# **Introduction**

# **Material and Methods**

## Collection methods, data sources

Our data are facts of species presence and abundance on various islands and the surrounding mainland. For all Aegean islands included in our analysis except Skyros we used published data, especially from a series of papers by Volker Assing on the staphylinid fauna of the islands Chios (Assing, 2015a, 2016a), Corfu (Assing et al., 2018), Crete (Assing, 2013a, 2015b), Cyprus (Assing, 2017b, 2017a; Assing and Wunderle, 2001), Ikaria (Assing, 2017c), Karpathos (Assing, 2016b), Kos (Assing, 2017a), Lesbos (Assing, 2005, 2016a, 2017a), Rhodes (Assing, 2013b, 2017a) and Samos (Assing, 2015c, 2016a, 2017a). All islands we used for our analyses have been marked on the map below (figure 1), where Skyros is marked with red. For Skyros we used specimens collected in during two trips, one in 2015 and one in 2019 (Other paper).

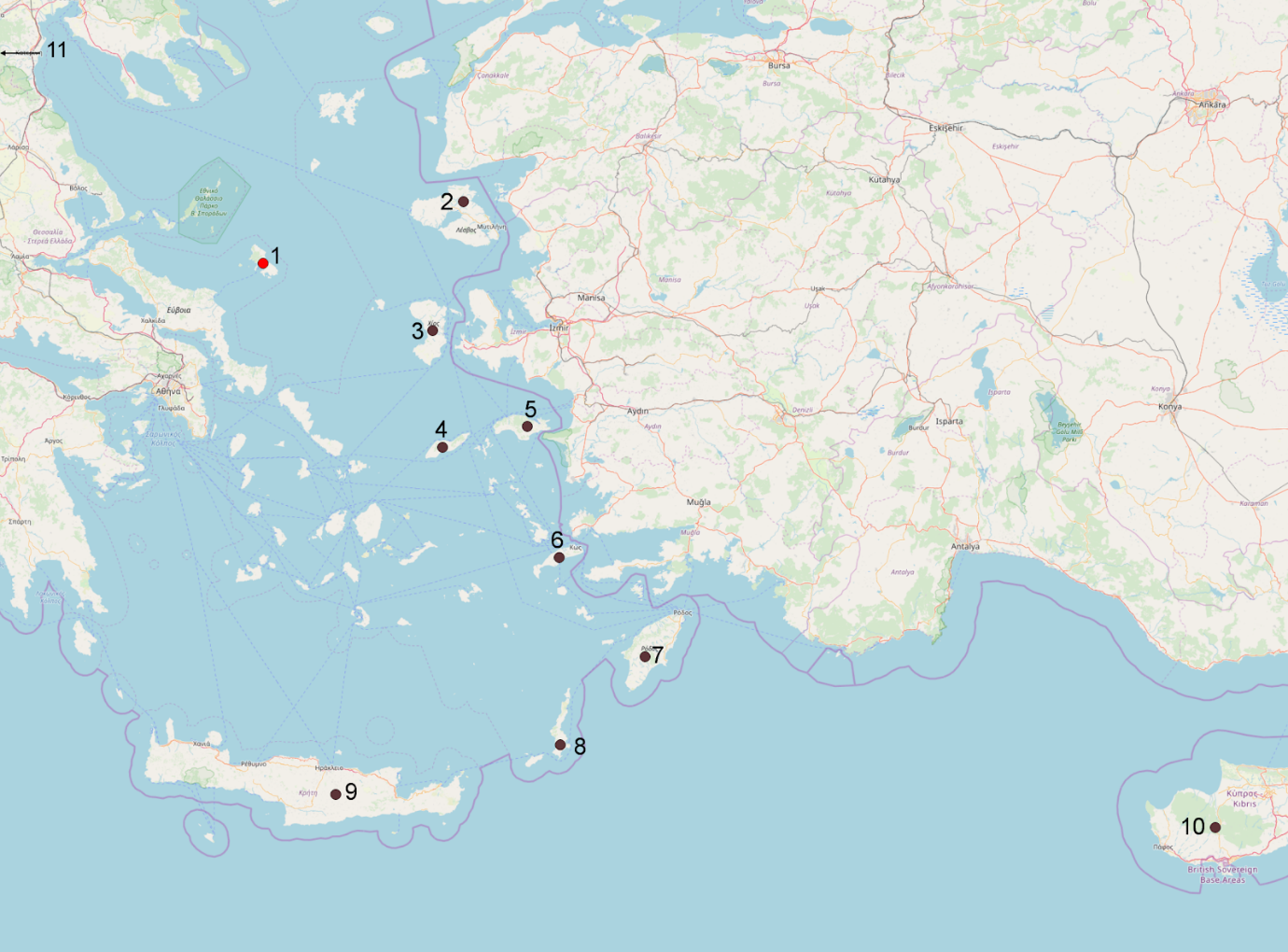


Figure 1. Aegean islands with data on Staphylinidae used for our analyses. Marked in red is our focal island Skyros (1). Marked in black are the islands Lesbos (2), Chios (3), Ikaria (4), Samos (5), Kos (6), Rhodos (7), Karpathos (8), Crete (9) and Cyprus (10). Corfu (11) is located on the opposite side of the greek mainland, and a direction towards its location has been indicated by an arrow.

## Rarefaction Curves

When we work with the species lists of any given islands, we assume they are representative of the actual fauna. In order to estimate whether this is actually true for each of the islands investigated, we use rarefaction curves that show whether the current sampling effort on the Aegean islands under consideration has been sufficient or whether some islands will need additional sampling before we can get something approaching a true species list. In relation to this, we will also attempt to compare the amount of work needed to get truly representative species lists for the islands. To do this, we will use rarefaction curves. They are a way to estimate the species richness of an area based on the current sampling effort and the absolute abundance of the collected species. These curves will follow a logarithmic curve, as fewer new species will be found with increased sampling effort as all of the most common species have already been found (Heck et al., 1975). The calculation of these rarefaction curves was carried out in R using the R package ‘sprex’ (Archer, 2016) with a newer algorithm developed in 2014. This algorithm estimates a lower bound of the species richness to prevent excessive overestimation of the number of species, in addition to extrapolating the potential known species richness resulting from some increased amount of effort (Chiu et al., 2014).

## Species richness and diversity

The core data for our species diversity analyses were the species composition of Staphylinidae on each of the compared islands. We defined our species richness (α-diversity) as the average number of staphylinid species found on a given island. β-diversity, an index which shows the compositional dissimilarity between the islands was then calculated using the Sørensen dissimilarity index (βsor) (Baselga, 2010). β-diversity is a measure of the difference in the species composition between communities, in this case the Aegean islands. It compares the species lists of various areas in a way that allows us to make a dissimilarity matrix showing all the pairwise comparisons between the islands. We then further divided the β-diversity into two components, the Simpson turnover (βsim) and the nestedness (βnes). The turnover describes the dissimilarity which is caused by species replacement, and thus it does not depend on the difference in the number of species (species richness) on the compared islands(Baselga, 2010). The nestedness describes the dissimilarity which is caused by the differences in species richness (Baselga, 2010). All these indexes were calculated using the ‘betapart’ R package (Baselga and Orme, 2012). We also calculated the ζ-diversity, which is a method for analysing the diversity of multi-site (2+) assemblies, as using only α- and β-diversity are insufficient for a complete view of the intersite diversity (Hui and McGeoch, 2014).  ζ-diversity allows us to explore how the incidence-based composition of the sites changes with different scales and different numbers of sites involved in the comparison. Different ζ-orders, ζi, indicates the mean number of species shared by i sites. Thus, we can estimate the amount of common and widespread (shared by several islands) and rare species within our study area on any particular island. Additionally, we will make some different estimates of both α- and β-diversity, as ζ1 (the average amount of species found on a given island) will be an estimate of α-diversity while ζ2 (the amount of species shared among two given islands) will be an estimate of β-diversity. We use the ζ-diversity to quantify the relative turnover rate of rare versus common species (Latombe et al., 2017). To calculate ζ-diversity we used the ‘zetadiv’ R package (Latombe et al., 2018).

## Hierarchical clustering analysis

To infer factors affecting species composition of the islands, we first need to compare their fauna. To achieve this we computed a dissimilarity index using the ‘vegan’ R package (Oksanen et al., 2019) which, among other things, measures the difference in species composition among the islands. We then used UPGMA (Unweighted Pair Group Method with Arithmetic mean) (Sokal and Michener, 1958), a very common form of hierarchical clustering in ecology as well as other areas of biology. This method is included in the R-native ‘stats’ package and we used the hclust command(R Core Team, 2019). We used presence/absence data for all islands in our species list and made a dendrogram from a Jaccard dissimilarity matrix. We also made another dendrogram from the Simpson turnover matrix which we have computed using the ‘betapart’ R package (Baselga and Orme, 2012).

## Generalized Linear Models (GLM)

We used Generalized Linear Models to analyse which factors among the current geographic and climatic conditions of the islands (European Environment Agency (EEA), 2019) (Fick and Hijmans, 2017) were the greatest contributors to staphylinid species richness in the region. In our measure of species richness, we have also included currently unidentified morphospecies. We compared the area in km2 of the islands, the relative amount of forest on a given island (Others/Forest), the precipitation, the mean diurnal temperature variation as well as the average annual temperature variation. We have not included the amount of phrygana in the Others/Forest factor, as it is by far the most dominant vegetation type(>>50%) on all islands and would reduce the explanatory ability of this factor because it would minimize the information included in our model. As the amount of phrygana can be inferred when both the amount of forest and other vegetation is known, we do not believe that we reduce the amount of information fed to the GLM. The higher this factor is, the more forest is present on the island compared to other, non-phrygana vegetation. We also considered the distance of a given island to the nearest mainland (DCM). These various parameters are shown in table 1. Finally, we calculated an additional factor pertaining to our forest factor, the annual precipitation (OFxP) multiplied with the others/forest factor, as these parameters seem to interact to predict the type of forest present. Areas with less moisture will have a different type of forest than ones which are wetter, and this leads to an interaction between the factors. While not included in the table below, it is easily calculated from the data which have been included. To analyse our GLM we used the dredge function from the ‘MuMIn’ R package (Barton, 2018). Using the dredge function, we generated information criteria to investigate which model, and thus which parameters, performed the best. The main information criteria which we used was the Akaike Information Criterion (AICc) (Akaike, 1974) which was developed to help model selection by investigating how well the specific model explains our data relative to the other models. While this number is not particularly useful in itself, comparing it to the values of other models in our GLM gives us a comparative method to investigate the best model - the lower the AICc is, the better it explains our data.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Species | Area of island | Distance to mainland | Annual Precipitation | Mean Diurnal Range | Annual Temperature range | Other Vegetation/ Forest | Endemic species |
| Skyros | 65 | 209 | 38 | 425.9 | 6.17 | 22.13 | 0.199 | 1 |
| Ikaria | 69 | 257.96 | 57 | 574 | 7.66 | 23.75 | 0.043 | 5 |
| Kos | 55 | 289.06 | 5.5 | 740 | 8.14 | 23.98 | 1.865 | 0 |
| Karpathos | 61 | 301 | 97 | 678 | 6.59 | 20.77 | 0.457 | 2 |
| Samos | 157 | 477.4 | 1.6 | 714.7 | 8.33 | 25.1 | 0.460 | 9 |
| Corfu | 236 | 610.9 | 4.7 | 1097.3 | 8.41 | 23.61 | 4.489 | 14 |
| Chios | 42 | 851.13 | 6.8 | 520 | 8.25 | 25.11 | 0.857 | 0 |
| Rhodes | 123 | 1400 | 18.5 | 703 | 8.1 | 23.39 | 0.412 | 9 |
| Lesbos | 200 | 1633 | 9.5 | 670.6 | 8.77 | 26.7 | 0.156 | 8 |
| Crete | 171 | 8303 | 100 | 483.2 | 7.86 | 22.87 | 1.779 | 69 |
| Cyprus | 148 | 9251 | 214 | 524.8 | 10.65 | 27.15 | 1.141 | 14 |

Table 1. An overview of the various factors included in our analyses. Included are two geographic variables (e.g. Area of the island and Distance to mainland), three climactic variables (e.g. Annual Precipitation, Mean Diurnal Range and Annual Temperature range) as well as a biological variable (The amount of other vegetation, not including phrygana, compared to the amount of forest on a given island). Finally, we also show the species richness, which we want our model to predict.

# **Results**

## Rarefaction curves

Using our species abundance lists we have generated rarefaction curves for Crete, Chios, Ikaria, Karpathos, Kos, Lesbos, Rhodes, Samos and Skyros, and plotted them in figure 2. We can see that some islands, especially Chios and Skyros seem to be relatively well sampled as they are approaching the horizontal asymptote. Additionally, these islands’ confidence intervals, which we have plotted as a dashed line, are quite close to our estimates. The remaining islands, especially Ikaria, Kos and Rhodes have either not quite reached their horizontal asymptote or have very wide confidence intervals indicating a large amount of uncertainty. A more easily comparative chart, which shows the rarefaction curves for every island in the same graph, is provided in figure 3.

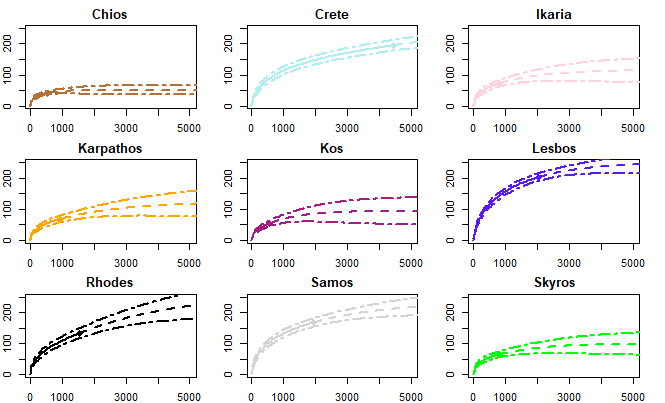


Figure 2. Rarefaction curves with error curves for the investigated Aegean Islands. We have plotted the number of species found (y-axis) against the number of individuals sampled (x-axis) and generated an extrapolation from this as well as error curves.

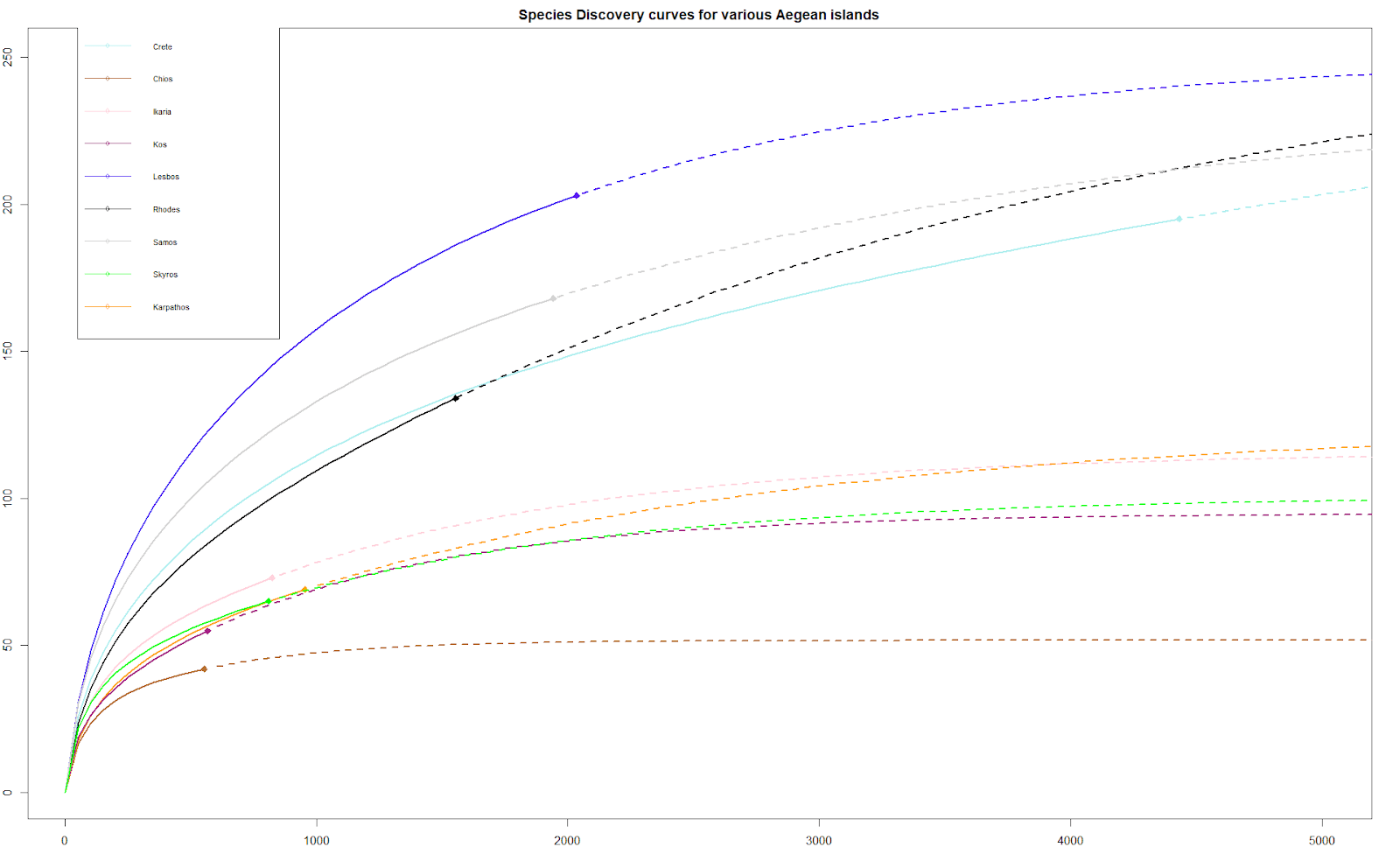
****

Figure 3. Rarefaction curves for all investigated Aegean Islands. This comparative chart shows the approximate species richness of all the islands. We have again plotted the number of species found (y-axis) against the number of sampled individuals (x-axis), but we have excluded the confidence intervals for ease of viewing.

## Species richness and diversity

Using our species list, we have calculated various richness and diversity measures. **B-diversity??**

When looking at the ζ-diversity for the islands (again excluding Corfu as it is an outlier in every metric), it becomes clear that very few species are widespread among all 10 considered islands, and almost no species are present everywhere (ζ10=1). This number might not be precise, as the species inventories on many islands are probably incomplete as shown in our rarefaction curves. Still it should give a somewhat representative view of the species dynamics, e.g. the rarity of a given found species in the region, as the most common and thus most widespread species should have been found in the early sampling efforts. Thus we see many uncommon or rare species (ζ1=100.1,ζ2=25.56) with a rapid decline in ζ-diversity when looking for more widespread species (ζ-order>2, see fig. 5). This suggests a difference in the staphylinid fauna development of the islands, and as such either different sources for colonisation by the staphylinids or different climactic/biotic niches present on the islands which will affect which species can be present on a given island.

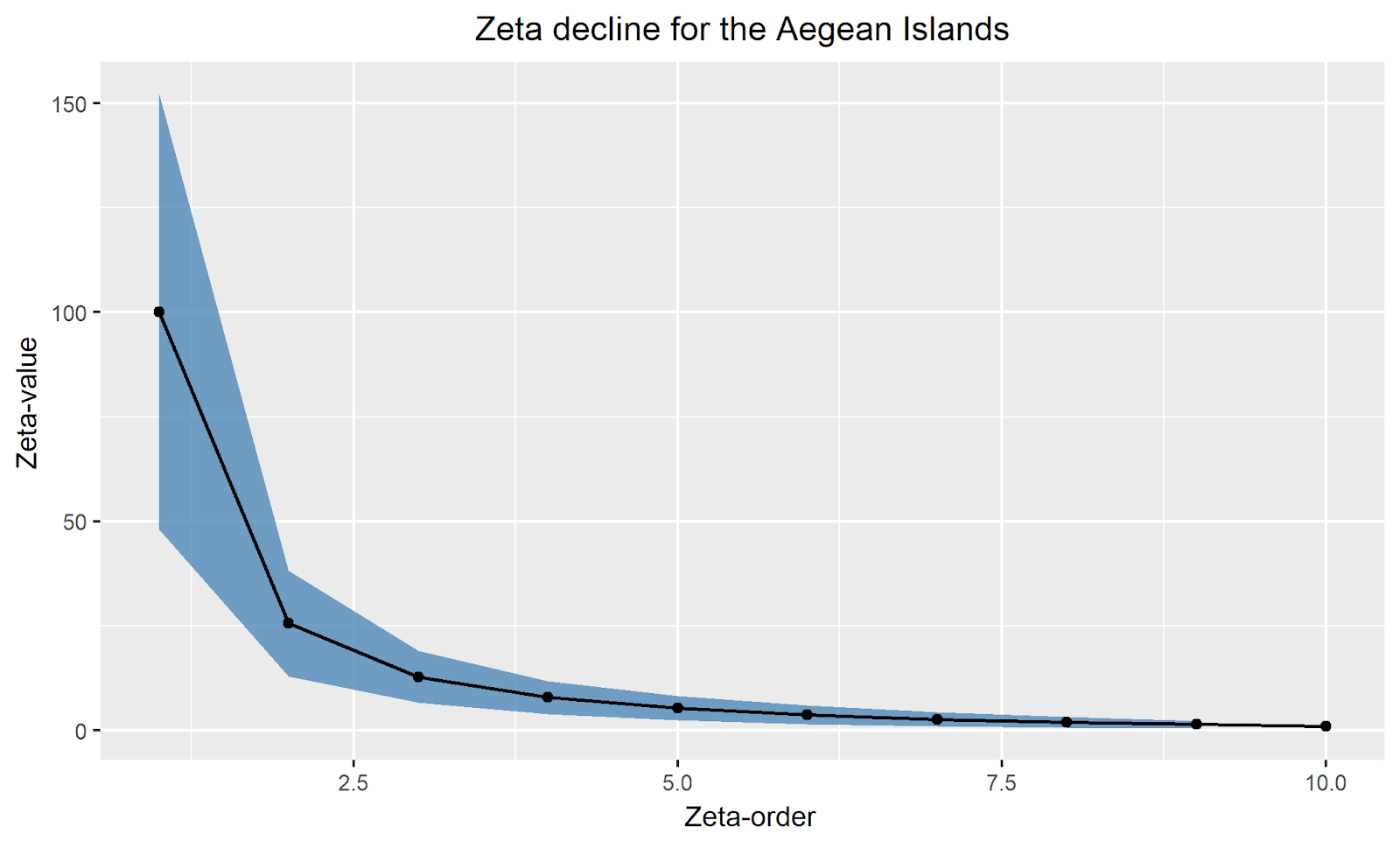


Figure 5. Zeta decline for the chosen Aegean Islands. Here we have calculated and plotted the Zeta-value on the x-axis, e.g the mean number of species in i localities, against the Zeta-order on the y-axis, e.g the number of localities compared.

## Hierarchical clustering

Using our Sørensen Dissimilarity matrices we have made a hierarchical clustering analysis on our presence/absence data and illustrated this in a dendrogram (figure 4). Here we see several clusters corresponding to the mutual location of the islands in the Aegean. Skyros is a sister group to a cluster formed by all the other islands, presumably because it is the only island in the western Aegean closest to the Greek mainland. We also see that Cyprus and Crete cluster together, which we will refer to as the southern cluster, and they are the largest and the southernmost islands in our study in addition to also being some of the islands which are the furthest away from the mainland. The remaining islands also cluster together, yet this cluster is split in two: one is composed of Rhodes, Samos and Lesbos and another composed of Ikaria, Chios, Kos and Karpathos. We will refer to this big cluster as the eastern cluster. When making a hierarchical clustering analysis using turnover instead, which helps correct for varying sample size, we see the same overall clusters, although Cyprus and Crete are more separate now. Here we no longer see the largest islands cluster together in the eastern cluster, and it is instead separated into three subclusters - one consisting of Rhodes and Karpathos (subcluster a), one of Samos and Chios (subcluster b), one of Lesbos and Kos (subcluster c) and with Ikaria being outside a clear cluster.

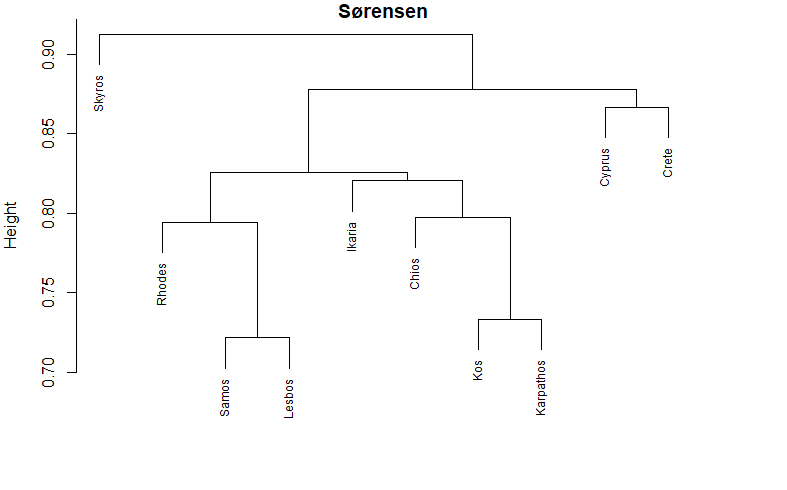


Figure 6. Hierarchical clustering analysis of the Aegean islands using Sørensens Dissimilarity. Here we see the islands clustered using UPGMA on our species occurrence lists. Several clusters are apparent, which are mostly geographically determined. The clusters containing Rhodes, Samos, Lesbos, Ikaria, Chios, Kos and Karpathos does not quite fit the geography of the islands, though they are all close to Turkey.

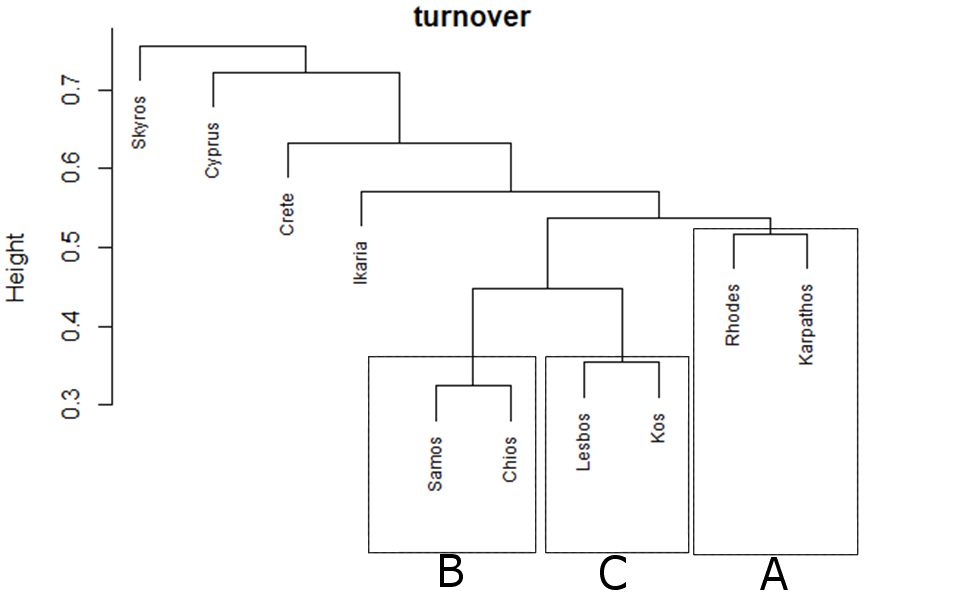
****

Figure 7. Hierarchical clustering analysis of the Aegean islands using a Simpson Turnover matrix. Here we see the islands clustered using UPGMA on our species occurrence lists. We see several apparent clusters which are mostly geographically determined.

## Generalized Linear Models

Using the dredge function in R, we calculated the Akaike Information Criterion correlated for a small sample size (AICc). A list of the four best explanatory models are shown in table 2. From this we gather that the model which best explains the Aegean species richness is one of the form:

*Species = 3.1705 + Area + Distance to Closest Mainland + Annual Precipitation + Others/Forest + 1*

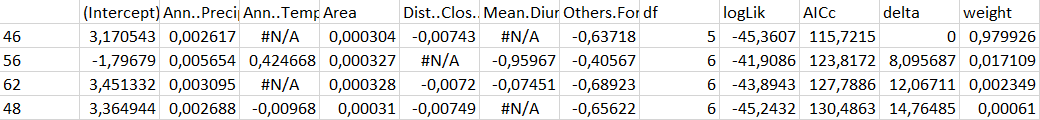
****

Table 3. Here our four best models are shown. The first column shows numbers indicating the different models. The following nine columns show our different factors, with the ones excluded by a given model being marked #N/A. The last five columns show us information about how well the model performed, and we see both the degrees of freedom in the model as well as the log likelihood, the Akaike Information Criterion (AICc), the delta value and the model weight.

Additionally we see that only the top model has a substantial level of empirical support as none of the other models has a delta value below two, which has been suggested as an indicator for the amount of support compared to the other models by Burnham and Anderson (2002). The evidence ratio between the two top models is 0.9799/0.0171 = 57.28 suggests a high amount of support for the top model. This still suggests that both the area of the island, the distance to the closest mainland, the amount of annual precipitation as well as the relative amount of forest and other vegetation all influence the species richness.

The percentage deviance explained by this best model is 90.2%, which is very high. The variables output by the top model had a significant explanatory effect on the species richness (p<0.001, see table 3).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | Coefficient (+- SE) | Upper 95% CI bound | Lower 95% CI bound | Z value |
| Intercept | 3.1706 (+-0.2699) | 3.338 | 3.0034 | 11.75\*\*\* |
| Precip | 0.0026 (+-0.000397) | 0.0029 | 0.0024 | 6.59\*\*\* |
| Area | 0.0003 (+-0.00002) | 0.00031 | 0.00029 | 14.29\*\*\* |
| DCM | -0.0074 (+-0.00086) | -0.0069 | -0.0079 | -8.67\*\*\* |
| Others/Forest | -0.6372 (+-0.0723) | -0.5924 | -0.682 | -8.82\*\*\* |

\*\*\**p<0.001*

Table 4. Overview of the output of our GLM. Precip is the annual precipitation, Area is the area of a given island in square kilometers, DCM is the distance to the closest mainland and others/forest is the relative amount of forest on a given island.