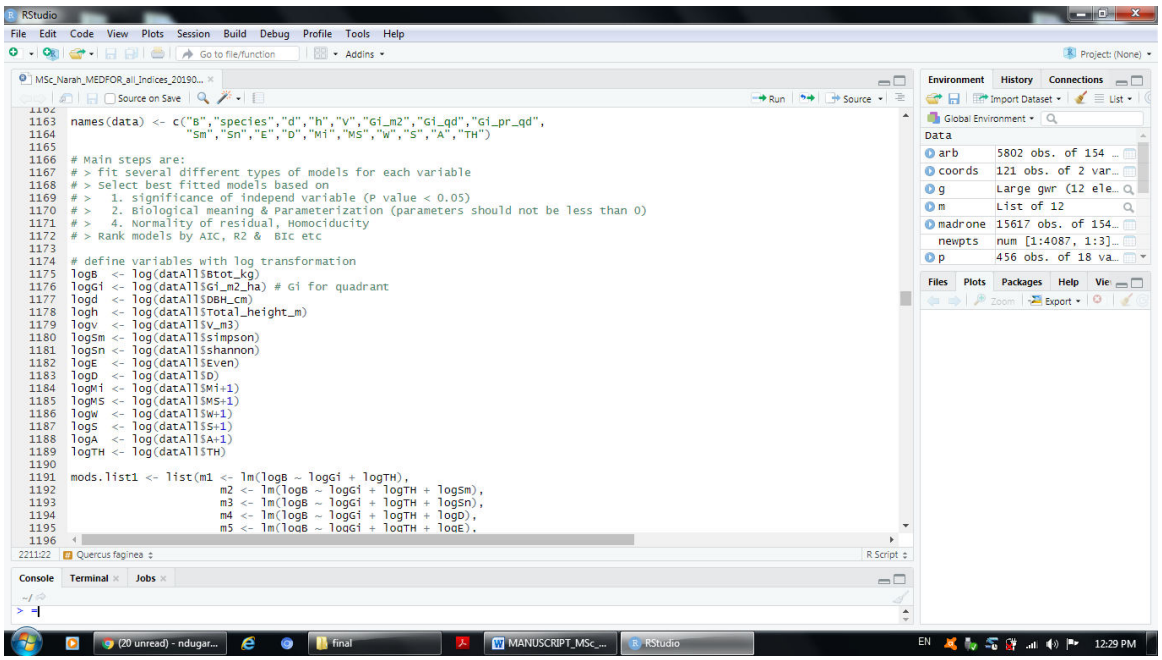
```
# write.csv(data, file="VerProfInd.csv")

################################### Height difference index (TH) ###############################

packages <- c("dplyr")
lapply(packages, require, character.only =TRUE)

# it is a single tree variable and be calulated for each pair of neighbours trees and reference/t.
# also it can be calculated at a whole stand level (subpopulation) & at species level by
# summing up TH values & divided it by the number of trees or individuals of the subpopulation.

# Formula is: Th = 1-min(Hi, Hj)/max(Hi, Hj); where Hi is height of reference tree
# Hj is height of neighbour tree. For each pair the one with min value is divided by the one
# with max value.

# 0 < TH < 1:
# if TH=1, neighbour trees have high differentiation in height,
# if TH=0, neighbtour trees have equal height.

# Procedures:
# >  define nearest neighbour trees
# >  calculate TH index

# we have already calculated distances between reference tree and nearest neighbour trees.
# using that, lets calculate TH

##### ================= Calculate height differ index TH:

### compute TH for each pairs of reference tree and nearest neighbors. For example:
# if we choose 3 neihgbour trees, it will be calculated like Ti1, Ti2, Ti3 which mean that
# i is reference tree to 1st neighbour
# i reference tree to 2nd,
# i reference tree to 3rd neighbour
```



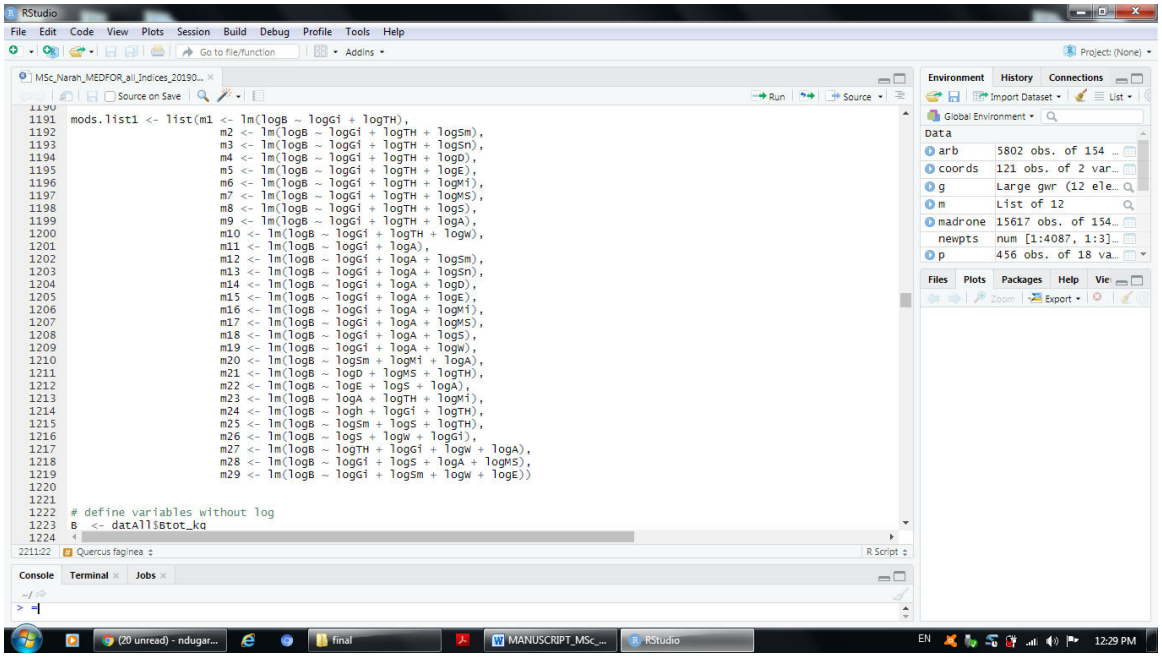
```
# pmin () and pmax() returns parallel min or max of two or more input vectors

names(dist)

outputTH <- dist %>%
  select(tree, sp.tree, dbh.tree, h.tree, x.tree, y.tree, distij, THij) %>%
  group_by(tree1 = gl(n()/4, 4), sp.tree) %>%
  mutate(TH = mean(THij), dbh.tree = dbh.tree[1], h.tree = h.tree[1],
         x.tree=x.tree[1], y.tree = y.tree[1], distij = distij[1])  %>%
  slice(1)

### the average of T by species
TH.sp <- outputTH  %>%
  select(sp.tree, tree1, TH)  %>%
  group_by(sp.tree)  %>%
  summarise(TH.sp = mean(TH)); TH.sp

### merge all outputs into one single data

# delete variables that have the same name but with different values. Otherwise
# when merged datas, the length increases a lot.
outputMi <- subset(outputMi, select= -c(MVij))
outputS <- subset(outputS, select = -c(tree1))
outputTH <- subset(outputTH, select = -c(tree1, THij))

out.list <- list(data, outputMi, outputW, outputS, outputTH)
results <- Reduce(function(x, y) merge(x, y, all = TRUE), out.list)

# delete some midd calc variables and just keep main variables
final.results <- subset(results, select = -c(sp.tree, dbh.tree, h.tree, x.tree, y.tree,distij,
                                             Hzones.st, Atmp, A.st, Hzones.qd, Gi.sp, Gi.qd,
                                             nn, Vij, tree1))
```



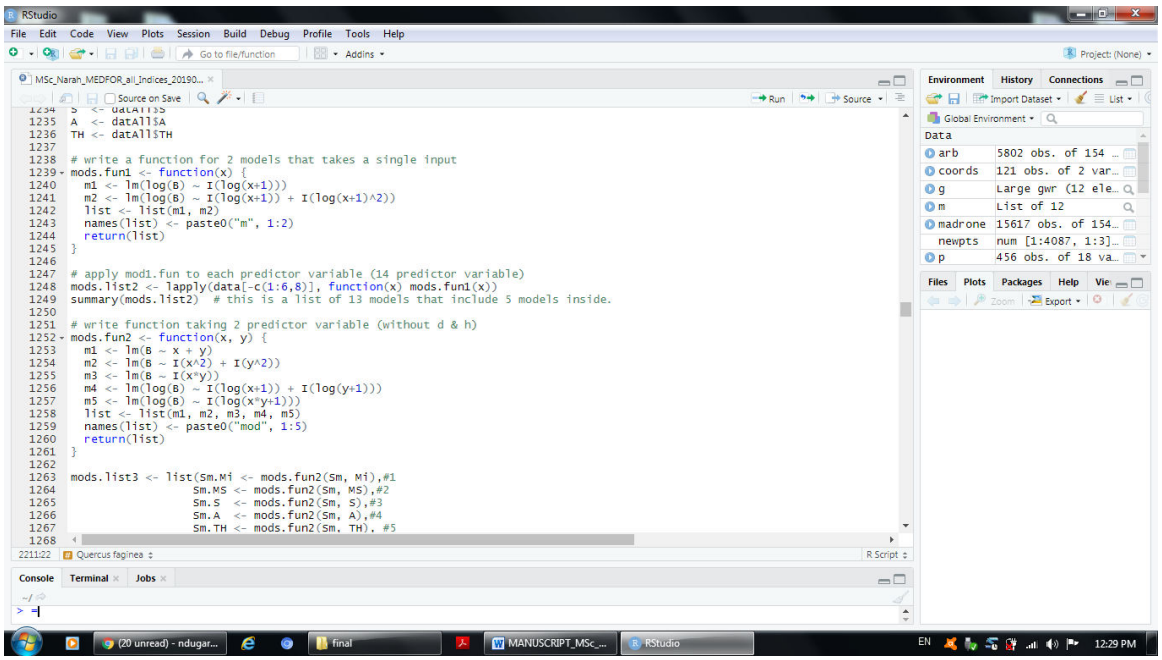
```
write.csv(final.results, file= all_indices.csv )

##################################### STATISTICAL ANALYSIS ######################################

################################## statistical analysis at stand level ###############

# https://rpubs.com/Marcelobn/many_regressions
# https://stackoverflow.com/questions/37395059/running-several-linear-regressions-from-a-single-dataframe-in-r
# https://stackoverflow.com/questions/25036007/linear-regression-loop-for-each-independent-variable-individually-against-d
# https://sebastiansauer.github.io/multiple-lm-purrr2/
# https://stackoverflow.com/questions/3822535/fitting-polynomial-model-to-data-in-r
# https://stackoverflow.com/questions/49344507/save-individual-model-output-from-a-list-of-model-summary
# https://sejdemyr.github.io/r-tutorials/basics/tables-in-r/

packages <- c("MASS","dplyr","broom","purrr","tidyverse","stargazer","sjPlot","splines", "nortest")
lapply(packages, require, character.only = TRUE)

setwd("D:/MEDFOR_classes/MScThesis_MEDFOR_UVa/data/processed")
datAll <- read.csv("all_indices.csv", header = TRUE)

# subset variables that will be used for corr analyze
data <- datAll[,c("Btot_kg","species","DBH_cm","Total_height_m","V_m3","Gi_m2",
                  "Gi_m2_ha","Gi.pr.sp.qd","simpson","shannon","Even","D","Mi","MS",
                  "W", "S", "A", "TH")]

names(data) <- c("B","species","d","h","V","Gi_m2","Gi_qd","Gi_pr_qd",
                 "Sm","Sn","E","D","Mi","MS","W","S","A","TH")

# Main steps are:
# > fit several different types of models for each variable
# > Select best fitted models based on
# >   1. significance of independ variable (P value < 0.05)
# >   2. Biological meaning & Parameterization (parameters should not be less than 0)
# >   4. Normality of residual, Homociducity
```



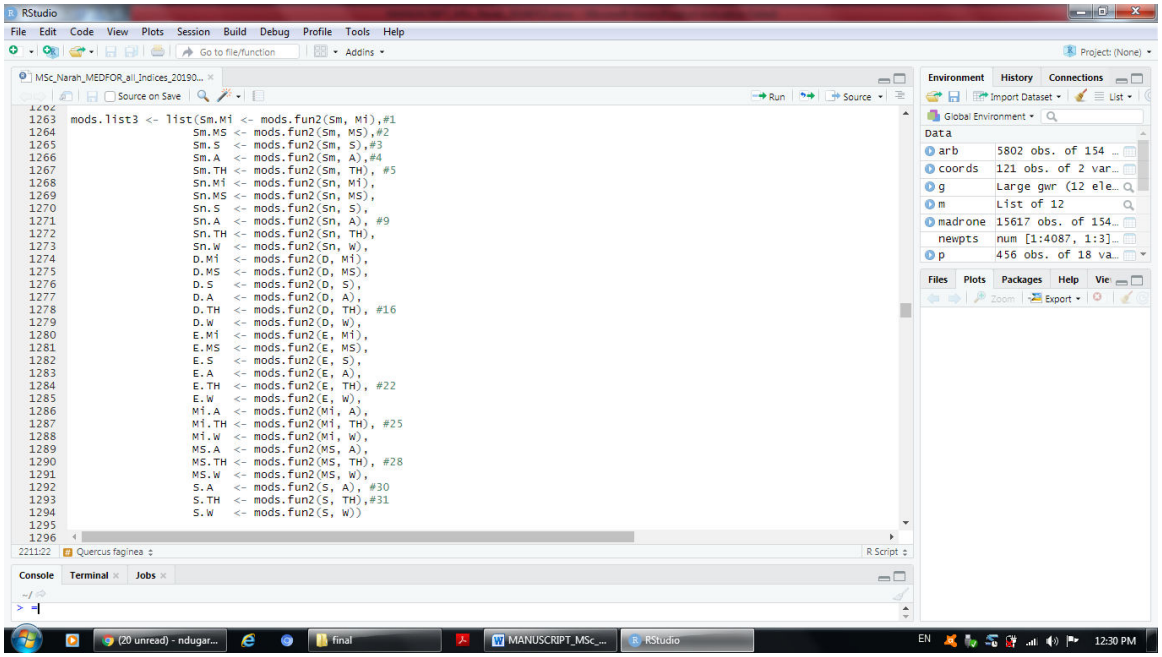
```
names(data) <- c("B","species","d","h","V","Gi_m2","Gi_qd","Gi_pr_qd",
                 "Sm","Sn","E","D","Mi","MS","W","S","A","TH")

# Main steps are:
# > fit several different types of models for each variable
# > Select best fitted models based on
# >   1. significance of independ variable (P value < 0.05)
# >   2. Biological meaning & Parameterization (parameters should not be less than 0)
# >   4. Normality of residual, Homociducity
# > Rank models by AIC, R2 &  BIc etc

# define variables with log transformation
logB  <- log(datAll$Btot_kg)
logGi <- log(datAll$Gi_m2_ha) # Gi for quadrant
logd  <- log(datAll$DBH_cm)
logh  <- log(datAll$Total_height_m)
logv  <- log(datAll$V_m3)
logSm <- log(datAll$simpson)
logSn <- log(datAll$shannon)
logE  <- log(datAll$Even)
logD  <- log(datAll$D)
logMi <- log(datAll$Mi+1)
logMS <- log(datAll$MS+1)
logW  <- log(datAll$W+1)
logS  <- log(datAll$S+1)
logA  <- log(datAll$A+1)
logTH <- log(datAll$TH)

mods.list1 <- list(m1 <- lm(logB ~ logGi + logTH),
                   m2 <- lm(logB ~ logGi + logTH + logSm),
                   m3 <- lm(logB ~ logGi + logTH + logSn),
                   m4 <- lm(logB ~ logGi + logTH + logD),
                   m5 <- lm(logB ~ logGi + logTH + logE),
```

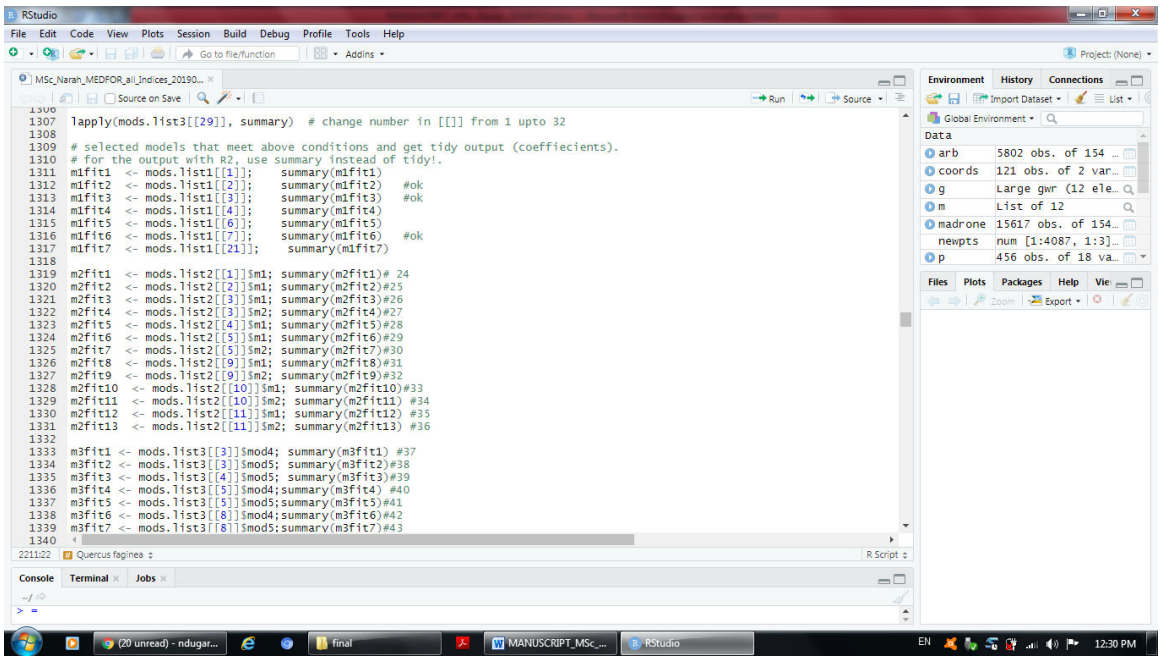


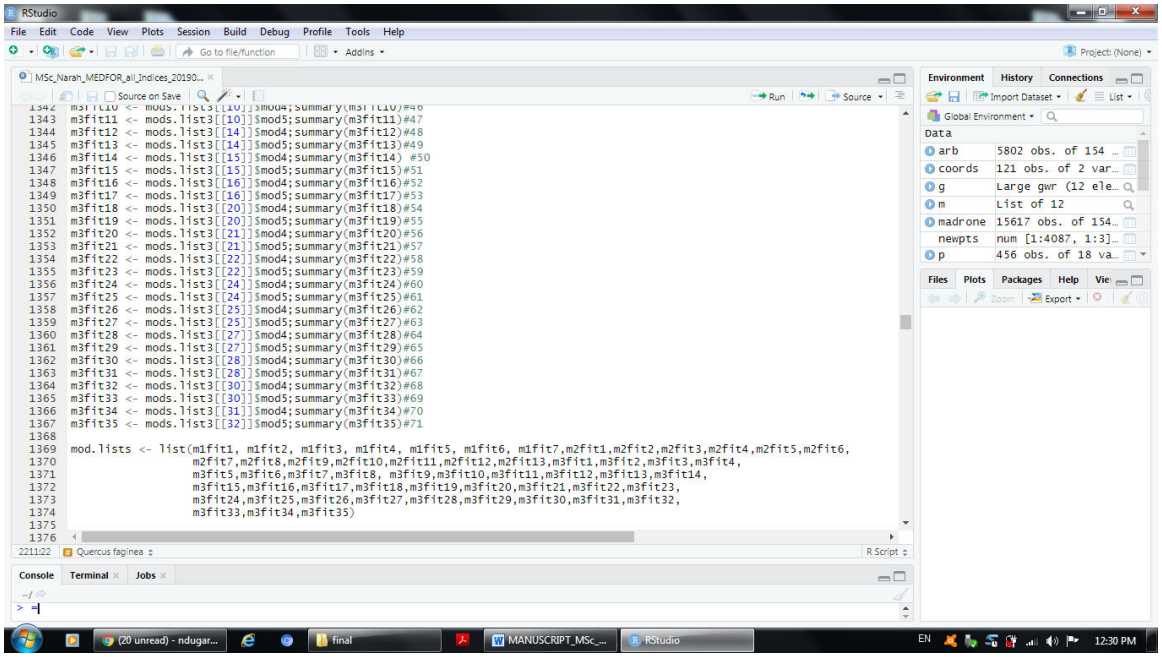
```
mods.list1 <- list(m1 <- lm(logB ~ logGi + logTH),
                   m2 <- lm(logB ~ logGi + logTH + logSm),
                   m3 <- lm(logB ~ logGi + logTH + logSn),
                   m4 <- lm(logB ~ logGi + logTH + logD),
                   m5 <- lm(logB ~ logGi + logTH + logE),
                   m6 <- lm(logB ~ logGi + logTH + logMi),
                   m7 <- lm(logB ~ logGi + logTH + logMS),
                   m8 <- lm(logB ~ logGi + logTH + logS),
                   m9 <- lm(logB ~ logGi + logTH + logA),
                   m10 <- lm(logB ~ logGi + logTH + logW),
                   m11 <- lm(logB ~ logGi + logA),
                   m12 <- lm(logB ~ logGi + logA + logSm),
                   m13 <- lm(logB ~ logGi + logA + logSn),
                   m14 <- lm(logB ~ logGi + logA + logD),
                   m15 <- lm(logB ~ logGi + logA + logE),
                   m16 <- lm(logB ~ logGi + logA + logMi),
                   m17 <- lm(logB ~ logGi + logA + logMS),
                   m18 <- lm(logB ~ logGi + logA + logS),
                   m19 <- lm(logB ~ logGi + logA + logW),
                   m20 <- lm(logB ~ logSm + logMi + logA),
                   m21 <- lm(logB ~ logD + logMS + logTH),
                   m22 <- lm(logB ~ logE + logS + logA),
                   m23 <- lm(logB ~ logA + logTH + logMi),
                   m24 <- lm(logB ~ logh + logGi + logTH),
                   m25 <- lm(logB ~ logSm + logS + logTH),
                   m26 <- lm(logB ~ logS + logW + logGi),
                   m27 <- lm(logB ~ logTH + logGi + logW + logA),
                   m28 <- lm(logB ~ logGi + logS + logA + logMS),
                   m29 <- lm(logB ~ logGi + logSm + logW + logE))


# define variables without log
B  <- dataAll$Btot_kg
```



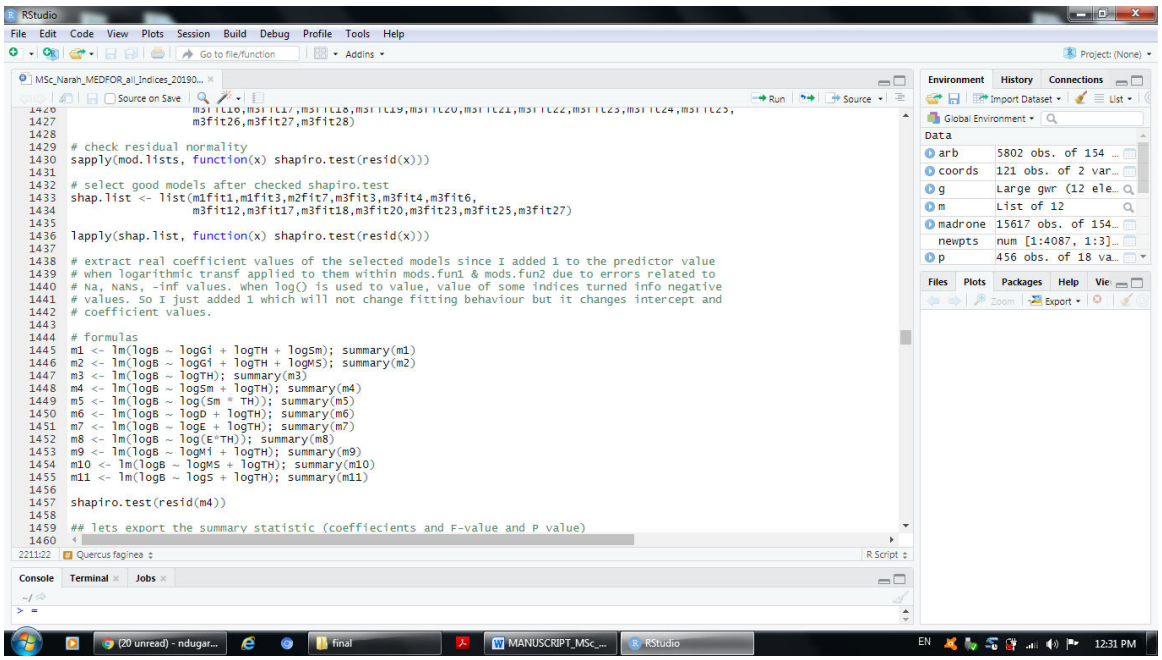
```
A  <- dataAll$A
TH <- dataAll$TH

# write a function for 2 models that takes a single input
mods.fun1 <- function(x) {
  m1 <- lm(log(B) ~ I(log(x+1)))
  m2 <- lm(log(B) ~ I(log(x+1)) + I(log(x+1)^2))
  list <- list(m1, m2)
  names(list) <- paste0("m", 1:2)
  return(list)
}

# apply mod1.fun to each predictor variable (14 predictor variable)
mods.list2 <- lapply(data[-c(1:6,8)], function(x) mods.fun1(x))
summary(mods.list2)  # this is a list of 13 models that include 5 models inside.

# write function taking 2 predictor variable (without d & h)
mods.fun2 <- function(x, y) {
  m1 <- lm(B ~ x + y)
  m2 <- lm(B ~ I(x^2) + I(y^2))
  m3 <- lm(B ~ I(x*y))
  m4 <- lm(log(B) ~ I(log(x+1)) + I(log(y+1)))
  m5 <- lm(log(B) ~ I(log(x*y+1)))
  list <- list(m1, m2, m3, m4, m5)
  names(list) <- paste0("mod", 1:5)
  return(list)
}

mods.list3 <- list(Sm.Mi <- mods.fun2(Sm, Mi),#1
                   Sm.MS <- mods.fun2(Sm, MS),#2
                   Sm.S  <- mods.fun2(Sm, S),#3
                   Sm.A  <- mods.fun2(Sm, A),#4
                   Sm.TH <- mods.fun2(Sm, TH), #5
```


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
mods.list3 <- list(Sm.Mi <- mods.fun2(Sm, Mi),#1
Sm.MS <- mods.fun2(Sm, MS),#2
Sm.S <- mods.fun2(Sm, S),#3
Sm.A <- mods.fun2(Sm, A),#4
Sm.TH <- mods.fun2(Sm, TH), #5
Sn.Mi <- mods.fun2(Sn, Mi),
Sn.MS <- mods.fun2(Sn, MS),
Sn.S <- mods.fun2(Sn, S),
Sn.A <- mods.fun2(Sn, A), #9
Sn.TH <- mods.fun2(Sn, TH),
Sn.W <- mods.fun2(Sn, W),
D.Mi <- mods.fun2(D, Mi),
D.MS <- mods.fun2(D, MS),
D.S <- mods.fun2(D, S),
D.A <- mods.fun2(D, A),
D.TH <- mods.fun2(D, TH), #16
D.W <- mods.fun2(D, W),
E.Mi <- mods.fun2(E, Mi),
E.MS <- mods.fun2(E, MS),
E.S <- mods.fun2(E, S),
E.A <- mods.fun2(E, A),
E.TH <- mods.fun2(E, TH), #22
E.W <- mods.fun2(E, W),
Mi.A <- mods.fun2(Mi, A),
Mi.TH <- mods.fun2(Mi, TH), #25
Mi.W <- mods.fun2(Mi, W),
MS.A <- mods.fun2(MS, A),
MS.TH <- mods.fun2(MS, TH), #28
MS.W <- mods.fun2(MS, W),
S.A <- mods.fun2(S, A), #30
S.TH <- mods.fun2(S, TH),#31
S.W <- mods.fun2(S, W))
Environment History Connections
Import Dataset
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
Console Terminal Jobs
Quercus faginea
R Script
(20 unread) - ndugar...
final
MANUSCRIPT_MSc_...
RStudio
EN
12:30 PM

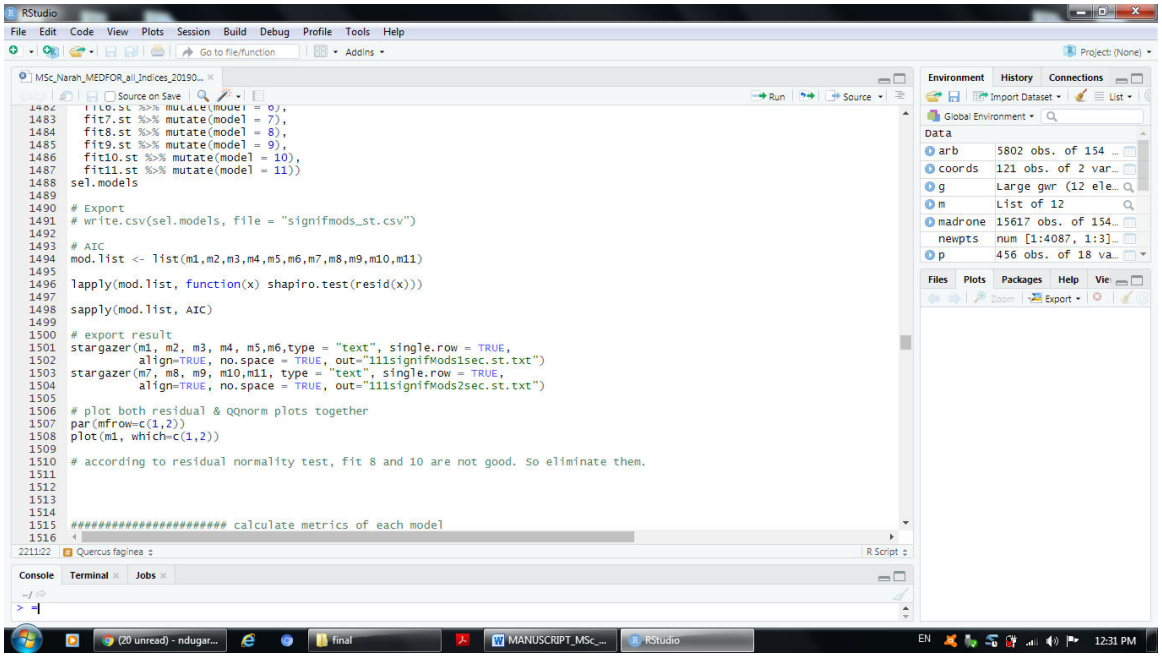

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
lapply(mods.list3[[29]], summary) # change number in [[]] from 1 upto 32
# selected models that meet above conditions and get tidy output (coeffiecients).
# for the output with R2, use summary instead of tidy!.
m1fit1 <- mods.list1[[1]]; summary(m1fit1)
m1fit2 <- mods.list1[[2]]; summary(m1fit2) #ok
m1fit3 <- mods.list1[[3]]; summary(m1fit3) #ok
m1fit4 <- mods.list1[[4]]; summary(m1fit4)
m1fit5 <- mods.list1[[6]]; summary(m1fit5)
m1fit6 <- mods.list1[[7]]; summary(m1fit6) #ok
m1fit7 <- mods.list1[[21]]; summary(m1fit7)
m2fit1 <- mods.list2[[1]]$m1; summary(m2fit1)# 24
m2fit2 <- mods.list2[[2]]$m1; summary(m2fit2)#25
m2fit3 <- mods.list2[[3]]$m1; summary(m2fit3)#26
m2fit4 <- mods.list2[[3]]$m2; summary(m2fit4)#27
m2fit5 <- mods.list2[[4]]$m1; summary(m2fit5)#28
m2fit6 <- mods.list2[[5]]$m1; summary(m2fit6)#29
m2fit7 <- mods.list2[[5]]$m2; summary(m2fit7)#30
m2fit8 <- mods.list2[[9]]$m1; summary(m2fit8)#31
m2fit9 <- mods.list2[[9]]$m2; summary(m2fit9)#32
m2fit10 <- mods.list2[[10]]$m1; summary(m2fit10)#33
m2fit11 <- mods.list2[[10]]$m2; summary(m2fit11) #34
m2fit12 <- mods.list2[[11]]$m1; summary(m2fit12) #35
m2fit13 <- mods.list2[[11]]$m2; summary(m2fit13) #36
m3fit1 <- mods.list3[[3]]$mod4; summary(m3fit1) #37
m3fit2 <- mods.list3[[3]]$mod5; summary(m3fit2)#38
m3fit3 <- mods.list3[[4]]$mod5; summary(m3fit3)#39
m3fit4 <- mods.list3[[5]]$mod4;summary(m3fit4) #40
m3fit5 <- mods.list3[[5]]$mod5;summary(m3fit5)#41
m3fit6 <- mods.list3[[8]]$mod4;summary(m3fit6)#42
m3fit7 <- mods.list3[[8]]$mod5;summary(m3fit7)#43
Environment History Connections
Import Dataset
Global Environment
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Files Plots Packages Help Vie
Export
Console Terminal Jobs
Quercus faginea
R Script
(20 unread) - ndugar...
final
MANUSCRIPT_MSc_...
RStudio
EN
12:30 PM

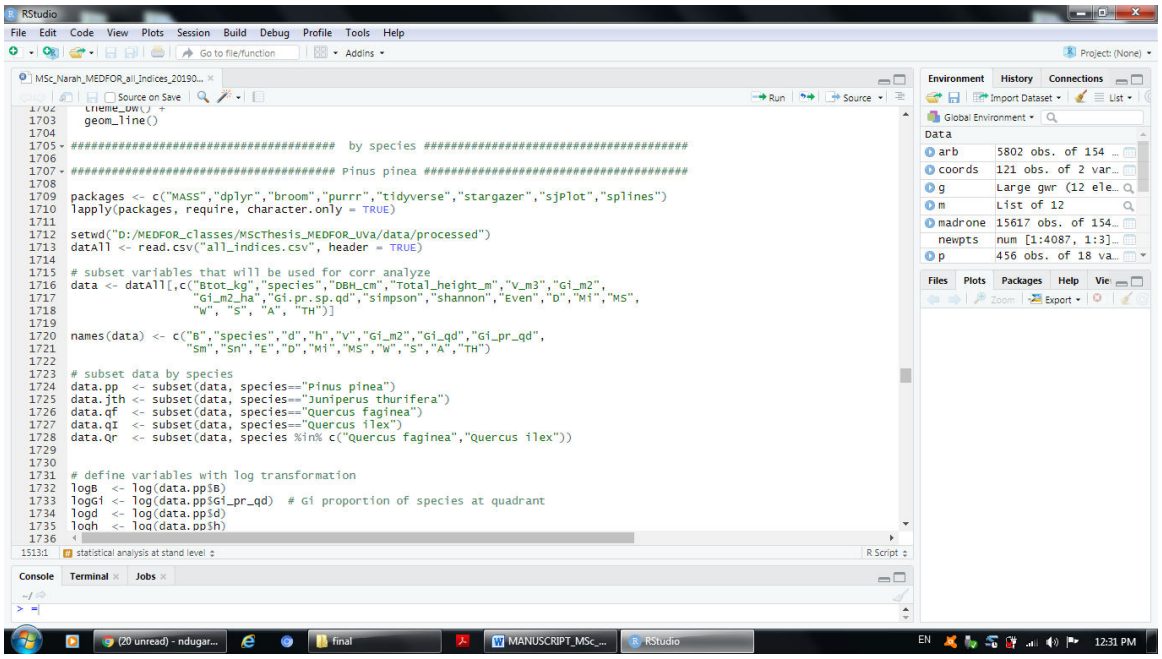


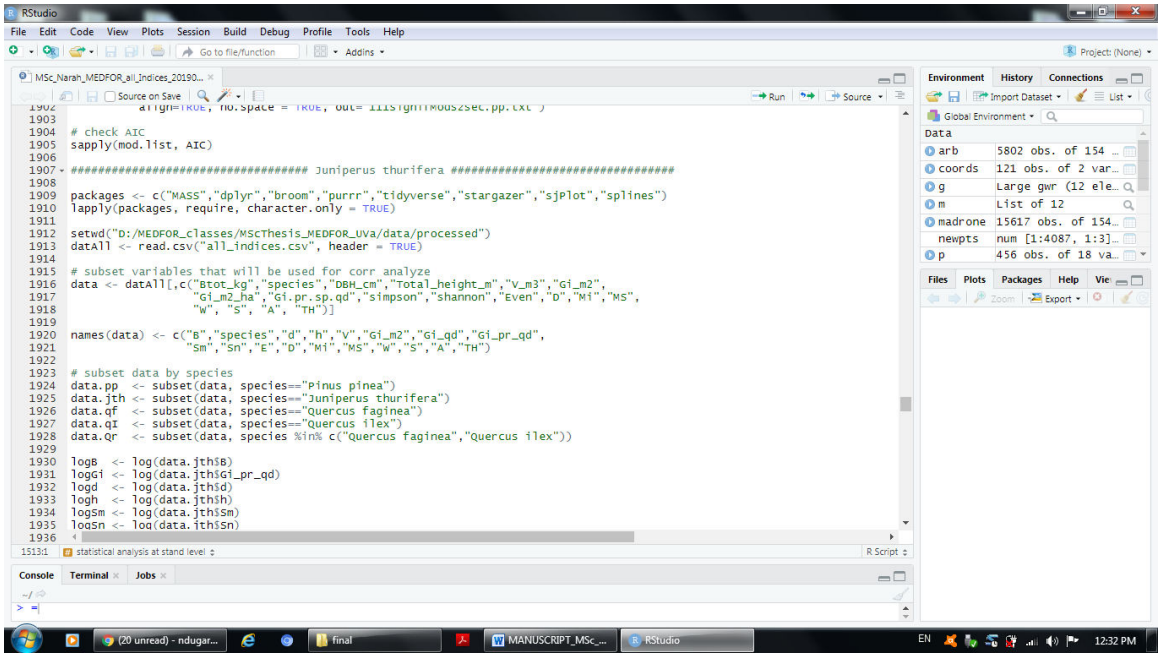
```
m3fit10 <- mods.list3[[10]]$mod4;summary(m3fit10)#46
m3fit11 <- mods.list3[[10]]$mod5;summary(m3fit11)#47
m3fit12 <- mods.list3[[14]]$mod4;summary(m3fit12)#48
m3fit13 <- mods.list3[[14]]$mod5;summary(m3fit13)#49
m3fit14 <- mods.list3[[15]]$mod4;summary(m3fit14) #50
m3fit15 <- mods.list3[[15]]$mod5;summary(m3fit15)#51
m3fit16 <- mods.list3[[16]]$mod4;summary(m3fit16)#52
m3fit17 <- mods.list3[[16]]$mod5;summary(m3fit17)#53
m3fit18 <- mods.list3[[20]]$mod4;summary(m3fit18)#54
m3fit19 <- mods.list3[[20]]$mod5;summary(m3fit19)#55
m3fit20 <- mods.list3[[21]]$mod4;summary(m3fit20)#56
m3fit21 <- mods.list3[[21]]$mod5;summary(m3fit21)#57
m3fit22 <- mods.list3[[22]]$mod4;summary(m3fit22)#58
m3fit23 <- mods.list3[[22]]$mod5;summary(m3fit23)#59
m3fit24 <- mods.list3[[24]]$mod4;summary(m3fit24)#60
m3fit25 <- mods.list3[[24]]$mod5;summary(m3fit25)#61
m3fit26 <- mods.list3[[25]]$mod4;summary(m3fit26)#62
m3fit27 <- mods.list3[[25]]$mod5;summary(m3fit27)#63
m3fit28 <- mods.list3[[27]]$mod4;summary(m3fit28)#64
m3fit29 <- mods.list3[[27]]$mod5;summary(m3fit29)#65
m3fit30 <- mods.list3[[28]]$mod4;summary(m3fit30)#66
m3fit31 <- mods.list3[[28]]$mod5;summary(m3fit31)#67
m3fit32 <- mods.list3[[30]]$mod4;summary(m3fit32)#68
m3fit33 <- mods.list3[[30]]$mod5;summary(m3fit33)#69
m3fit34 <- mods.list3[[31]]$mod4;summary(m3fit34)#70
m3fit35 <- mods.list3[[32]]$mod5;summary(m3fit35)#71

mod.lists <- list(m1fit1, m1fit2, m1fit3, m1fit4, m1fit5, m1fit6, m1fit7,m2fit1,m2fit2,m2fit3,m2fit4,m2fit5,m2fit6,
                  m2fit7,m2fit8,m2fit9,m2fit10,m2fit11,m2fit12,m2fit13,m3fit1,m3fit2,m3fit3,m3fit4,
                  m3fit5,m3fit6,m3fit7,m3fit8, m3fit9,m3fit10,m3fit11,m3fit12,m3fit13,m3fit14,
                  m3fit15,m3fit16,m3fit17,m3fit18,m3fit19,m3fit20,m3fit21,m3fit22,m3fit23,
                  m3fit24,m3fit25,m3fit26,m3fit27,m3fit28,m3fit29,m3fit30,m3fit31,m3fit32,
                  m3fit33,m3fit34,m3fit35)
```



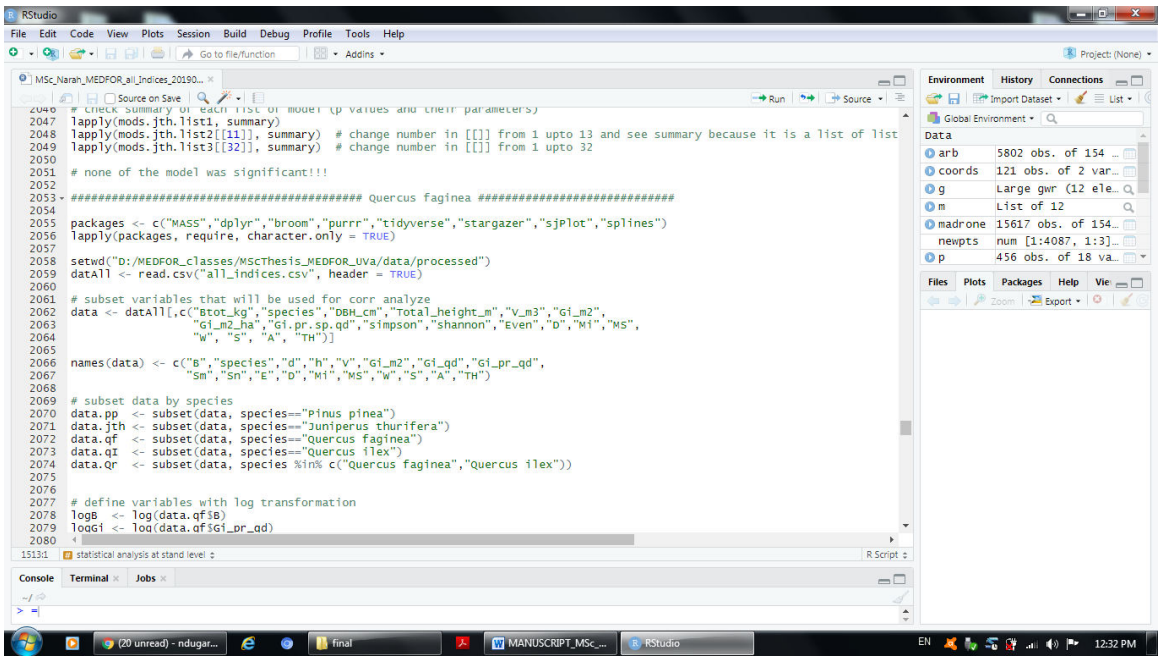
```
                  m3fit26,m3fit27,m3fit28)

# check residual normality
sapply(mod.lists, function(x) shapiro.test(resid(x)))

# select good models after checked shapiro.test
shap.list <- list(m1fit1,m1fit3,m2fit7,m3fit3,m3fit4,m3fit6,
                  m3fit12,m3fit17,m3fit18,m3fit20,m3fit23,m3fit25,m3fit27)

lapply(shap.list, function(x) shapiro.test(resid(x)))

# extract real coefficient values of the selected models since I added 1 to the predictor value
# when logarithmic transf applied to them within mods.fun1 & mods.fun2 due to errors related to
# Na, NaNs, -inf values. When log() is used to value, value of some indices turned info negative
# values. So I just added 1 which will not change fitting behaviour but it changes intercept and
# coefficient values.

# formulas
m1 <- lm(logB ~ logGi + logTH + logSm); summary(m1)
m2 <- lm(logB ~ logGi + logTH + logMS); summary(m2)
m3 <- lm(logB ~ logTH); summary(m3)
m4 <- lm(logB ~ logSm + logTH); summary(m4)
m5 <- lm(logB ~ log(Sm * TH)); summary(m5)
m6 <- lm(logB ~ logD + logTH); summary(m6)
m7 <- lm(logB ~ logE + logTH); summary(m7)
m8 <- lm(logB ~ log(E*TH)); summary(m8)
m9 <- lm(logB ~ logMi + logTH); summary(m9)
m10 <- lm(logB ~ logMS + logTH); summary(m10)
m11 <- lm(logB ~ logS + logTH); summary(m11)

shapiro.test(resid(m4))

## lets export the summary statistic (coeffiecients and F-value and P value)
```


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
fit7.st %>% mutate(model = 7),
fit8.st %>% mutate(model = 8),
fit9.st %>% mutate(model = 9),
fit10.st %>% mutate(model = 10),
fit11.st %>% mutate(model = 11))
sel.models
# Export
# write.csv(sel.models, file = "signifmods_st.csv")
# AIC
mod.list <- list(m1,m2,m3,m4,m5,m6,m7,m8,m9,m10,m11)
lapply(mod.list, function(x) shapiro.test(resid(x)))
sapply(mod.list, AIC)
# export result
stargazer(m1, m2, m3, m4, m5,m6,type = "text", single.row = TRUE,
align=TRUE, no.space = TRUE, out="111signifMods1sec.st.txt")
stargazer(m7, m8, m9, m10,m11, type = "text", single.row = TRUE,
align=TRUE, no.space = TRUE, out="111signifMods2sec.st.txt")
# plot both residual & QQnorm plots together
par(mfrow=c(1,2))
plot(m1, which=c(1,2))
# according to residual normality test, fit 8 and 10 are not good. So eliminate them.
###################### calculate metrics of each model
Environment History Connections
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Console Terminal Jobs
(20 unread) - ndugar...
final
MANUSCRIPT_MSc_...
RStudio
12:31 PM

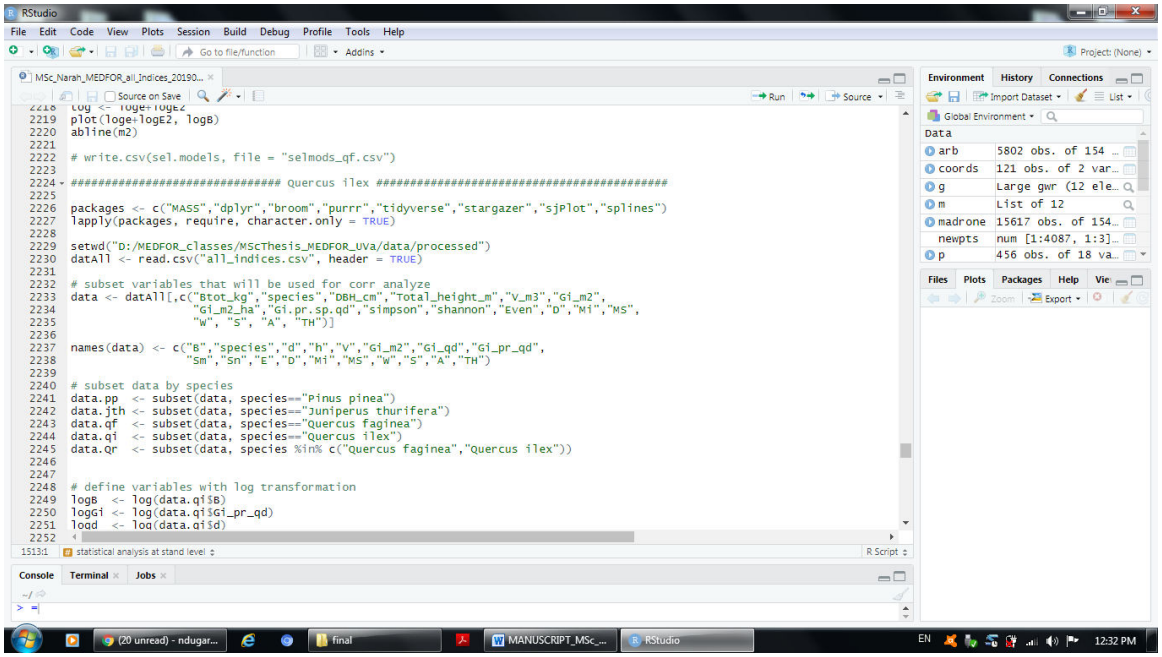

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
MSc_Narah_MEDFOR_all_Indices_20190...
geom_line()
#################################### by species ####################################
#################################### Pinus pinea ####################################
packages <- c("MASS","dplyr","broom","purrr","tidyverse","stargazer","sjPlot","splines")
lapply(packages, require, character.only = TRUE)
setwd("D:/MEDFOR_classes/MScThesis_MEDFOR_UVa/data/processed")
dataAll <- read.csv("all_indices.csv", header = TRUE)
# subset variables that will be used for corr analyze
data <- dataAll[,c("Btot_kg","species","DBH_cm","Total_height_m","V_m3","Gi_m2",
"Gi_m2_ha","Gi.pr.sp.qd","simpson","shannon","Even","D","Mi","MS",
"W", "S", "A", "TH")]
names(data) <- c("B","species","d","h","V","Gi_m2","Gi_qd","Gi_pr_qd",
"Sm","Sn","E","D","Mi","MS","W","S","A","TH")
# subset data by species
data.pp <- subset(data, species=="Pinus pinea")
data.jth <- subset(data, species=="Juniperus thurifera")
data.qf <- subset(data, species=="Quercus faginea")
data.qI <- subset(data, species=="Quercus ilex")
data.Qr <- subset(data, species %in% c("Quercus faginea","Quercus ilex"))
# define variables with log transformation
logB <- log(data.pp$B)
logGi <- log(data.pp$Gi_pr_qd) # Gi proportion of species at quadrant
logd <- log(data.pp$d)
logh <- log(data.pp$h)
Environment History Connections
Data
arb 5802 obs. of 154 ...
coords 121 obs. of 2 var...
g Large gwr (12 ele...
m List of 12
madrone 15617 obs. of 154...
newpts num [1:4087, 1:3]...
p 456 obs. of 18 va...
Console Terminal Jobs
(20 unread) - ndugar...
final
MANUSCRIPT_MSc_...
RStudio
12:31 PM



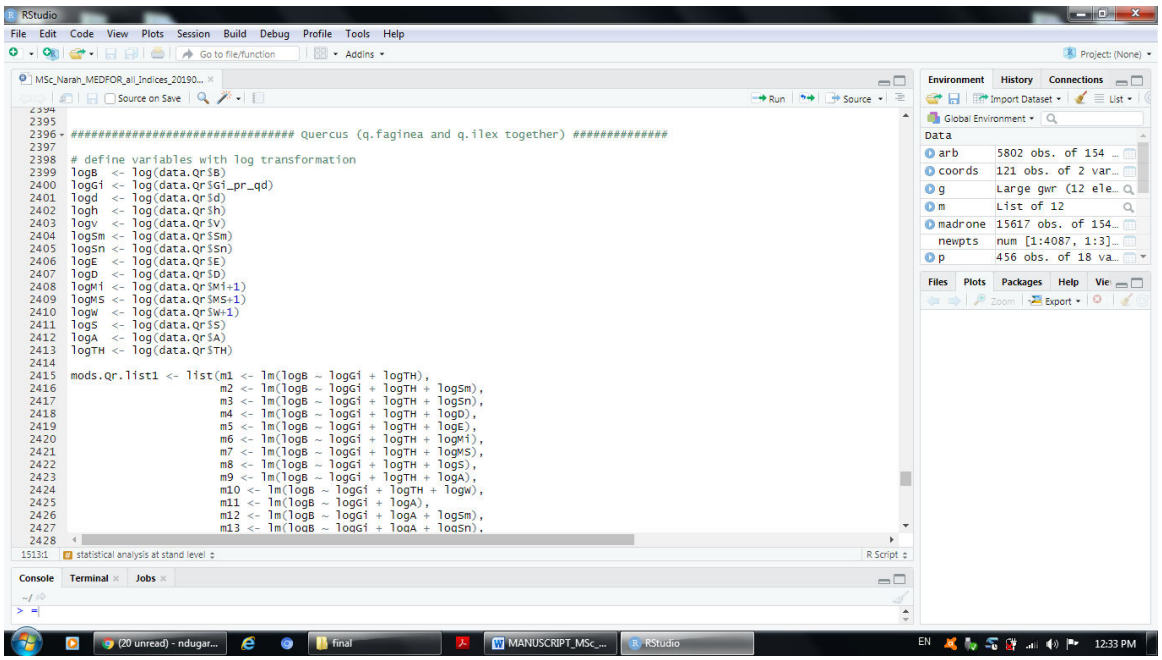
```
# check AIC
sapply(mod.list, AIC)

################################## Juniperus thurifera ##############################

packages <- c("MASS","dplyr","broom","purrr","tidyverse","stargazer","sjPlot","splines")
lapply(packages, require, character.only = TRUE)

setwd("D:/MEDFOR_classes/MScThesis_MEDFOR_UVa/data/processed")
dataAll <- read.csv("all_indices.csv", header = TRUE)

# subset variables that will be used for corr analyze
data <- dataAll[,c("Btot_kg","species","DBH_cm","Total_height_m","V_m3","Gi_m2",
                   "Gi_m2_ha","Gi.pr.sp.qd","simpson","shannon","Even","D","Mi","MS",
                   "W", "S", "A", "TH")]

names(data) <- c("B","species","d","h","V","Gi_m2","Gi_qd","Gi_pr_qd",
                 "Sm","Sn","E","D","Mi","MS","W","S","A","TH")

# subset data by species
data.pp  <- subset(data, species=="Pinus pinea")
data.jth <- subset(data, species=="Juniperus thurifera")
data.qf  <- subset(data, species=="Quercus faginea")
data.qI  <- subset(data, species=="Quercus ilex")
data.Qr  <- subset(data, species %in% c("Quercus faginea","Quercus ilex"))

logB  <- log(data.jth$B)
logGi <- log(data.jth$Gi_pr_qd)
logd  <- log(data.jth$d)
logh  <- log(data.jth$h)
logSm <- log(data.jth$Sm)
logSn <- log(data.jth$Sn)
```



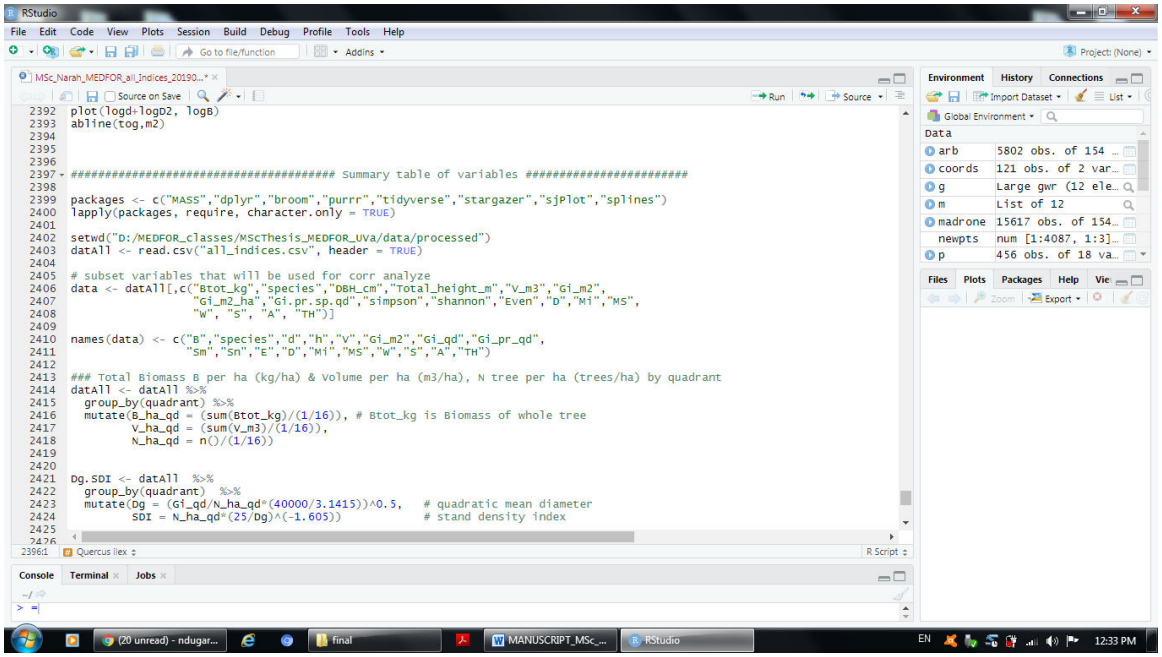
```
# check summary of each list of model (p values and their parameters)
lapply(mods.jth.list1, summary)
lapply(mods.jth.list2[[11]], summary)  # change number in [[]] from 1 upto 13 and see summary because it is a list of list
lapply(mods.jth.list3[[32]], summary)  # change number in [[]] from 1 upto 32

# none of the model was significant!!!

######################################## Quercus faginea ############################

packages <- c("MASS","dplyr","broom","purrr","tidyverse","stargazer","sjPlot","splines")
lapply(packages, require, character.only = TRUE)

setwd("D:/MEDFOR_classes/MScThesis_MEDFOR_UVa/data/processed")
dataAll <- read.csv("all_indices.csv", header = TRUE)

# subset variables that will be used for corr analyze
data <- dataAll[,c("Btot_kg","species","DBH_cm","Total_height_m","V_m3","Gi_m2",
                   "Gi_m2_ha","Gi.pr.sp.qd","simpson","shannon","Even","D","Mi","MS",
                   "W", "S", "A", "TH")]

names(data) <- c("B","species","d","h","V","Gi_m2","Gi_qd","Gi_pr_qd",
                 "Sm","Sn","E","D","Mi","MS","W","S","A","TH")

# subset data by species
data.pp  <- subset(data, species=="Pinus pinea")
data.jth <- subset(data, species=="Juniperus thurifera")
data.qf  <- subset(data, species=="Quercus faginea")
data.qI  <- subset(data, species=="Quercus ilex")
data.Qr  <- subset(data, species %in% c("Quercus faginea","Quercus ilex"))


# define variables with log transformation
logB  <- log(data.qf$B)
logGi <- log(data.qf$Gi_pr_qd)
```



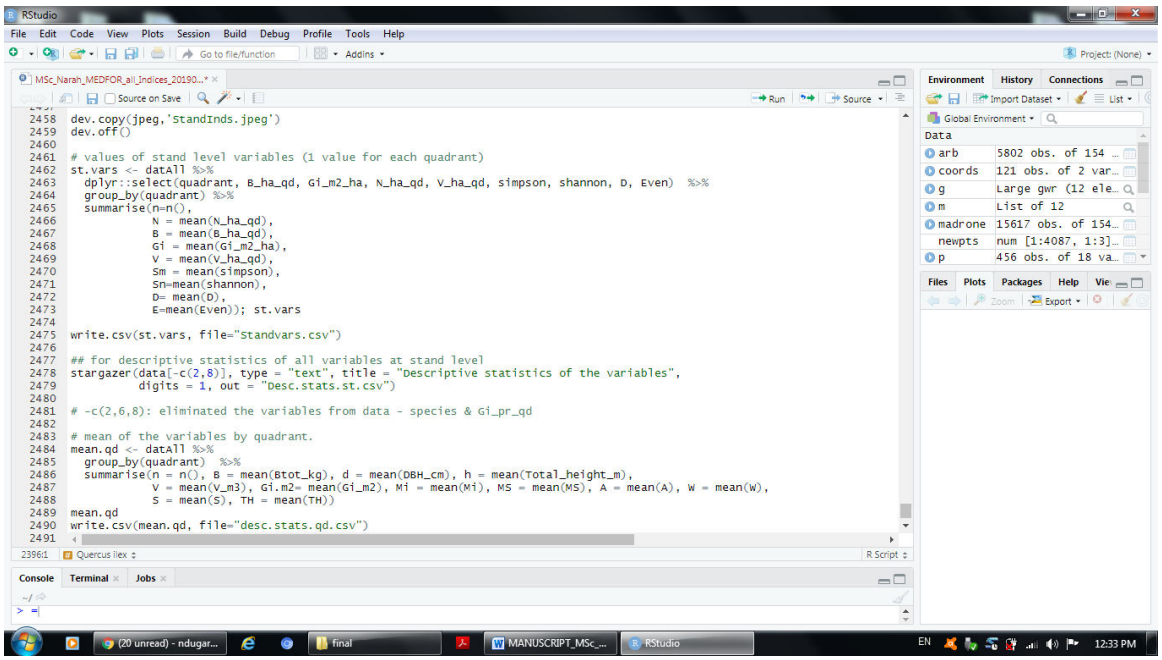
```
plot(loge+loge2, logB)
abline(m2)

# write.csv(sel.models, file = "selmods_qf.csv")

############################## Quercus ilex ########################################

packages <- c("MASS","dplyr","broom","purrr","tidyverse","stargazer","sjPlot","splines")
lapply(packages, require, character.only = TRUE)

setwd("D:/MEDFOR_classes/MScThesis_MEDFOR_UVa/data/processed")
dataAll <- read.csv("all_indices.csv", header = TRUE)

# subset variables that will be used for corr analyze
data <- dataAll[,c("Btot_kg","species","DBH_cm","Total_height_m","V_m3","Gi_m2",
                   "Gi_m2_ha","Gi.pr.sp.qd","simpson","shannon","even","D","Mi","MS",
                   "W", "S", "A", "TH")]

names(data) <- c("B","species","d","h","V","Gi_m2","Gi_qd","Gi_pr_qd",
                 "Sm","Sn","E","D","Mi","MS","W","S","A","TH")

# subset data by species
data.pp  <- subset(data, species=="Pinus pinea")
data.jth <- subset(data, species=="Juniperus thurifera")
data.qf  <- subset(data, species=="Quercus faginea")
data.qi  <- subset(data, species=="Quercus ilex")
data.Qr  <- subset(data, species %in% c("Quercus faginea","Quercus ilex"))


# define variables with log transformation
logB  <- log(data.qi$B)
logGi <- log(data.qi$Gi_pr_qd)
logd  <- log(data.qi$d)
```



```
################################## Quercus (q.faginea and q.ilex together) ##############

# define variables with log transformation
logB  <- log(data.Qr$B)
logGi <- log(data.Qr$Gi_pr_qd)
logd  <- log(data.Qr$d)
logh  <- log(data.Qr$h)
logv  <- log(data.Qr$V)
logSm <- log(data.Qr$Sm)
logSn <- log(data.Qr$Sn)
logE  <- log(data.Qr$E)
logD  <- log(data.Qr$D)
logMi <- log(data.Qr$Mi+1)
logMS <- log(data.Qr$MS+1)
logW  <- log(data.Qr$W+1)
logS  <- log(data.Qr$S)
logA  <- log(data.Qr$A)
logTH <- log(data.Qr$TH)

mods.Qr.list1 <- list(m1 <- lm(logB ~ logGi + logTH),
                      m2 <- lm(logB ~ logGi + logTH + logSm),
                      m3 <- lm(logB ~ logGi + logTH + logSn),
                      m4 <- lm(logB ~ logGi + logTH + logD),
                      m5 <- lm(logB ~ logGi + logTH + logE),
                      m6 <- lm(logB ~ logGi + logTH + logMi),
                      m7 <- lm(logB ~ logGi + logTH + logMS),
                      m8 <- lm(logB ~ logGi + logTH + logS),
                      m9 <- lm(logB ~ logGi + logTH + logA),
                      m10 <- lm(logB ~ logGi + logTH + logW),
                      m11 <- lm(logB ~ logGi + logA),
                      m12 <- lm(logB ~ logGi + logA + logSm),
                      m13 <- lm(logB ~ logGi + logA + logSn),
```



```
plot(logd+logD2, logB)
abline(tog,m2)


########################################## Summary table of variables ######################

packages <- c("MASS","dplyr","broom","purrr","tidyverse","stargazer","sjPlot","splines")
lapply(packages, require, character.only = TRUE)

setwd("D:/MEDFOR_classes/MScThesis_MEDFOR_UVa/data/processed")
datAll <- read.csv("all_indices.csv", header = TRUE)

# subset variables that will be used for corr analyze
data <- datAll[,c("Btot_kg","species","DBH_cm","Total_height_m","V_m3","Gi_m2",
                  "Gi_m2_ha","Gi.pr.sp.qd","simpson","shannon","Even","D","Mi","MS",
                  "W", "S", "A", "TH")]

names(data) <- c("B","species","d","h","V","Gi_m2","Gi_qd","Gi_pr_qd",
                 "Sm","Sn","E","D","Mi","MS","W","S","A","TH")

### Total Biomass B per ha (kg/ha) & Volume per ha (m3/ha), N tree per ha (trees/ha) by quadrant
datAll <- datAll %>%
  group_by(quadrant) %>%
  mutate(B_ha_qd = (sum(Btot_kg)/(1/16)), # Btot_kg is Biomass of whole tree
         V_ha_qd = (sum(V_m3)/(1/16)),
         N_ha_qd = n()/(1/16))


Dg.SDI <- datAll  %>%
  group_by(quadrant)  %>%
  mutate(Dg = (Gi_qd/N_ha_qd*(40000/3.1415))^0.5,    # quadratic mean diameter
         SDI = N_ha_qd*(25/Dg)^(-1.605))             # stand density index
```



```
dev.copy(jpeg,'StandInds.jpeg')
dev.off()

# values of stand level variables (1 value for each quadrant)
st.vars <- datAll %>%
  dplyr::select(quadrant, B_ha_qd, Gi_m2_ha, N_ha_qd, V_ha_qd, simpson, shannon, D, Even)  %>%
  group_by(quadrant) %>%
  summarise(n=n(),
            N = mean(N_ha_qd),
            B = mean(B_ha_qd),
            Gi = mean(Gi_m2_ha),
            V = mean(V_ha_qd),
            Sm = mean(simpson),
            Sn=mean(shannon),
            D= mean(D),
            E=mean(Even)); st.vars

write.csv(st.vars, file="Standvars.csv")

## for descriptive statistics of all variables at stand level
stargazer(data[-c(2,8)], type = "text", title = "Descriptive statistics of the variables",
          digits = 1, out = "Desc.stats.st.csv")

# -c(2,6,8): eliminated the variables from data - species & Gi_pr_qd

# mean of the variables by quadrant.
mean.qd <- datAll %>%
  group_by(quadrant)  %>%
  summarise(n = n(), B = mean(Btot_kg), d = mean(DBH_cm), h = mean(Total_height_m),
            V = mean(V_m3), Gi.m2= mean(Gi_m2), Mi = mean(Mi), MS = mean(MS), A = mean(A), W = mean(W),
            S = mean(S), TH = mean(TH))
mean.qd
write.csv(mean.qd, file="desc.stats.qd.csv")
```