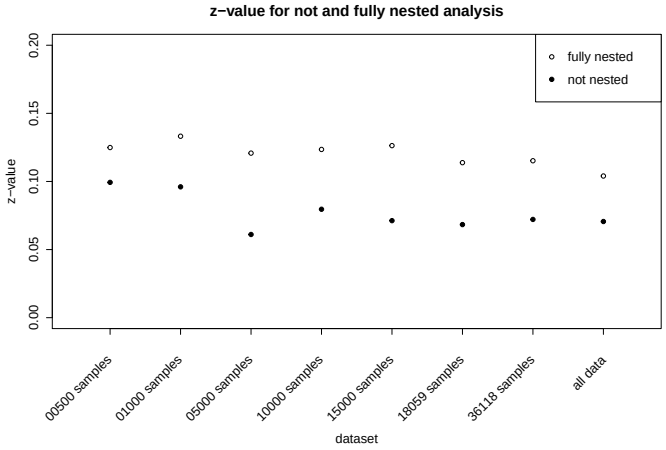
A statistical approach to species-area relationships for range data

Karina Reckling

Bachelor thesis (Student ID3504252)
submitted to
the Faculty of Environment & Natural Resources
at the Albert-Ludwigs-University Freiburg



Supervisor: Prof. Dr. Carsten F. Dormann,
Department of Biometry and Environmental System Analysis
Co-Supervisor: Dr. Michael Staab,
Chair of Nature Conservation and Landscape Ecology

Freiburg, December 2015

Abstract

This thesis is analyzing the scale-dependence of the z-values of the species-area relationship (SAR). The increase in the number of species when adding of area is considered a general law in ecology. Within the discussion of SAR the terms species richness, species density and species diversity are used interchangeably, although their meanings as defined by Whittaker vary. Richness refers to the number of species per sample, whereas density refers to the number of species per sampling area. Diversity, meanwhile is divided into inventory diversity and differentiation diversity. A common interpretation of the differentiation diversity is the beta diversity, which is considered to be the rate of rise of the SAR linear relationship in logarithmic space, the z-value. The scale-dependence of this parameter is debated intensely. Most studies, such as this one, use a log-log transformation of the power law as introduced by Arrhenius in 1921: $log(S) = log(c) + z * log(A)$ This mathematical function was applied on the rodents data of the Terrestrial Species dataset of the International Union for Conservation of Nature (IUCN) for different levels of nestedness. Grid sizes of 50 km by 50 km, 100 km by 100 km, 200 km by 200 km, 500 km by 500 km and 1000 km by 1000 km and the MCMCglmm analytical model of the R statistical software was used. Also subsets of 500, 1000, 5000, 10000, 15000, 18059 and 36118 random samples were analyzed. The results of this process showed little variation within the different levels of nesting, for which the z-values remain in the magnitude of 0.1. A difference of one magnitude was detected for z-values of analysis without any nesting; here they were at 0.01 magnitude. For the c-values little variation was found. This difference in magnitude shows the scale-dependence of z-values. Since biological consequences and the thus missing linearity are not studied here, further reading and research is encouraged.

Zusammenfassung

Die species-area relationship (SAR) wird generell als grundlegend für die Ökologie anerkannt. Es handelt sich dabei um die Zunahme der Speziesanzahl bei Vergrößerung der Untersuchungsfläche. Dieser Zusammenhang wird durch eine Vielzahl an Faktoren auf verschiedenen Ebenen beeinflußt. Neben dem generellen Konsens über die Existenz der SAR herrscht große Diskussion über die Maßstabsabhängigkeit der z-values. Dieser Wert gibt die Steigung der linearen Approximation im log-log plot an und wird mit der beta diversity in Zusammenhang gebracht, welche die Veränderung der Spezieszusammenstellung bezeichnet. In dieser Arbeit wird die Maßstabsabhängigkeit der z-values untersucht. Die Analyse basiert auf der logarithmus-transformierten von Arrhenius in 1921 vorgeschlagenen und seither oft verwendeten Exponentialfunktion $log(S) = log(c) + z * log(A)$. Diese ermöglicht eine lineare Regression der arithmetischen Form $S = cA^z$, welche eine Parabel beschreibt. Mithilfe der MCMCglmm-Regression in R wurden die Daten der International Union for Conservation of Nature (IUCN) zur Verbreitung von Nagetieren weltweit gefittet. Grids mit 50 km, 100 km, 200 km, 500 km und 1000 km Seitenlänge wurden genutzt, um die Maßstabsabhängigkeit zu testen. Zudem wurden Teilmengen von 500, 1000, 5000, 10000, 15000, 18059 und 36118 zufällig gezogener Stichproben gefittet. Die c-values für alle Analysen liegen in der 1.1 Größenordnung. Die z-values hingegen sind eine Größenordnung kleiner für nicht genestete Analysen verglichen mit genesteten (0.01 anstatt 0.1). Innerhalb der verschiedenen Tiefen der genesteten Analysen gibt es jedoch kaum Unterschiede. Somit lässt sich erschließen, dass die z-values maßstabsabhängig sind. Wie tief genested wird, ist jedoch wenig ausschlaggebend. Dies führt zu Konsequenzen in der biologischen Interpretation und zu Zweifeln an der Verwendung der linearen Approximation. Generell jedoch sind die gewonnenen z-values nicht untypisch, wenn auch leicht geringer als der typische Rahmen von 0.2 bis 0.4. Dies ist ein Indiz für die Validitätsverlässlichkeit der Vorgehensweise. Zudem ist zu erkennen, dass Teilmengen bis zur Hälfte der Gesamtmenge der Stichproben nicht aussagekräftig genug sind, um z-values korrekt zu bestimmen. Alles in allem ist ein Maßstabseffekt festzustellen. Die terminologische Uneinigkeit kann zudem zu einer Verkomplizierung der Probennahme und fehlerhaften Interpretation führen. Teilmengen der Stichproben sind nicht aussagekräftig. Schlussendlich ist eine weitere Recherche in den Bereichen sigmodial beziehungsweise triphasic SAR und direkter Probennahme anstelle von Verbreitungsgebietskarten wünschenswert.

Contents

List of Figures

List of Tables

List of Listings

Chapter 1

Introduction

As stated in many publications, the species-area relationship (SAR) is seen as a fundamental law in ecology [Lomolino, 2001, Scheiner, 2003] and as one of the first and most general ecological patterns. [Ladle and Whittaker, 2011] The idea of an increase in the number of species with added area is generally accepted, although details vary. [Drakare et al., 2006, Lomolino, 2001, Rosenzweig, 1995, Scheiner, 2003] Within this consensus, the synonymous use of the terms species richness, species density, and species diversity is leading to confusion and debate. In this thesis, however, scale is very important, thus the definitions of Whittaker need to be mentioned, as he distinguishes between richness and density by area. The former is the number of species per sample, whereas the latter refers to the number per area unit. Since in the study undertaken all sample areas were given specific contiguous dimensions, the terms species richness and species density can be used interchangeably. Species diversity, on the other hand, is defined as a index weighting richness and the "evenness of abundance across species". [Whittaker et al., 2001] Diversity is further generally accepted to refer to several scales as point, alpha (local), gamma (regional) and epsilon (landscape) diversity, which are generally known as inventory diversity. Seldomly mentioned is the differentiation diversity, which denominates the turnover of species or the change in the species composition. Generally, only beta diversity is used and refers to the turnover between alpha and gamma diversity. There is also internal beta diversity (between point and alpha) and delta diversity (between gamma and epsilon). The complexity of species diversity varies for each publication, exemplifying the heterogeneity in the field. [Arita and Rodrguez, 2002, Gray et al., 2004, Scheiner, 2003, Scheiner, 2004, Whittaker et al., 2001]

1.1 Curve types

In contrast to the large consensus shown before that only included minor disagreements, the discussion on the shape that the curve exhibits is more comprehensive. In his 2003 publication, Scheiner classified species-area relationships into six types, depending on the underlying sampling design. This classification is widely referred to and adapted from. His Type I curve is based on a nested sampling approach of single data points. Types II and III are samples of contiguous and non-contiguous grids respectively, each with two versions - Version A for spatially explicit sampling, and a Version B for sampling without spatial reference. Islands, as a special case, are classified as Type IV SARs. In Figure 1.1, visualizations of this classification are provided. For all types, he considered the intercept to be the alpha diversity and the slope (in 2004 corrected to rate of rise) to be the beta diversity. Appointing the term species-area relationship to all six types is very debated, and several authors adapted the classification to their own ideas. Gray, for example, only accepted Type IV as a species-area relationship, and changed Type I and Types II and III to species-accumulation and species-rarefaction curves respectively. Hereby, the

number of species is not plotted in respect to sample area directly, but to the accumulation of area with increased sampling effort. Dengler adds the term species-sampling relationship (SSR) and leaves out the species-accumulation relationship, even though his new term also separates SAR and SSR by the use of contiguous area, which is only given in species-area relationships. [Whittaker and Fernandez-Palacios, 2007, Dengler, 2009, Gray et al., 2004, Lomolino, 2001, Scheiner, 2003, Scheiner, 2004]

In accordance with the debate on which shape the curve should have, there is further debate on the scale-dependence of the species-area relationship. Differences in scale are explained by factors and processes, which affect the SAR at different scales. Area is generally accepted as the most important factor, with residuals being explained by habitat diversity, latitudinal gradient, isolation, extinction risk, elevation or altitude, specialization, immigration, disturbances, sampling, energy and resources, and other factors. A variety of hypotheses are also applied on different scales, complicating the underlying simple relationship of the number of species to area. For further details on the hypotheses and factors involved in shaping the species-area relationship, I suggest the study of the cited publications, such as Stiles, which deal with this topic in greater depth. [Whittaker and Fernandez-Palacios, 2007, Connor and McCoy, 1979, Drakare et al., 2006, Lomolino, 2001, Gerstner et al., 2014, Preston, 1960, Rosenzweig, 1995, Scheiner, 2004, Stiles and Scheiner, 2010, Ladle and Whittaker, 2011]

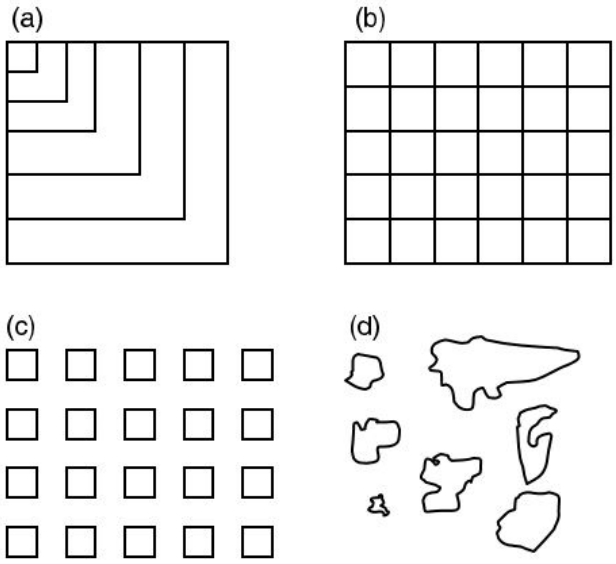


Fig. 1 Species-area curves can be built from four general sampling schemes: (a) strictly nested quadrats (Type I curves); (b) quadrats arrayed in a contiguous grid (Type II curves); (c) quadrats arrayed in a regular but noncontiguous grid (Type III curves); or (d) areas of varying size, often islands (Type IV curves).

Figure 1.1: Scheiners classification of species-area relationships (Scheiner2003)

1.2 Mathematical model

In the 1920s, the first and still most used mathematical models for the species-area relationship were published by Arrhenius (1921) and Gleason (1922). Although there are many other models available, Arrhenius' power law is still recommended as the best model with two parameters. Its formula in arithmetic space $S = cA^z$ is log-log-transformed to $log(S) = log(c) + z * log(A)$. This transformation resulted in the linearization of the parabola in arithmetic space and the normalization of the residuals. Additionally, the fitted z and c-values approximated the actual ones. Lomolino preferred this approach since the c-values can be seen as the initial trajectory, and the z-values as the rate of change of this trajectory, with both features seen as necessary for an ideal model. Often neglected, due to the separated views on c and z, is the fact that the arithmetic slope is determined by both c and z together, and the z-value only being the slops in log-log space. Furthermore, Arrhenius never implied any biological significance to the parameters; he considered them arbitrary. Only in more recent years was a biological implication added and z-values came to represent the turnover rate, the beta diversity. [Connor and McCoy, 1979, Crawley and Harral, 2001, Dengler, 2009, Lomolino, 2001, Preston, 1960, Rosenzweig, 1995, Scheiner, 2003, Ladle and Whittaker, 2011]

Since this time, many analyses have evolved around z-values. One of which is on its scale-dependence. Considering this and the broad discussion on the effect of scale, and the handling of area within the sampling design, this thesis analyzes the scale-dependence of the z-values. By using rodent distribution maps, the species density for each grid cell of a given area is obtained. This information for the nested approach of different grid sizes will be used for scale-dependence analysis.

Chapter 2

Methods

For this analysis, the r packages *maptools* version 0.8-36 [Bivand and Lewin-Koh, 2015] and *rgdal* version 1.0-7 [Bivand et al., 2015] were used for loading and preparing the shapefile. With *raster* version 2.4-20 [Hijmans, 2015] and *MCMCglmm* version 2.22 [Hadfield, 2010], the raster preparation and analysis were done. The manually specified function *extract.species* by Dormann as it was applied is attached in Appendix B. The full R script can be found in Appendix C.

2.1 Data preparation

The data used for this analysis was the Terrestrial Mammals shapefile of the International Union for Conservation of Nature (IUCN). [IUCN, 2014] The projection was transformed to the equal-area, pseudo-cylindrical Mollweide one, so that spatial operation can be executed without any distortions. With the *extract.species* function, the spatial information on all rodent species was extracted. Additionally, the data was rasterized to a 50 km by 50 km grid. To achieve a nested design, this raster was aggregated to 100 km, 200 km, 500 km and 1000 km side lengths, resulting in five separate but nested files. Subsequently, the data was extracted into data frames for all grid sizes respectively. The values extracted are the number of the last polygon, which was found in the raster cell. Or, for the aggregations, the mean of these numbers. Since only the occurrence of a species is relevant, the mean function could be used. So only if there is no species in any of the 50 km by 50 km cells, the NA value is transferred to the aggregated cell. To make further data use more comprehensible, all NA values were replaced by 0, and all cells with a number (symbolizing species occurrence) by 1. At this stage, data frames for each grid size are achieved including x and y coordinates of the center of the respective raster cell and one column per species with a 0 or 1 value, referring to this species occurrence in the cell. Accordingly, identification columns were added to refer to the smaller cells a larger cell consists of. To do this, a 50 km by 50 km raster with running numbers was created and aggregated. This time the minimum value of the subset cells is used to keep the indices small. The indices were added as new columns to the data frames, whereby one column was appointed for each grid size. So at the data frame for the 100 km by 100 km grid size, the 'ID100' column consists of subsequent numbers, and all identification columns for smaller grid sizes (ID50) consist of NAs. For larger grid sizes (ID200, ID500, ID1000), the raster files were overlaid to appoint the same value to the four 100 km by 100 km cells contributing to the respective larger cell. Only for the 'ID500' column of the 200 km by 200 km grid size data frame an exception was made. Here NA was appointed due to the fact that it would take two and a half cell in each direction to aggregate to the 500 km by 500 km grid. (This rather complex approach consumed much time and is probably not yet perfectly solved. An improvement is desirable for any further study.) To be able to compute the species-area relationship as log(species) to log(area), a summation of the species occurring in each cell and a reference to the area occupied

by it is needed. Consequently, a sum and a resolution (short res) column were added. The sum of all species occurrences was applied to the sum column and the side length of the grid to the res one. Finally, the separate columns for each species were discarded and the single data frames combined into one, which includes all grid sizes. Considering that only the cells in which any species occurs is needed for the analysis and to reduce file size, all cells (represented by rows) with a sum value of 0 were discarded as well. The final data frame consists of columns for x and y coordinates of the center of the cell, an 'ID' column for each grid size and a 'sum' and a 'res' column referring to the amount of species within the cell and its side length. Each row represents a single cell. Table 2.1 shows selected rows of the final data frame.

Additionally to the general dataset subsets of 500, 1000, 5000, 10000, 15000, 18059 and 36118, randomly selected rows (thus raster cells) of the final data frame were generated. 36118 samples are half and 18059 are one-fourth of the total number of rows. The smaller the grid size, the more likely it is that a row of that particular grid is selected for such a subset, due to the fact that they provide more cells (and thus rows). The intention is to simulate smaller sampling effort.

Table 2.1: Selected rows of the final data frame

	x	y	ID50	ID100	ID200	ID500	ID1000	sum	res
1	-890447.7	8744750	3886	878	262	34	17	1	50
2	-790225.2	8744750	3888	879	263	34	17	1	50
3	-740114.0	8744750	3889	880	263	34	17	1	50
4	-1191115.1	8694696	4590	1230	261	33	17	1	50
5	-1141003.9	8694696	4591	1231	261	34	17	1	50
52819	-915503.3	8769777	NA	878	262	34	17	1	100
52820	-815280.8	8769777	NA	879	263	34	17	1	100
52821	-715058.4	8769777	NA	880	263	34	17	1	100
52822	-1717283.2	8669669	NA	1225	258	32	16	1	100
52823	-1617060.7	8669669	NA	1226	259	33	17	1	100
52824	-1516838.2	8669669	NA	1227	259	33	17	1	100
67130	-1767394.4	8719723	NA	NA	258	NA	17	1	200
67131	-1566949.4	8719723	NA	NA	259	NA	17	1	200
67132	-1366504.5	8719723	NA	NA	260	NA	17	1	200
67133	-1166059.5	8719723	NA	NA	261	NA	17	1	200
67134	-965614.6	8719723	NA	NA	262	NA	17	1	200
67135	-765169.6	8719723	NA	NA	263	NA	18	1	200
71169	-2919953.0	8769777	NA	NA	NA	30	15	1	500
71170	-1917728.1	8769777	NA	NA	NA	32	16	1	500
71171	-1416615.7	8769777	NA	NA	NA	33	17	1	500
71172	-915503.3	8769777	NA	NA	NA	34	17	1	500
71173	-414390.9	8769777	NA	NA	NA	35	18	1	500
71174	-5425515.0	8269235	NA	NA	NA	96	13	1	500
71978	-5174958.8	8519506	NA	NA	NA	NA	13	2	1000
71979	-4172734.0	8519506	NA	NA	NA	NA	14	2	1000
71980	-3170509.2	8519506	NA	NA	NA	NA	15	1	1000
71981	-2168284.3	8519506	NA	NA	NA	NA	16	1	1000
71982	-1166059.5	8519506	NA	NA	NA	NA	17	1	1000

2.2 Analysis

Initially, a fit for the species-area relationship with Linear Mixed Effect Models (lmer) was attempted. This approach did not give any results due to computational capacity problems, so one with Markov chain Monte Carlo Sampler for Multivariate Generalised Linear Mixed Models (MCMCglmm) was used instead. The different level of nesting were included as random effects. The deepness reached 50 km, 100 km, 200 km and 500 km respectively. Also, one fit was run with no random variable, which represents the lack of any application of nesting to the data analysis. An example of an output from a summary of such a fit is shown in Listing 2.1. Hereby, the Intercept and $\mathbf{log}(res2)$ equal the c and z-value respectively.

Listing 2.1: Example of a summary of a MCMCglmm fitting

```
¿ summary(fm contiguous0)

Iterations = 3001:12991
Thinning interval  = 10
Sample size  = 1000

DIC: 157347.9

R-structure:   units

post.mean l-95% CI u-95% CI eff.samp
units    0.5169   0.5115   0.5219     1000

Location effects: log(sum)   log(res 2 )

post.mean l-95% CI u-95% CI eff.samp  pMCMC
(Intercept)   1.92032  1.87362  1.96644     1000 ¡0.001 ***
log(res 2 )    0.07060  0.06509  0.07599     1000 ¡0.001 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .
    0.1         1
```

Chapter 3

Results

C and z-values were obtained after a MCMCglmm fit of the data frame of which a shorted version is shown in Table 2.1. The data frame is the result of overlaying the range size polygons of the IUCN with rasters of different grid sizes. Since only the c and z-values of the fit are important for this analysis, Table 3.1 summarizes those values for all datasets. The level of nestedness is reported by the extension of the data frame title; .0 represents no, and .1 to .4 increasing nesting. It reaches a maximum in the 500 km by 500 km, 200 km by 200 km, 100 km by 100 km and 50 km by 50 km grid cell size for the .1 to .4 extensions respectively. Considering that the power function is determined by both variables c and z, both are listed in the table. The c-value represents the intercept of the species-area relationship in log-log space and the z-value the rate of rise for the linear fit in log-log space.

When fitting all data without nesting, c is 1.92032 and z is 0.07060. If nested to 500 km grid side length, c is 1.5021 and z is 0.1087. For the nesting to 200 km, 100 km and 50 km grid side lengths,

Table 3.1: List of all c and z-values

	c	z		c	z
AllData.0	1.92032	0.07060	Sample10000.0	1.83780	0.07960
AllData.1	1.5021	0.1087	Sample10000.1	1.3699	0.1118
AllData.2	1.2606	0.1131	Sample10000.2	1.3401	0.1153
AllData.3	1.4196	0.1072	Sample10000.3	1.3177	0.1194
AllData.4	1.4995	0.1040	Sample10000.4	1.2798	0.1235
Sample500.0	1.65745	0.09933	Sample15000.0	1.91137	0.07123
Sample500.1	1.42183	0.12139	Sample15000.1	1.2958	0.1152
Sample500.2	1.39674	0.12244	Sample15000.2	1.2592	0.1183
Sample500.3	1.34841	0.12836	Sample15000.3	1.2547	0.1227
Sample500.4	1.37689	0.12490	Sample15000.4	1.2278	0.1263
Sample1000.0	1.70964	0.09604	Sample18059.0	1.93654	0.06834
Sample1000.1	1.41287	0.11977	Sample18059.1	1.39072	0.10372
Sample1000.2	1.43463	0.11709	Sample18059.2	1.3281	0.1099
Sample1000.3	1.2963	0.1328	Sample18059.3	1.3320	0.1116
Sample1000.4	1.2923	0.1332	Sample18059.4	1.3149	0.1138
Sample5000.0	1.99132	0.06108	Sample36118.0	1.91056	0.07212
Sample5000.1	1.3845	0.1089	Sample36118.1	1.3068	0.1084
Sample5000.2	1.3648	0.1096	Sample36118.2	1.2475	0.1141
Sample5000.3	1.2750	0.1208	Sample36118.3	1.3054	0.1157
Sample5000.4	1.2741	0.1208	Sample36118.4	1.3299	0.1152

c equals 1.2606, 1.4196, and 1.4995, and z equals 0.1131, 0.1072, and 0.1040 respectively. For the fit of the 500 randomly selected cells, c is 1.65745 and z is 0.09933 for a non-nested approach. For an approach with nesting to 500 km, 200 km, 100 km and 50 km grid side length c equals 1.42183, 1.39674, 1.34841, and 1.37689, and z is 0.12139, 0.12244, 0.12836, and 0.12490 respectively. When using 1000 samples, c-values vary with increased nestedness from 1.70964 to 1.41287, 1.43463, 1.2963, and 1.2923. Z-values meanwhile are 0.09604, 0.11977, 0.11709, 0.1328, and 0.1332. The non-nested 5000 sample fit gives c=1.99132 and z= 0.06108. For nesting, these numbers changed to c-values of 1.3845, 1.3648, 1.2750, and 1.2741, and z-values of 0.1089, 0.1096, 0.1208, and 0.1208. C-values of 1.83780, 1.3699, 1.3401, 1.3177, and 1.2798, and z-values of 0.07960, 0.1118, 0.1153, 0.1194, and 0.1235 were fitted for the 10000 samples. And for the 15000 ones, the following values were obtained: c=1.91137, 1.2958, 1.2592, 1.2547, and 1.2278 and z=0.07123, 0.1152, 0.1183, 0.1227, 0.1263. The amount of 18059 samples was chosen because this represents one-quarter of the total amount of cells. It resulted in c equals 1.93654, 1.39072, 1.3281, 1.3320, and 1.3149 with respective z of 0.06834, 0.10372, 0.1099, 0.1116 and 0.1138. Finally, an additional fit with 36118 samples was ran, equaling half the cells of the whole dataset. Here, c-values for increased nesting are 1.91056, 1.3068, 1.2475, 1.3054, and 1.3299 and z-values are 0.07212, 0.1084, 0.1141, 0.1157, and 0.1152. For consistency and ease of understanding, I always mentioned the non-nested results first, followed by increasing levels of nesting. The species-area relationships for all datasets are shown in Figures 3.1 and 3.2.

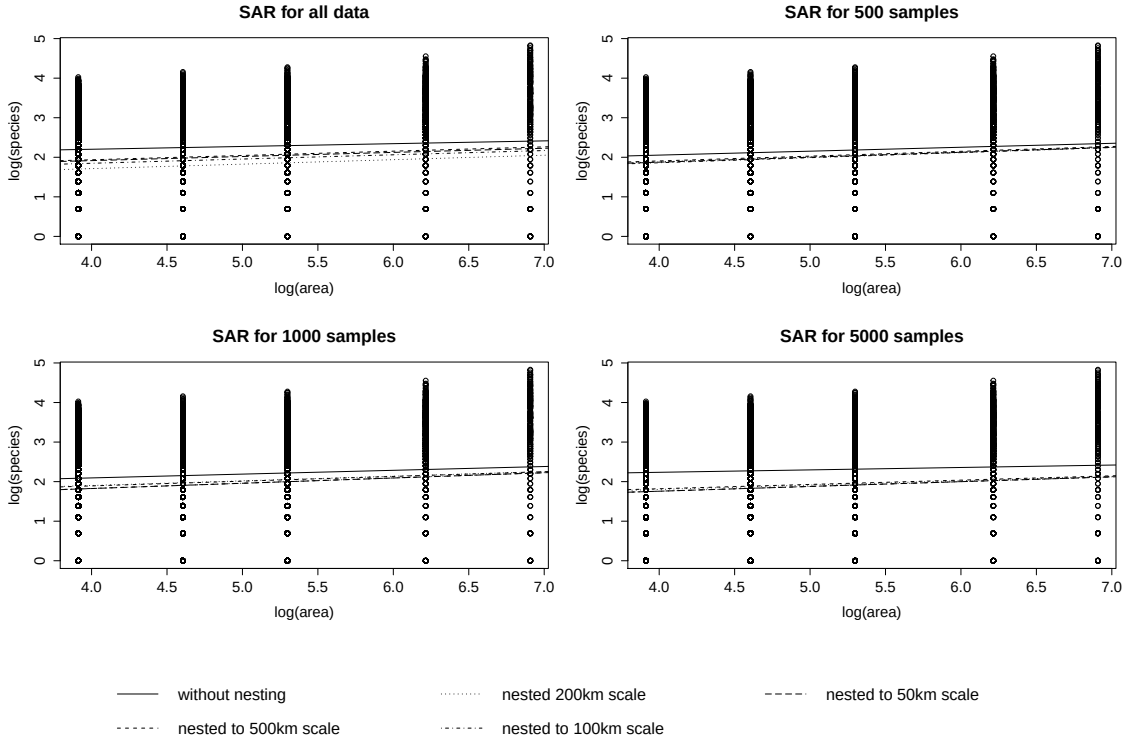


Figure 3.1: Species-area relationships for the whole dataset, 500 samples, 1000 samples, and 5000 samples

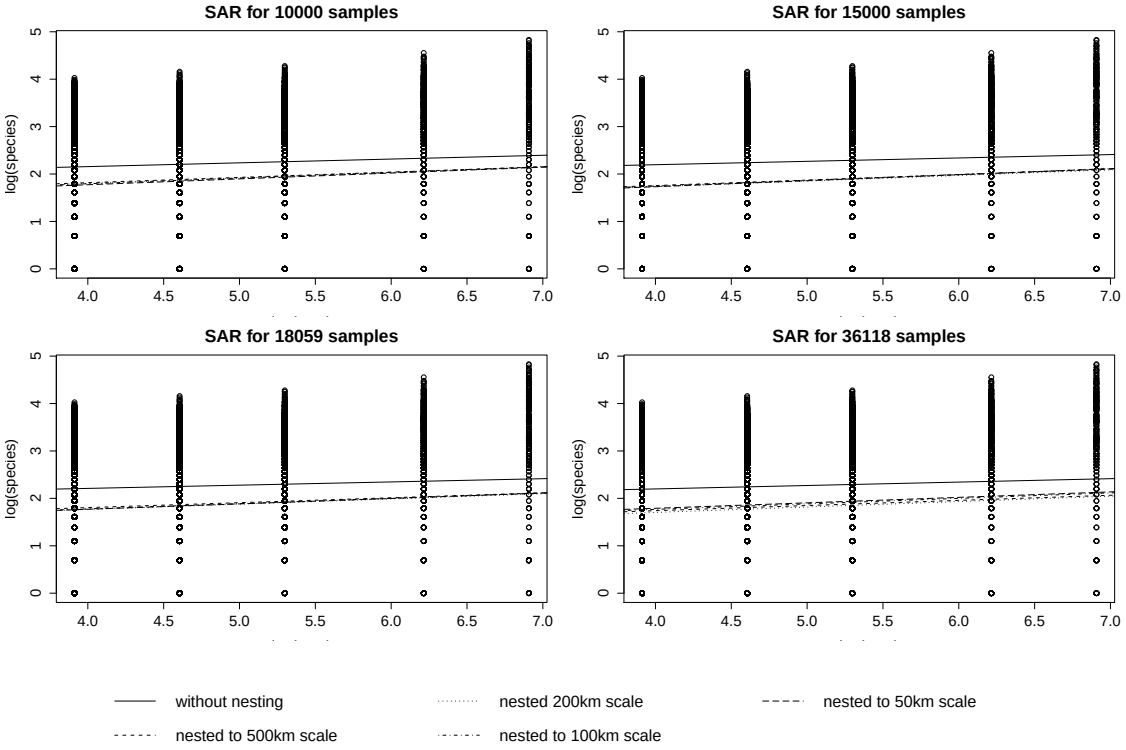


Figure 3.2: Species-area relationships for 10000 samples, 15000 samples, 18059 samples, and 36118 samples

Chapter 4

Discussion

Overall, the results listed in Table 3.1 display higher c-values for the non-nested approach than for the nested ones while still being within the same magnitude. For z-values, this difference is even more drastic with values for the non-nested approach being one magnitude lower than the nested ones. For both, there is relatively little variation within the nested approaches. Thus, the effect of nestedness on the z-values is generally detectable and also visible in Figures 4.1 and 4.2.

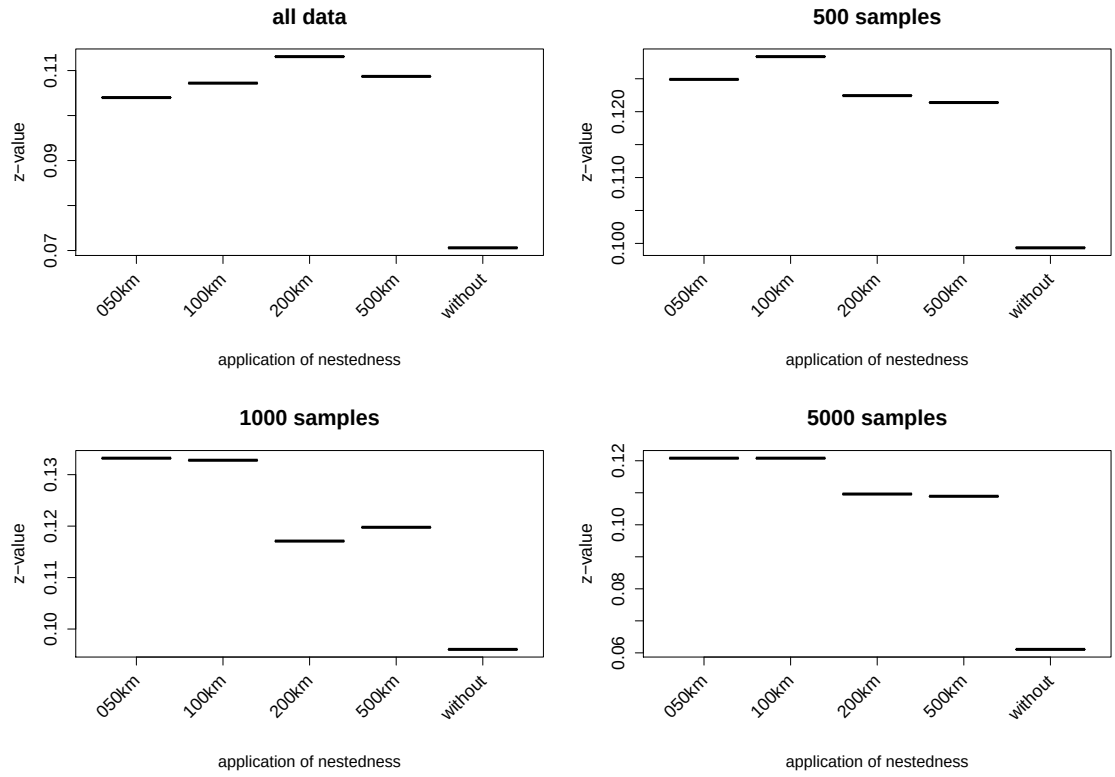


Figure 4.1: Z-values for the whole dataset, 500 samples, 1000 samples, and 5000 samples

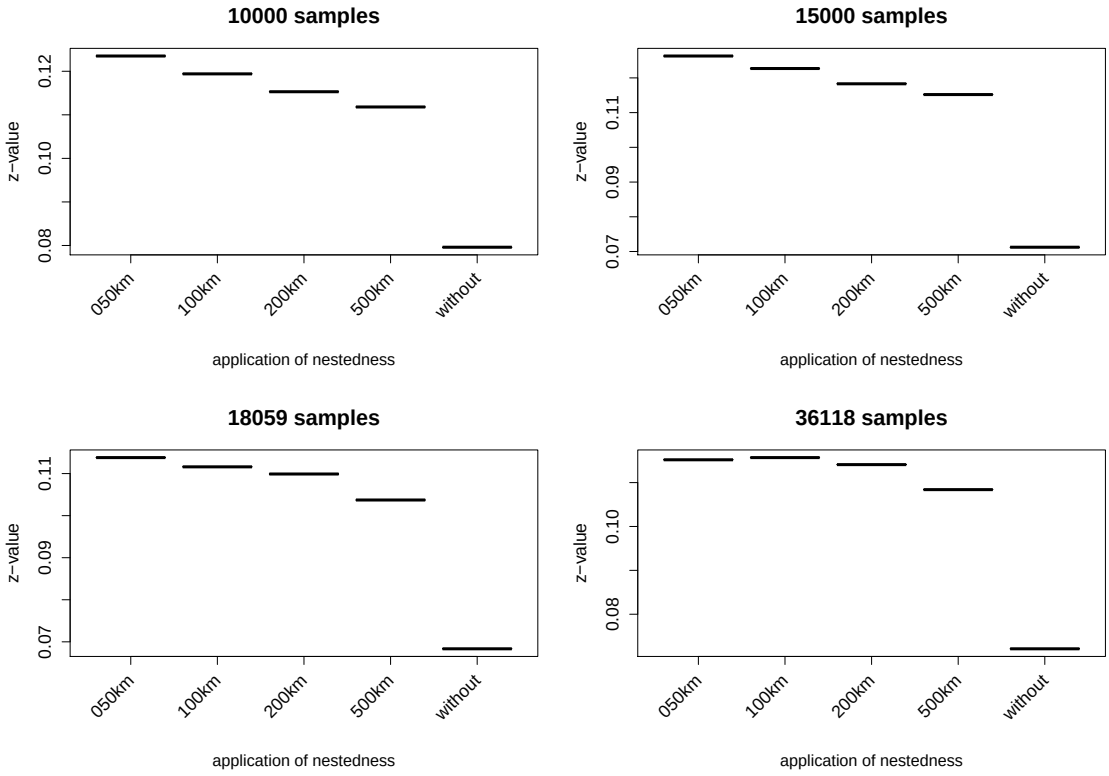


Figure 4.2: Z-values for 10000 samples, 15000 samples, 18059 samples, and 36118 samples

4.1 Slope

It is important to keep in mind that for this whole study the use of the term 'slope' is used in reference to the log-log space or clearly stated otherwise. In the arithmetic space, both values (z and c) determine the slope of the parabola. [Lomolino, 2001] Thus, it is difficult to connect them with diversities directly. In fact, the idea of MacArthur to equal the c-value with the alpha diversity (actually point diversity) and the z-value with the beta diversity is proven to be incorrect. [Connor and McCoy, 1979] The z-values themselves, being the slope in logarithmic space, are relatively low while ranging from 0.06108 to 0.1332. The typical values from other publications range from 0.1 to 0.4, with a typical range of 0.2 to 0.4. [Connor and McCoy, 1979, Gerstner et al., 2014, Lomolino, 2001] Still, they do not seem to be unreasonable, compared to Rosenzweigs publication of low z-values. He includes several values of the 0.01 magnitude, one being as low as 0.0475. [Rosenzweig, 1995] Setting aside the fact that the z-values are at the lower end of the spectrum, relatively higher slopes were obtained for the random samples. For example is the highest slope of 0.1332 the values obtained for 1000 random samples and a full nesting to the 50 km by 50 km scale. This conclusion overlaps with Goodall, Greig-Smith, and Kobayashi' finding as cited by Connor. [Connor and McCoy, 1979]

4.2 Scale-dependence

In this thesis, the nesting of the 1000 km by 1000 km, 500 km by 500 km, 200 km by 200 km, 100 km by 100 km and 50 km by 50 km grids was used as a proxy for scale-dependence, so that any influence at any level of nestedness on the fitted z-value could be interpreted as a scale-dependence. Since the values for nested approaches compared to the non-nested one differ by one magnitude, this dependence should be considered to be in effect. Additionally, it is supported by the studies of Preston and Arita, who also detected scale-dependence in their analyses. [Ladle and Whittaker, 2011, Arita and Rodrguez, 2002] This implies a convex upward curve instead of a linear relationship in log-log space, thus the linearity of the power function in log-log transformation would be doubted in its usefulness as it only approximates the fitted values. [Arita and Rodrguez, 2002, Connor and McCoy, 1979, Rosenzweig, 1995] It might have severe implications for biodiversity patterns, which are not discussed here. [Storch et al., 2012] Whittaker mentioned an obscurity of scale-dependence based on this transformation, so that this study might be doubted in its general effectiveness. [Whittaker et al., 2001] The log-log transformation on the other hand was suggested by Dengler and Lomolino and is used widely. Considering this, it is logical to use it for comparability and due to its proven usefulness. The convex upward shape is minimal enough to still allow a reasonable fit with the linear approach. [Dengler, 2009, Lomolino, 2001, Rosenzweig, 1995] Furthermore, the steeper slopes for nested approaches such as the z-value of 0.1040 instead of 0.07060 (non-nested) for the whole dataset are not uncommon. Drakare further obtained steeper slopes for a nested one (0.36 to 0.24). He also found out, that a nested approach results in a better fit than independent sampling. Hereby, he defined independent sampling as using ”spatially non-overlapping areas”. [Drakare et al., 2006]

4.3 Subsets

For the subsets, there is no difference in c-values to be seen compared to the whole dataset, since there is a high variation within the samples. The z-values on the other hand, show a slight tendency to approximate the values for the whole dataset with increasing sample size for each level of nestedness. For easier comprehensibility, Figure 4.3 only shows the values for full nesting to the 50 km by 50 km grid and non-nesting for each sample size. Although an approximation is detectable and slightly visible, it does not seem to be adequate to consider any of the sample sizes large enough to substitute for the whole dataset. The non-nested approach for the whole dataset gives a z-value of 0.07060, whereas for the largest sample size of 36118 samples it still is 0.07212. For the fully nested approach, it is 0.1040 for the whole dataset and 0.1152 for the 36118 samples. The total sample size is 72236 cells so that 36118 samples represent exactly half of the total samples. For a quarter of the total sample size (18059 samples), the z is 0.06834 for the non-nested approach, and 0.1138 for the nested approach to 50 km by 50 km. So the values show the ambiguity of the approximation. With the pronounced variation, it is not surprising to find outliers such as the very low 0.06108 for non-nested 5000 samples.

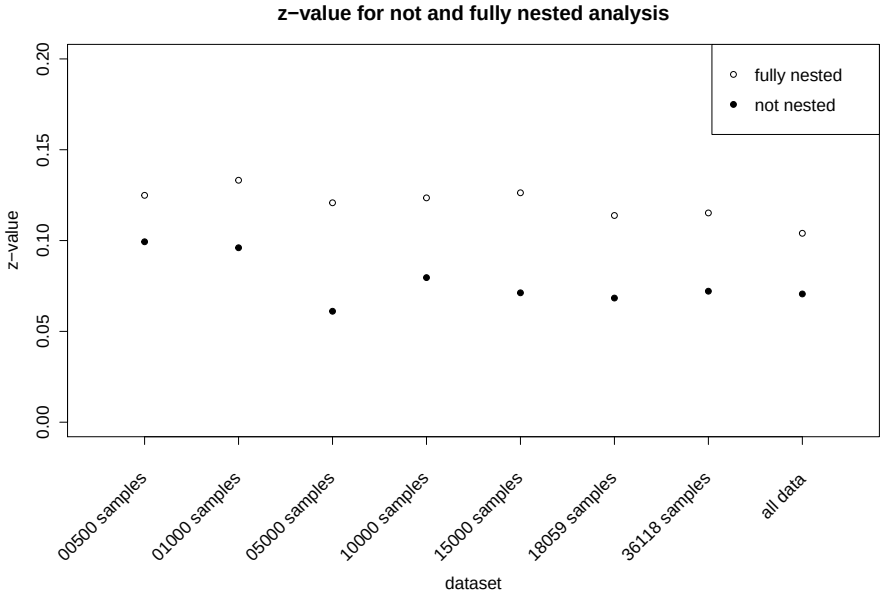


Figure 4.3: Comparison of z-values for fully and non-nested analysis

Chapter 5

Conclusion

The obtained values of the MCMCglmm fitting of the rodents ranges at different levels of nestedness are all within a logical range, even though they are relatively low. The difference of one magnitude between non-nested and nested approaches indicates scale-dependence (0.01 instead of 0.1 for non-nested compared to nested). Additionally for the different sampling sizes, the larger samples slightly improve the fitted values. Still, a sample size of up to half the total grid cells is not enough to predict the z or c-value for the whole dataset. For further research the idea of a sigmodial or triphasic species-area relationship as mentioned by Chaloner, Connor&McCoy, Dengler and Lomolino should be studied, as well as the use of direct sampling instead of range maps. The terminological confusion as well could use some clarification due to its effect of difficult and erroneous conclusions, inappropriate sampling [Dengler, 2009], and, based on my own experience, difficulty of understanding and comparison of former research.

Chapter 6

Acknowledgments

I'd like to thank my supervisor Prof. Dormann for the support during the last three month. Especially for the fact that I had the feeling that no question, however simple, would be considered stupid to ask. Also much thanks to Dr. Staab for co-correcting and Lara Budic for her help on the R programming. Last but not leased I thank my friends and family for supporting me during this time of constant talk about the thesis and limited social availability.

Bibliography

[Arita and Rodrguez, 2002] Arita, H. T. and Rodrguez, P. (2002). Geographic range, turnover rate and the scaling of species diversity. *Ecography*, 25(5):541–550.

[Bivand et al., 2015] Bivand, R., Keitt, T., and Rowlingson, B. (2015). *rgdal: Bindings for the Geospatial Data Abstraction Library*. R package version 1.0-7.

[Bivand and Lewin-Koh, 2015] Bivand, R. and Lewin-Koh, N. (2015). *maptools: Tools for Reading and Handling Spatial Objects*. R package version 0.8-36.

[Connor and McCoy, 1979] Connor, E. F. and McCoy, E. D. (1979). The Statistics and Biology of the Species-Area Relationship. *The American Naturalist*, 113(6):pp. 791–833.

[Crawley and Harral, 2001] Crawley, M. J. and Harral, J. E. (2001). Scale Dependence in Plant Biodiversity. *Science*, 291(5505):864–868.

[Dengler, 2009] Dengler, J. (2009). Which function describes the species-area relationship best? A review and empirical evaluation. *Journal of Biogeography*, 36(4):728–744.

[Drakare et al., 2006] Drakare, S., Lennon, J. J., and Hillebrand, H. (2006). The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecology Letters*, 9(2):215–227.

[Gerstner et al., 2014] Gerstner, K., Dormann, C. F., Vclavk, T., Kreft, H., and Seppelt, R. (2014). Accounting for geographical variation in species-area relationships improves the prediction of plant species richness at the global scale. *Journal of Biogeography*, 41(2):261–273.

[Gray et al., 2004] Gray, J. S., Ugland, K. I., and Lambshead, J. (2004). Species accumulation and species area curves - a comment on Scheiner (2003). *Global Ecology and Biogeography*, 13(5):473–476.

[Hadfield, 2010] Hadfield, J. D. (2010). Mcmc methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33.

[Hijmans, 2015] Hijmans, R. J. (2015). *raster: Geographic Data Analysis and Modeling*. R package version 2.4-20.

[IUCN, 2014] IUCN (2014). The IUCN Red List of Threatened Species. http://www.iucnredlist.org. Downloaded on Sep 2nd 2015.

[Ladle and Whittaker, 2011] Ladle, R. and Whittaker, R. (2011). *Conservation Biogeography*. Wiley.

[Lomolino, 2001] Lomolino, M. (2001). The species-area relationship: new challenges for an old pattern. *Progress in Physical Geography*, 25(1):1 – 21.

[Preston, 1960] Preston, F. W. (1960). Time and Space and the Variation of Species. *Ecology*, 41(4):612–627.

[Rosenzweig, 1995] Rosenzweig, M. L. (1995). *Species Diversity in Space and Time*. Cambridge University Press.

[Scheiner, 2003] Scheiner, S. M. (2003). Six types of species-area curves. *Global Ecology and Biogeography*, 12(6):441–447.

[Scheiner, 2004] Scheiner, S. M. (2004). A mélange of curves - further dialogue about species-area relationships. *Global Ecology and Biogeography*, 13(6):479–484.

[Stiles and Scheiner, 2010] Stiles, A. and Scheiner, S. M. (2010). A multi-scale analysis of fragmentation effects on remnant plant species richness in Phoenix, Arizona. *Journal of Biogeography*, 37(9):1721–1729.

[Storch et al., 2012] Storch, D., Keil, P., and Jetz, W. (2012). Universal species-area and endemics-area relationships at continental scales. , 488:78–81.

[Whittaker and Fernandez-Palacios, 2007] Whittaker, R. and Fernandez-Palacios, J. (2007). *Island Biogeography: Ecology, Evolution, and Conservation*. Oxford biology. OUP Oxford.

[Whittaker et al., 2001] Whittaker, R. J., Willis, K. J., and Field, R. (2001). Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, 28(4):453–470.

Appendix A

Selbstständigkeitserklärung

Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen habe ich unter Angabe der Quellen kenntlich gemacht. Die digitale entspricht der gedruckten Fassung. Die Arbeit wurde noch in keiner Form einem anderen Prüfungsamt vorgelegt oder veröffentlicht.

(I hereby declare that I have composed this document unassistedly and that I only used the sources and devices I declared. Passages taken verbatim or in meaning from other sources are identified as such and the sources are acknowledged and cited. The digital and printed version are alike. The thesis is no part of any other examination and was not yet published.)

Karina Reckling, December 2015

Appendix B

extract.species function

Listing B.1: Extract.species function by Carsten F. Dormann

```
extract.species ¡- function(mams, species)-
        # SPDF is the spatialPolygonsDataFrame object
        # species.name of the species to be extracted
        # author: Carsten F. Dormann
        if ("SCI NAME" %in% names(mams@data))
                ind ¡- which(mams@data$SCI NAME == species)
        if ("BINOMIAL" %in% names(mams@data))
                ind ¡- which(mams@data$BINOMIAL == species)
        if ("binomial" %in% names(mams@data))
                ind ¡- which(mams@data$binomial == species)
        selected.species ¡- mams
        selected.species@data ¡- mams@data[ind,]
        selected.species@polygons ¡- mams@polygons[ind]
        selected.species@plotOrder ¡- seq along(ind)
        #plot(selected.species, col=my.colours[i], add=T)
        #print(i)
        return(selected.species)
```

Appendix C

R script

Listing C.1: Complete R script

```
library(maptools); library(rgdal); library(raster)
library(MCMCglmm); library(lattice); library(grid)

#load and prepare shapefile
mams ¡- readShapePoly(" /Data/TERRESTRIAL MAMMALS.shp")
mams@proj4string ¡- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no defs +
    towgs84=0,0,0")
mams ¡- spTransform(mams, CRS('+proj=moll'))

#create empty raster with extant from mams
r ¡- raster(nrow=310, ncols=710, xmn=-17702769, xmx=17876212, ymn=-6496757, ymx
    =9020048, crs=CRS('+proj=moll'))

#extract polygons function
extract.species ¡- function(mams, species)-
# SPDF is the spatialPolygonsDataFrame object
# species.name of the species to be extracted
# author: Carsten F. Dormann
if ("SCI NAME" %in% names(mams@data)) ind ¡- which(mams@data$SCI NAME ==
    species)
if ("BINOMIAL" %in% names(mams@data)) ind ¡- which(mams@data$BINOMIAL ==
    species)
if ("binomial" %in% names(mams@data)) ind ¡- which(mams@data$binomial ==
    species)
selected.species ¡- mams
selected.species@data ¡- mams@data[ind,]
selected.species@polygons ¡- mams@polygons[ind]
selected.species@plotOrder ¡- seq along(ind)
#plot(selected.species, col=my.colours[i], add=T)
#print(i)
return(selected.species)


#extracting species and save
species ¡- unique(mams@data$binomial[which(mams@data$order name=="RODENTIA")])
raster.rodentia ¡- rasterize(extract.species(mams, species[1]), r)
for (i in 2:length(species))-
raster.rodentia2 ¡- rasterize(extract.species(mams, species[i]), r)
raster.rodentia ¡- addLayer(raster.rodentia, raster.rodentia2)
print(paste("Finished the ", i, "th species: ", species[i], sep=""))

names(raster.rodentia) ¡- species
raster50 ¡- raster.rodentia; save(raster50, file=" /raster50grid.Rdata")
```

```
rm(i, raster.rodentia2, mams, r, species, extract.species); gc()

#aggregating raster
#original 50, fact: 2(100), 4(200), 10(500), 20(1000)
raster.rodentia <- raster50
raster100 <- aggregate(raster.rodentia, fact=2, fun=mean, na.rm=T)
save(raster100, file=" /raster100grid.Rdata")
raster200 <- aggregate(raster.rodentia, fact=4, fun=mean, na.rm=T)
save(raster200, file=" /raster200grid.Rdata")
raster500 <- aggregate(raster.rodentia, fact=10, fun=mean, na.rm=T)
save(raster500, file=" /raster500grid.Rdata")
raster1000 <- aggregate(raster.rodentia, fact=20, fun=mean, na.rm=T)
save(raster1000, file=" /raster1000grid.Rdata")
rm(raster.rodentia); gc()

#create data frames and save
coord50 <- as.data.frame(xyFromCell(raster50, 1:ncell(raster50)))
cover50 <- cbind.data.frame(coord50, values(raster50))
rm(coord50, raster50); gc()
save(cover50, file=" /cover50grid.Rdata")
coord100 <- as.data.frame(xyFromCell(raster100, 1:ncell(raster100)))
cover100 <- cbind.data.frame(coord100, values(raster100))
rm(coord100, raster100); gc()
save(cover100, file=" /cover100grid.Rdata")
coord200 <- as.data.frame(xyFromCell(raster200, 1:ncell(raster200)))
cover200 <- cbind.data.frame(coord200, values(raster200))
rm(coord200, raster200); gc()
save(cover200, file=" /cover200grid.Rdata")
coord500 <- as.data.frame(xyFromCell(raster500, 1:ncell(raster500)))
cover500 <- cbind.data.frame(coord500, values(raster500))
rm(coord500, raster500); gc()
save(cover500, file=" /cover500grid.Rdata")
coord1000 <- as.data.frame(xyFromCell(raster1000, 1:ncell(raster1000)))
cover1000 <- cbind.data.frame(coord1000, values(raster1000))
rm(coord1000, raster1000); gc()
save(cover1000, file=" /cover1000grid.Rdata")

#replace all NAs by 0, all non-NAs by 1 and save
cover50[, -c(1,2)] <- apply(cover50[, -c(1,2), drop=F], 2, function(x) ifelse(
    is.na(x), 0, 1))
table(rowSums(cover50[, -c(1,2)]))
save(cover50, file=" /cover50grid.Rdata")
cover100[, -c(1,2)] <- apply(cover100[, -c(1,2), drop=F], 2, function(x) ifelse
    (is.na(x), 0, 1))
table(rowSums(cover100[, -c(1,2)]))
save(cover100, file=" /cover100grid.Rdata")
cover200[, -c(1,2)] <- apply(cover200[, -c(1,2), drop=F], 2, function(x) ifelse
    (is.na(x), 0, 1))
table(rowSums(cover200[, -c(1,2)]))
save(cover200, file=" /cover200grid.Rdata")
cover500[, -c(1,2)] <- apply(cover500[, -c(1,2), drop=F], 2, function(x) ifelse
    (is.na(x), 0, 1))
table(rowSums(cover500[, -c(1,2)]))
save(cover500, file=" /cover500grid.Rdata")
cover1000[, -c(1,2)] <- apply(cover1000[, -c(1,2), drop=F], 2, function(x)
    ifelse(is.na(x), 0, 1))
table(rowSums(cover1000[, -c(1,2)]))
save(cover1000, file=" /cover1000grid.Rdata")

#getting ID rasters
r50 <- raster(nrow=310, ncols=710, xmn=-17702769, xmx=17876212, ymn=-6496757,
```

```
    ymx=9020048, crs=CRS('+proj=moll'))
r50[] <- 1:ncell(r50)
r100 <- aggregate(r50, fact=2, fun=min, na.rm=F)
r200 <- aggregate(r50, fact=4, fun=min, na.rm=F)
r500 <- aggregate(r50, fact=10, fun=min, na.rm=F)
r1000 <- aggregate(r50, fact=20, fun=min, na.rm=F)

#create ID columns
#ID50
cover50 <- cbind(cover50, 'ID50'= seq(1,220100))
cover100 <- cbind(cover100, 'ID50'= NA)
cover200 <- cbind(cover200, 'ID50'= NA)
cover500 <- cbind(cover500, 'ID50'= NA)
cover1000 <- cbind(cover1000, 'ID50'= NA)

#ID100
Poly100 <- rasterToPolygons(r100, na.rm=F)
a <- extract(r50, Poly100, df=T)
b <- a[order(a$layer),]
cover50 <- cbind(cover50, 'ID100'= b$ID)
cover100 <- cbind(cover100, 'ID100'= seq(1,55025))
cover200 <- cbind(cover200, 'ID100'= NA)
cover500 <- cbind(cover500, 'ID100'= NA)
cover1000 <- cbind(cover1000, 'ID100'= NA)
rm(a, b, Poly100); gc()

#ID200
Poly200 <- rasterToPolygons(r200, na.rm=F)
a <- extract(r50, Poly200, df=T)
b <- a[order(a$layer),]
cover50 <- cbind(cover50, 'ID200'= b$ID)
c <- extract(r100, Poly200, df=T)
d <- c[order(c$layer),]
cover100 <- cbind(cover100, 'ID200'= d$ID)
cover200 <- cbind(cover200, 'ID200'= seq(1,13884))
cover500 <- cbind(cover500, 'ID200'= NA)
cover1000 <- cbind(cover1000, 'ID200'= NA)
rm(a, b, c, d, Poly200); gc()

#ID500
Poly500 <- rasterToPolygons(r500, na.rm=F)
a <- extract(r50, Poly500, df=T)
b <- a[order(a$layer),]
cover50 <- cbind(cover50, 'ID500'= b$ID)
c <- extract(r100, Poly500, df=T)
d <- c[order(c$layer),]
cover100 <- cbind(cover100, 'ID500'= d$ID)
cover200 <- cbind(cover200, 'ID500'= NA)
cover500 <- cbind(cover500, 'ID500'= seq(1,2201))
cover1000 <- cbind(cover1000, 'ID500'= NA)
rm(a, b, c, d, Poly500); gc()

#ID1000
Poly1000 <- rasterToPolygons(r1000, na.rm=F)
a <- extract(r50, Poly1000, df=T)
b <- a[order(a$layer),]
cover50 <- cbind(cover50, 'ID1000'= b$ID)
c <- extract(r100, Poly1000, df=T)
d <- c[order(c$layer),]
cover100 <- cbind(cover100, 'ID1000'= d$ID)
e <- extract(r200, Poly1000, df=T)
```

```
f ¡- e[order(e$layer),]
cover200 ¡- cbind(cover200, 'ID1000'=f$ID)
g ¡- extract(r500, Poly1000, df=T)
h ¡- g[order(g$layer),]
cover500 ¡- cbind(cover500, 'ID1000'=h$ID)
cover1000 ¡- cbind(cover1000, 'ID1000'= seq(1,576))
rm(a,b,c,d,e,f,g,h, Poly1000, r50, r100, r200, r500, r1000); gc()

#saving data frames
save(cover50, file=" /cover50gridID.Rdata")
save(cover100, file=" /cover100gridID.Rdata")
save(cover200, file=" /cover200gridID.Rdata")
save(cover500, file=" /cover500gridID.Rdata")
save(cover1000, file=" /cover1000gridID.Rdata")

#add sum and res columns and save
cover50 ¡- cbind(cover50, 'sum'=c(1:nrow(cover50)))
for ( i in 1:nrow(cover50))-cover50$sum[i] ¡- sum(cover50[i,c(3:2220)])
cover50 ¡- cbind(cover50, 'res'=50); rm(i); gc()
save(cover50, file=" /cover50gridFinal.Rdata")
cover100 ¡- cbind(cover100, 'sum'=c(1:nrow(cover100)))
for ( i in 1:nrow(cover100))-cover100$sum[i] ¡- sum(cover100[i,c(3:2220)])
cover100 ¡- cbind(cover100, 'res'=100); rm(i); gc()
save(cover100, file=" /cover100gridFinal.Rdata")
cover200 ¡- cbind(cover200, 'sum'=c(1:nrow(cover200)))
for ( i in 1:nrow(cover200))-cover200$sum[i] ¡- sum(cover200[i,c(3:2220)])
cover200 ¡- cbind(cover200, 'res'=200); rm(i); gc()
save(cover200, file=" /cover200gridFinal.Rdata")
cover500 ¡- cbind(cover500, 'sum'=c(1:nrow(cover500)))
for ( i in 1:nrow(cover500))-cover500$sum[i] ¡- sum(cover500[i,c(3:2220)])
cover500 ¡- cbind(cover500, 'res'=500); rm(i); gc()
save(cover500, file=" /cover500gridFinal.Rdata")
cover1000 ¡- cbind(cover1000, 'sum'=c(1:nrow(cover1000)))
for ( i in 1:nrow(cover1000))-cover1000$sum[i] ¡- sum(cover1000[i,c(3:2220)])
cover1000 ¡- cbind(cover1000, 'res'=1000); rm(i); gc()
save(cover1000, file=" /cover1000gridFinal.Rdata")

#selecting columns and save
attach(cover50)
cover50 ¡- cbind.data.frame(x, y, ID50, ID100, ID200, ID500, ID1000, sum,res)
detach(cover50); save(cover50, file=" /cover50gridFinal.Rdata"); attach(
    cover100)
cover100 ¡- cbind.data.frame(x, y, ID50, ID100, ID200, ID500, ID1000, sum,res)
detach(cover100); save(cover100, file=" /cover100gridFinal.Rdata"); attach(
    cover200)
cover200 ¡- cbind.data.frame(x, y, ID50, ID100, ID200, ID500, ID1000, sum,res)
detach(cover200); save(cover200, file=" /cover200gridFinal.Rdata"); attach(
    cover500)
cover500 ¡- cbind.data.frame(x, y, ID50, ID100, ID200, ID500, ID1000, sum,res)
detach(cover500); save(cover500, file=" /cover500gridFinal.Rdata"); attach(
    cover1000)
cover1000 ¡- cbind.data.frame(x, y, ID50, ID100, ID200, ID500, ID1000, sum,res)
detach(cover1000); save(cover1000, file=" /cover1000gridFinal.Rdata")

#combining data frames and save
cover.all ¡- rbind.data.frame(cover50, cover100, cover200, cover500, cover1000)
save(cover.all, file=" /cover.all.Rdata")
cover.small ¡- cover.all[cover.all$sum != 0,]
save(cover.small, file=" /cover.small.Rdata")
rm(cover.small, cover.all, cover50, cover100, cover200, cover500, cover1000);
    gc()
```

```
#selecting islands and save
load(" /cover.small.Rdata")
row.names ¡- row.names(cover.small)
sample500 ¡- sample(row.names, 500)
sample500 ¡- subset(cover.small, row.names%in%sample500)
save(sample500, file=" /sample500.Rdata")
sample1000 ¡- sample(row.names, 1000)
sample1000 ¡- subset(cover.small, row.names%in%sample1000)
save(sample1000, file=" /sample1000.Rdata")
sample5000 ¡- sample(row.names, 5000)
sample5000 ¡- subset(cover.small, row.names%in%sample5000)
save(sample5000, file=" /sample5000.Rdata")
sample10000 ¡- sample(row.names, 10000)
sample10000 ¡- subset(cover.small, row.names%in%sample10000)
save(sample10000, file=" /sample10000.Rdata")
sample15000 ¡- sample(row.names, 15000)
sample15000 ¡- subset(cover.small, row.names%in%sample15000)
save(sample15000, file=" /sample15000.Rdata")
sample18059 ¡- sample(row.names, 18059)
sample18059 ¡- subset(cover.small, row.names%in%sample18059)
save(sample18059, file=" /sample18059.Rdata")
sample36118 ¡- sample(row.names, 36118)
sample36118 ¡- subset(cover.small, row.names%in%sample36118)
save(sample36118, file=" /sample36118.Rdata")
rm(cover.small, row.names, sample500, sample1000, sample5000, sample10000,
    sample15000, sample18059, sample36118); gc()

#fitting contiguous cells (Type II)
load(" /cover.small.Rdata")
fm contiguous0 ¡- MCMCglmm(fixed=log(sum) log(res 2), data=cover.small)
fm contiguous1 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID500, data=cover
    .small)
fm contiguous2 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500,
    data=cover.small)
fm contiguous3 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200, data=cover.small)
fm contiguous4 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200+ID100, data=cover.small)
summary(fm contiguous0); summary(fm contiguous1); summary(fm contiguous2)
summary(fm contiguous3); summary(fm contiguous4)
save(fm contiguous0, file=" /fm contiguous0.Rdata")
save(fm contiguous1, file=" /fm contiguous1.Rdata")
save(fm contiguous2, file=" /fm contiguous2.Rdata")
save(fm contiguous3, file=" /fm contiguous3.Rdata")
save(fm contiguous4, file=" /fm contiguous4.Rdata")
rm(fm contiguous0, fm contiguous1, fm contiguous2, fm contiguous3, fm
    contiguous4, cover.small); gc()

#fitting islands (Type IV)
load(" /sample500.Rdata")
fm sample500.0 ¡- MCMCglmm(fixed=log(sum) log(res 2), data=sample500)
fm sample500.1 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000, data=
    sample500)
fm sample500.2 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500,
    data=sample500)
fm sample500.3 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200, data=sample500)
fm sample500.4 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200+ID100, data=sample500)
summary(fm sample500.0); summary(fm sample500.1); summary(fm sample500.2)
```

```
summary(fm sample500.3); summary(fm sample500.4)
save(fm sample500.0, file=" /fm sample500.0.Rdata")
save(fm sample500.1, file=" /fm sample500.1.Rdata")
save(fm sample500.2, file=" /fm sample500.2.Rdata")
save(fm sample500.3, file=" /fm sample500.3.Rdata")
save(fm sample500.4, file=" /fm sample500.4.Rdata")
rm(sample500, fm sample500.0, fm sample500.1, fm sample500.2, fm sample500.3,
    fm sample500.4); gc()

load(" /sample1000.Rdata")
fm sample1000.0 ¡- MCMCglmm(fixed=log(sum) log(res 2), data=sample1000)
fm sample1000.1 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000, data=
    sample1000)
fm sample1000.2 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500,
    data=sample1000)
fm sample1000.3 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200, data=sample1000)
fm sample1000.4 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200+ID100, data=sample1000)
summary(fm sample1000.0); summary(fm sample1000.1); summary(fm sample1000.2)
summary(fm sample1000.3); summary(fm sample1000.4)
save(fm sample1000.0, file=" /fm sample1000.0.Rdata")
save(fm sample1000.1, file=" /fm sample1000.1.Rdata")
save(fm sample1000.2, file=" /fm sample1000.2.Rdata")
save(fm sample1000.3, file=" /fm sample1000.3.Rdata")
save(fm sample1000.4, file=" /fm sample1000.4.Rdata")
rm(sample1000, fm sample1000.0, fm sample1000.1, fm sample1000.2, fm sample1000
    .3, fm sample1000.4); gc()

load(" /sample5000.Rdata")
fm sample5000.0 ¡- MCMCglmm(fixed=log(sum) log(res 2), data=sample5000)
fm sample5000.1 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000, data=
    sample5000)
fm sample5000.2 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500,
    data=sample5000)
fm sample5000.3 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200, data=sample5000)
fm sample5000.4 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200+ID100, data=sample5000)
summary(fm sample5000.0); summary(fm sample5000.1); summary(fm sample5000.2)
summary(fm sample5000.3); summary(fm sample5000.4)
save(fm sample5000.0, file=" /fm sample5000.0.Rdata")
save(fm sample5000.1, file=" /fm sample5000.1.Rdata")
save(fm sample5000.2, file=" /fm sample5000.2.Rdata")
save(fm sample5000.3, file=" /fm sample5000.3.Rdata")
save(fm sample5000.4, file=" /fm sample5000.4.Rdata")
rm(sample5000, fm sample5000.0, fm sample5000.1, fm sample5000.2, fm sample5000
    .3, fm sample5000.4); gc()

load(" /sample10000.Rdata")
fm sample10000.0 ¡- MCMCglmm(fixed=log(sum) log(res 2), data=sample10000)
fm sample10000.1 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000, data=
    sample10000)
fm sample10000.2 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500,
    data=sample10000)
fm sample10000.3 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200, data=sample10000)
fm sample10000.4 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200+ID100, data=sample10000)
summary(fm sample10000.0); summary(fm sample10000.1); summary(fm sample10000.2)
summary(fm sample10000.3); summary(fm sample10000.4)
```

```
save(fm sample10000.0, file=" /fm sample10000.0.Rdata")
save(fm sample10000.1, file=" /fm sample10000.1.Rdata")
save(fm sample10000.2, file=" /fm sample10000.2.Rdata")
save(fm sample10000.3, file=" /fm sample10000.3.Rdata")
save(fm sample10000.4, file=" /fm sample10000.4.Rdata")
rm(sample10000, fm sample10000.0, fm sample10000.1, fm sample10000.2, fm
    sample10000.3, fm sample10000.4); gc()

load(" /sample15000.Rdata")
fm sample15000.0 ¡- MCMCglmm(fixed=log(sum) log(res 2), data=sample15000)
fm sample15000.1 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000, data=
    sample15000)
fm sample15000.2 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500,
    data=sample15000)
fm sample15000.3 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200, data=sample15000)
fm sample15000.4 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200+ID100, data=sample15000)
summary(fm sample15000.0); summary(fm sample15000.1); summary(fm sample15000.2)
summary(fm sample15000.3); summary(fm sample15000.4)
save(fm sample15000.0, file=" /fm sample15000.0.Rdata")
save(fm sample15000.1, file=" /fm sample15000.1.Rdata")
save(fm sample15000.2, file=" /fm sample15000.2.Rdata")
save(fm sample15000.3, file=" /fm sample15000.3.Rdata")
save(fm sample15000.4, file=" /fm sample15000.4.Rdata")
rm(sample15000, fm sample15000.0, fm sample15000.1, fm sample15000.2, fm
    sample15000.3, fm sample15000.4); gc()

load(" /sample18059.Rdata")
fm sample18059.0 ¡- MCMCglmm(fixed=log(sum) log(res 2), data=sample18059)
fm sample18059.1 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000, data=
    sample18059)
fm sample18059.2 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500,
    data=sample18059)
fm sample18059.3 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200, data=sample18059)
fm sample18059.4 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200+ID100, data=sample18059)
summary(fm sample18059.0); summary(fm sample18059.1); summary(fm sample18059.2)
summary(fm sample18059.3); summary(fm sample18059.4)
save(fm sample18059.0, file=" /fm sample18059.0.Rdata")
save(fm sample18059.1, file=" /fm sample18059.1.Rdata")
save(fm sample18059.2, file=" /fm sample18059.2.Rdata")
save(fm sample18059.3, file=" /fm sample18059.3.Rdata")
save(fm sample18059.4, file=" /fm sample18059.4.Rdata")
rm(sample18059, fm sample18059.0, fm sample18059.1, fm sample18059.2, fm
    sample18059.3, fm sample18059.4); gc()

load(" /sample36118.Rdata")
fm sample36118.0 ¡- MCMCglmm(fixed=log(sum) log(res 2), data=sample36118)
fm sample36118.1 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000, data=
    sample36118)
fm sample36118.2 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500,
    data=sample36118)
fm sample36118.3 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200, data=sample36118)
fm sample36118.4 ¡- MCMCglmm(fixed=log(sum) log(res 2), random= ID1000+ID500+
    ID200+ID100, data=sample36118)
summary(fm sample36118.0); summary(fm sample36118.1); summary(fm sample36118.2)
summary(fm sample36118.3); summary(fm sample36118.4)
save(fm sample36118.0, file=" /fm sample36118.0.Rdata")
```

```
save(fm sample36118.1, file=" /fm sample36118.1.Rdata")
save(fm sample36118.2, file=" /fm sample36118.2.Rdata")
save(fm sample36118.3, file=" /fm sample36118.3.Rdata")
save(fm sample36118.4, file=" /fm sample36118.4.Rdata")
rm(sample36118, fm sample36118.0, fm sample36118.1, fm sample36118.2, fm
    sample36118.3, fm sample36118.4); gc()

#Maps
load(" /cover50gridFinal.Rdata")
cover50 ¡- cover50[cover50$sum != 0,]
col ¡- colorRampPalette(c('lightgrey', 'darkgrey'))(150)
levelplot(sum x+y, data=cover50, main='Rodentia species richness for 50km grid'
    ,
xlab='Longitude [meters]', ylab='Latitude [meters]', col.regions=col)
trellis.focus("legend", side="right", clipp.off=TRUE, highlight=FALSE)
grid.text(expression(N(species)), 0.2, 0, hjust=0.5, vjust=1)
trellis.unfocus(); rm(cover50, col); gc()

load(" /cover100gridFinal.Rdata")
cover100 ¡- cover100[cover100$sum != 0,]
col ¡- colorRampPalette(c('lightgrey', 'darkgrey'))(150)
levelplot(sum x+y, data=cover100, main='Rodentia species richness for 100km
    grid',
xlab='Longitude [meters]', ylab='Latitude [meters]', col.regions=col)
trellis.focus("legend", side="right", clipp.off=TRUE, highlight=FALSE)
grid.text(expression(N(species)), 0.2, 0, hjust=0.5, vjust=1)
trellis.unfocus(); rm(cover100, col); gc()

load(" /cover200gridFinal.Rdata")
cover200 ¡- cover200[cover200$sum != 0,]
col ¡- colorRampPalette(c('lightgrey', 'darkgrey'))(150)
levelplot(sum x+y, data=cover200, main='Rodentia species richness for 200km
    grid',
xlab='Longitude [meters]', ylab='Latitude [meters]', col.regions=col)
trellis.focus("legend", side="right", clipp.off=TRUE, highlight=FALSE)
grid.text(expression(N(species)), 0.2, 0, hjust=0.5, vjust=1)
trellis.unfocus(); rm(cover200, col); gc()

load(" /cover500gridFinal.Rdata")
cover500 ¡- cover500[cover500$sum != 0,]
col ¡- colorRampPalette(c('lightgrey', 'darkgrey'))(150)
levelplot(sum x+y, data=cover500, main='Rodentia species richness for 500km
    grid',
xlab='Longitude [meters]', ylab='Latitude [meters]', col.regions=col)
trellis.focus("legend", side="right", clipp.off=TRUE, highlight=FALSE)
grid.text(expression(N(species)), 0.2, 0, hjust=0.5, vjust=1)
trellis.unfocus(); rm(cover500, col); gc()

load(" /cover1000gridFinal.Rdata")
cover1000 ¡- cover1000[cover1000$sum != 0,]
col ¡- colorRampPalette(c('lightgrey', 'darkgrey'))(150)
levelplot(sum x+y, data=cover1000, main='Rodentia species richness for 1000km
    grid',
xlab='Longitude [meters]', ylab='Latitude [meters]', col.regions=col)
trellis.focus("legend", side="right", clipp.off=TRUE, highlight=FALSE)
grid.text(expression(N(species)), 0.2, 0, hjust=0.5, vjust=1)
trellis.unfocus(); rm(cover1000, col); gc()
rm(cover.small, cover50, cover100, cover200, cover500, cover1000); rm()

#SAR
load(" /cover.small.Rdata")
```

```
attach(cover.small)
sum ¡- jitter(log(sum)); res ¡- log(res)
layout(matrix(c(1,2,3,4,5,5), 3, 2, byrow = TRUE), height=c(2.1,2.1,1))
plot(sum res, xlab="log(area)", ylab="log(species)", main="SAR for all data")
abline(1.92032, 0.07060, lty=1)
abline(1.5021, 0.1087, lty=2); abline(1.2606, 0.1131, lty=3)
abline(1.4196, 0.1072, lty=4); abline(1.4995, 0.1040, lty=5)
plot(sum res, xlab="log(area)", ylab="log(species)", main="SAR for 500 samples"
    )
abline(1.65745, 0.09933, lty=1)
abline(1.42183, 0.12139, lty=2); abline(1.39674,        0.12244, lty=3)
abline(1.34841, 0.12836, lty=4); abline(1.37689,        0.12490, lty=5)
plot(sum res, xlab="log(area)", ylab="log(species)", main="SAR for 1000 samples
    ")
abline(1.70964, 0.09604, lty=1)
abline(1.41287, 0.11977, lty=2); abline(1.43463,        0.11709, lty=3)
abline(1.2963, 0.1328, lty=4); abline(1.2923, 0.1332, lty=5)
plot(sum res, xlab="log(area)", ylab="log(species)", main="SAR for 5000 samples
    ")
abline(1.99132, 0.06108, lty=1)
abline(1.3845, 0.1089, lty=2); abline(1.3648, 0.1096, lty=3)
abline(1.2750, 0.1208, lty=4); abline(1.2741, 0.1208, lty=5)
par(mar=c(3,4,3,2))
plot(1, type = "n", axes=FALSE, xlab="", ylab="")
legend('center', c("without nesting", "nested to 500km scale", "nested 200km
    scale",
"nested to 100km scale", "nested to 50km scale"),
lty=c(1:5), ncol=3, cex=1.1, box.col="white")
plot(sum res, xlab="log(area)", ylab="log(species)", main="SAR for 10000
    samples")
abline(1.83780, 0.07960, lty=1)
abline(1.3699, 0.1118, lty=2); abline(1.3401, 0.1153, lty=3)
abline(1.3177, 0.1194, lty=4); abline(1.2798, 0.1235, lty=5)
plot(sum res, xlab="log(area)", ylab="log(species)", main="SAR for 15000
    samples")
abline(1.91137, 0.07123, lty=1)
abline(1.2958, 0.1152, lty=2); abline(1.2592, 0.1183, lty=3)
abline(1.2547, 0.1227, lty=4); abline(1.2278, 0.1263, lty=5)
plot(sum res, xlab="log(area)", ylab="log(species)", main="SAR for 18059
    samples")
abline(1.93654, 0.06834, lty=1)
abline(1.39072, 0.10372, lty=2); abline(1.3281, 0.1099, lty=3)
abline(1.3320, 0.1116, lty=4); abline(1.3149, 0.1138, lty=5)
plot(sum res, xlab="log(area)", ylab="log(species)", main="SAR for 36118
    samples")
abline(1.91056, 0.07212, lty=1)
abline(1.3068, 0.1084, lty=2); abline(1.2475, 0.1141, lty=3)
abline(1.3054, 0.1157, lty=4); abline(1.3299, 0.1152, lty=5)
par(mar=c(3,4,3,2))
plot(1, type = "n", axes=FALSE, xlab="", ylab="")
legend('center', c("without nesting", "nested to 500km scale",
"nested 200km scale", "nested to 100km scale", "nested to 50km scale"),
lty=c(1:5), ncol=3, cex=1.1, box.col="white")
par(mar=c(5.1,4.1,4.1,2.1)); rm(sum, res); gc()

#Znestedness
contiguous ¡- data.frame("nestedness"=c("without","500km","200km","100km","050
    km"),
"z"=c(0.07060, 0.1087, 0.1131, 0.1072, 0.1040))
sample500 ¡- data.frame("nestedness"=c("without","500km","200km","100km","050km
    "),
```

```
"z"=c(0.09933, 0.12139, 0.12244, 0.12836, 0.12490))
sample1000 <- data.frame("nestedness"=c("without","500km","200km","100km","050
    km"),
"z"=c(0.09604, 0.11977, 0.11709, 0.1328, 0.1332))
sample5000 <- data.frame("nestedness"=c("without","500km","200km","100km","050
    km"),
"z"=c(0.06108, 0.1089, 0.1096, 0.1208, 0.1208))
sample10000 <- data.frame("nestedness"=c("without","500km","200km","100km","050
    km"),
"z"=c(0.07960, 0.1118, 0.1153, 0.1194, 0.1235))
sample15000 <- data.frame("nestedness"=c("without","500km","200km","100km","050
    km"),
"z"=c(0.07123, 0.1152, 0.1183, 0.1227, 0.1263))
sample18059 <- data.frame("nestedness"=c("without","500km","200km","100km","050
    km"),
"z"=c(0.06834, 0.10372, 0.1099, 0.1116, 0.1138))
sample36118 <- data.frame("nestedness"=c("without","500km","200km","100km","050
    km"),
"z"=c(0.07212, 0.1084, 0.1141, 0.1157, 0.1152))
par(mfrow=c(2,2), mar=c(8,4.1,4.1,2.1))
plot(contiguous, ylab="z-value", main="all data", xaxt="n", xlab="")
axis(1, labels=FALSE); mtext(side = 1, text = "application of nestedness", line
    =6, cex=0.75)
text(x=contiguous$nestedness, y=par()$usr[3]-0.1*(par()$usr[4]-par()$usr[3]),
labels=contiguous$nestedness, srt=45, adj=1, xpd=TRUE)
plot(sample500, ylab="z-value", main="500 samples", xaxt="n", xlab="")
axis(1, labels=FALSE); mtext(side = 1, text = "application of nestedness", line
    =6, cex=0.75)
text(x=sample500$nestedness, y=par()$usr[3]-0.1*(par()$usr[4]-par()$usr[3]),
labels=sample500$nestedness, srt=45, adj=1, xpd=TRUE)
plot(sample1000, ylab="z-value", main="1000 samples", xaxt="n", xlab="")
axis(1, labels=FALSE); mtext(side = 1, text = "application of nestedness", line
    =6, cex=0.75)
text(x=sample1000$nestedness, y=par()$usr[3]-0.1*(par()$usr[4]-par()$usr[3]),
labels=sample1000$nestedness, srt=45, adj=1, xpd=TRUE)
plot(sample5000, ylab="z-value", main="5000 samples", xaxt="n", xlab="")
axis(1, labels=FALSE); mtext(side = 1, text = "application of nestedness", line
    =6, cex=0.75)
text(x=sample5000$nestedness, y=par()$usr[3]-0.1*(par()$usr[4]-par()$usr[3]),
labels=sample5000$nestedness, srt=45, adj=1, xpd=TRUE)
plot(sample10000, ylab="z-value", main="10000 samples", xaxt="n", xlab="")
axis(1, labels=FALSE); mtext(side = 1, text = "application of nestedness", line
    =6, cex=0.75)
text(x=sample10000$nestedness, y=par()$usr[3]-0.1*(par()$usr[4]-par()$usr[3]),
labels=sample10000$nestedness, srt=45, adj=1, xpd=TRUE)
plot(sample15000, ylab="z-value", main="15000 samples", xaxt="n", xlab="")
axis(1, labels=FALSE); mtext(side = 1, text = "application of nestedness", line
    =6, cex=0.75)
text(x=sample15000$nestedness, y=par()$usr[3]-0.1*(par()$usr[4]-par()$usr[3]),
labels=sample15000$nestedness, srt=45, adj=1, xpd=TRUE)
plot(sample18059, ylab="z-value", main="18059 samples", xaxt="n", xlab="")
axis(1, labels=FALSE); mtext(side = 1, text = "application of nestedness", line
    =6, cex=0.75)
text(x=sample18059$nestedness, y=par()$usr[3]-0.1*(par()$usr[4]-par()$usr[3]),
labels=sample18059$nestedness, srt=45, adj=1, xpd=TRUE)
plot(sample36118, ylab="z-value", main="36118 samples", xaxt="n", xlab="")
axis(1, labels=FALSE); mtext(side = 1, text = "application of nestedness", line
    =6, cex=0.75)
text(x=sample36118$nestedness, y=par()$usr[3]-0.1*(par()$usr[4]-par()$usr[3]),
labels=sample36118$nestedness, srt=45, adj=1, xpd=TRUE)
par(mfrow=c(1,1), mar=c(5.1, 4.1, 4.1, 2.1))
```

```
rm(contiguous, sample500, sample1000, sample5000, sample10000, sample15000,
    sample18059,sample36118); gc()

#Znot fully nested
data ¡- data.frame("data"=c("all data", "00500 samples", "01000 samples",
"05000 samples", "10000 samples", "15000 samples", "18059 samples", "36118
    samples"),
"non nested"=c(0.07060, 0.09933, 0.09604, 0.06108, 0.07960, 0.07123, 0.06834,
    0.07212),
"fully nested"=c(0.1040, 0.12490, 0.1332, 0.1208, 0.1235, 0.1263, 0.1138,
    0.1152))
par(mar=c(12,4.1,4.1,2.1))
plot(y=rep(0.3,nrow(data)), x=data$data, ylim=c(0,0.2), main="z-value for not
    and fully nested analysis", ylab="z-value", las=3, xaxt="n")
axis(1, at=data$data, labels=FALSE); mtext(side = 1, text = "dataset", line =
    9)
text(x=data$data, y=par()$usr[3]-0.1*(par()$usr[4]-par()$usr[3]), labels=data$
    data, srt=45, adj=1, xpd=TRUE)
points(data$data, data$non nested, pch=19); points(data$data, data$fully nested
    , pch=21)
legend('topright', c("fully nested", "not nested"), pch=c(21,19))
par(mar=c(5.1,4.1,4.1,2.1)); rm(data); gc()
```