Diversity and completeness of North American mammal assemblages

Joshua Aning – 3484645
e-mail: j.v.aning@students.uu.nl
tel: 06-18696672

Supervisor: dr. Mara Baudena, Utrecht University (in collaboration with dr. Miguel Angel Rodriguez, Alcalá University)

Abstract

Biodiversity can be defined in many ways. One of the more common ways of indicating an ecosystem's biodiversity is by measuring species richness; the number of different species that inhabit it. However, species richness follows a distinctive global pattern (generally decreasing from the tropics towards the poles), and gives only an absolute value for biodiversity. As such, while species richness may be useful for studying biogeographical patterns, it is not a particularly useful indicator to assess the health of ecosystems, let alone comparing the health of dissimilar ecosystems. For conservationists, indicators that measure how species-rich an ecosystem is relative to how species-rich it *could* be may be more useful. In this study, biodiversity patterns of a network of mammal assemblages (hexagonal cells of 7500 km^2) across extratropical North America were analyzed. The distribution patterns of species amongst the cells were used to calculate various biodiversity indicators.

The method of reflections was used to quantify the *potential* species richness of each cell, which was then compared to the *observed* species richness, yielding an indicator called the anomaly. Beal's probability index, a method that can be used to quantify with what probability a species might occur in a particular cell, was used to calculate each cell's *dark diversity*; the amount of species that are unexpectedly absent. The relationship between dark diversity and species richness yields the indicator *completeness*. Both *anomaly* and *completeness* relate to how species-rich a cell is relative to the potential that the network co-occurrence patterns of the species within the cell suggest; cells with a high anomaly or low completeness are far from their potential species richness. Unlike *completeness*, the anomaly does not require an explicit qualification of a cell's dark diversity and therefore does not need to make assumptions. Here we assess whether the indicator anomaly is a suitable substitute for the more established indicator completeness.

All biodiversity indicators were regressed against one another, and against models containing a variety of environmental factors commonly understood to affect biodiversity: precipitation, temperature, net primary productivity, elevation range, habitat homogeneity and human influence. We found that a combination of elevation range, mean annual precipitation and mean annual temperature was generally the most accurate predictor for these biodiversity indicators. Also, unlike species richness, which increased from the northeast to the southwest of the continent, *anomaly* and *completeness* were more closely tied to topographical features such as mountains and islands.

Introduction & Theory

Each species, plant or animal, has its unique distribution across the globe. Some species are cosmopolitan, and can be found nearly anywhere where conditions are suitable. Other species are more endemic, and occur only in a few particular areas, often despite the existence of other areas with similar ecological conditions[1]. This is not only the case on the global scale, but also on regional and local scales, such as continents, lakes, islands or patches of rainforests. Biogeography is the study of the distribution of organisms, and biodiversity, across geographic space. Patterns of biodiversity are influenced by various factors, such as climate, frequency of disturbance, productivity and biogeographic history[2]. Although it has been observed that, globally, biodiversity tends to follow a latitudinal gradient, and is correlated with climate (decreasing from the equator towards the poles)[3], local drivers of biodiversity are much less well understood. Global biodiversity is the result of the Earth's evolutionary history. Regional diversity is a subset of global diversity, and is shaped by migration, as well as speciation and extinction[4]. Local communities are in turn a subset of regional biodiversity filtered by local environmental conditions, and limited by dispersal as well as biotic interactions[5].

The distribution of species across a geographical region can be represented by a presence/absence matrix, which indicates which species occurs in which (local) area. Species within the region can be said to form a network, with each species being connected to the areas in which it occurs, and to the species with which it tends to co-occur. When considering such networks of species, ecologists are typically interested in what species occur where; that is, they tend to focus on the *presences* in an absence-presence matrix. However, a case can be made for considering absences as well. Not all absences are equal, and there are a variety of distinct reasons why a species should *not* occur in a particular area (here area is broadly defined as the smallest scale at which sampling takes place)[6]. For instance, a species can be absent from an area despite its suitable conditions, because it has never reached the region in which the area is situated; African rainforest species are not typically found in South American rainforests. In this scenario, it is biogeographical history that is the main limiting factor on a species' distribution. Conversely, a species may be present in a region, but absent in a given area, because the area does not meet the species' ecological needs. In such a case, environmental and ecological factors play the most important role.
The most interesting cases, however, are those where species are absent in an area while they do belong to the area's *species pool*. A species pool is a collection of species that are present in the region that, given their ecological requirements, would in theory be able to inhabit a particular area[7]. In other words, the species pool is the total regional (or *gamma*) diversity, filtered for the environmental conditions of the area. Part of the species pool will actually be present in the area; this is the area's observed (or realized) biodiversity. The remaining part of the species pool, which is not found in the area, is termed *dark diversity*[8]. Dark diversity is similar to, but distinct from, the concept of *beta diversity*, which quantifies the association between local (alpha) and regional (gamma) diversity; i.e. how much variation exists between local communities. By considering only those species whose ecological requirements are met by the area, dark diversity is more informative than beta diversity and can be used to quantify an area's relative richness[8].

There are various reasons why a species should belong to an area's dark, rather than observed, diversity. Within a given region, dispersal limitation plays an important role. Species with a low dispersal ability (e.g. few and/or heavy seeds in case of plants) are more prone to be a component of a particular site's dark diversity, as a result of their limited mobility. In case of plants, species that have low stress-tolerance to abiotic factors are also more likely to be a constituent of a site's dark diversity, as they are sensitive to stressors that can cause them to go extinct locally[9]. Biotic

interactions within local communities, such as cooperation and competition, may also play a major role[10]. Finally, human interference such as habitat fragmentation, can cause species to become locally extirpated, whilst remaining present in the region [11].

Dark diversity, together with species richness (i.e. *observed* diversity), can be used to quantify *assemblage completeness*. Completeness is an indicator of how much of an area's potential species richness is actually realized. It is defined as the natural logarithm of the observed-to-dark diversity ratio in the assemblage. Because it is the ratio between these two measures that determines completeness, its value is independent of the absolute species richness of the region. As such, this concept can be used to compare dissimilar ecosystems that may vary widely in terms of species richness (e.g. deserts, rivers and rainforests), and can be applied on any spatial scale[12]. It should be noted that this concept appears in the literature as *community completeness*[13,14]. However, as our analysis used large grid cells rather than small local communities, we have opted for the term *assemblage completeness*, or simply *completeness.*

Low assemblage completeness could indicate that an area has been disturbed, or is currently under stress. This allows for the possibility to identify vulnerable areas that could be targeted for conservation efforts. Moreover, quantifying dark diversity for a degraded area can also reveal how likely conservation efforts are to be successful. If the species that were previously present in the area are no longer a component of its dark diversity, it is unlikely they will return unless more drastic measures are taken[15]. Generally, degraded areas that score low on completeness (i.e. have a relatively large number of species as dark diversity) can be considered to have a high potential for restoration[8]. In this sense, *completeness* may be a more valuable metric to determine the success of conservation efforts than simply a change in species richness over time. And because it is a relative rather than absolute indicator, it may be used to compare patterns of biodiversity on a national and even continental scale.

This project investigates large-scale biodiversity patterns of terrestrial mammal assemblages in extratropical North America, and how the patterns for species richness differ from those of assemblage completeness. Mammals were selected as a focus group, as vertebrates in general are good indicators for the overall health of ecosystems[16], and mammals are a particularly well-documented and common subset of vertebrates. The local disappearance or decline of mammal populations may signify the loss of biological capital[17], which may have significant long-term ecological and economic consequences. By identifying mammal assemblages that are incomplete and understanding what they have in common, it may be possible to devise more effective conservation strategies than by looking at species richness alone.

A problematic issue, however, is how dark diversity may best be quantified. Dark diversity, by definition, cannot be directly observed. Various methods can be used to estimate it for a given area. The most straightforward way of finding out whether a species is able to live in an area would be to introduce it there and monitor it throughout its lifecycle. Of course, this method is not feasible for a variety of reasons: it is time-consuming, costly and laborious[9]; it is not easily applied on animals, which can move; and it may disrupt local ecosystems. A more common and feasible way of estimating the dark diversity of a large network of areas, is by inferring it from co-occurrence patterns of species throughout the network[18]. Species are not spread at random across the network, and similar species tend to occur together. These patterns enable one to calculate the probability of a species being able to live in a specific area (which may be accomplished using Beal's probability index), based on the other species that are present there[19,20]. While this is a useful method, it is difficult to precisely characterize an area's dark diversity, and different methods can yield dissimilar results[12,18]. Hence, a method that could quantify an assemblage's completeness without first needing

to qualify its dark diversity might be a real asset to conservation biologists. In this project one such method, the Method of Reflections[21], is investigated for its ability to do so.

The main research questions for this project are the following:

1. Is the Method of Reflections an appropriate method for yielding an indicator for completeness, without the need of explicitly identifying dark diversity size/composition? Are its completeness estimates similar to those yielded by Beal's index? If so, the Method of Reflections may be a useful alternative to such established methods, which yield absolute numbers for dark diversity that are subsequently used to calculate community completeness[22]. The Method of Reflections avoids the problem that these different methods can lead to quite different estimates of dark diversity (and therefore community completeness)[22].
2. What is the relationship between environmental factors (i.e. temperature, precipitation, net primary productivity, elevation range, habitat homogeneity and human influence index) on the one hand, and various biodiversity indices (i.e. observed and generalized species richness, the anomaly between the former, and assemblage completeness) on the other?

Methods

For this project, a large IUCN-dataset containing distribution data of terrestrial mammals across the Americas was used. While the dataset spanned all of the Americas, here only a part of North America was considered, spanning from the tropic of cancer to 60 °N. The dataset contained an estimated range of occurrence for every species[23].

In order to include environmental/climatic conditions and human impact in the analysis, various open-source databases and maps were used:

- The general human impact was quantified using the Last of the Wild Project,v2; a dataset produced by Wildlife Conservation Society (WCS) in collaboration with the Columbia University Center for International Earth Science Information Network (CIESIN). This dataset contains the Human Influence Index (hii), which is a measure of direct human influence on terrestrial ecosystems integrating data regarding human settlement (population density, urbanization), access (roads, railroads, rivers, coastline etc.), land use change and electric power infrastructure, on a 1-km^2 resolution[24].
- The EarthEnv project is a collaboration of biodiversity scientists and remote sensing experts, which produces various global, 1-km^2 resolution maps for monitoring biodiversity, ecosystems and climate. These maps were used for obtaining data on elevation range[25], habitat homogeneity and net primary productivity (in grams of C fixed/year)[26].
- For climate data, the open-source WorldClim database was utilized. This database also contains data for 1-km^2 pixels on various climatic variables[27]. Considered in this project were mean annual precipitation and mean annual temperature , two climatic factors that are frequently found to be important predictors of richness [28–30]

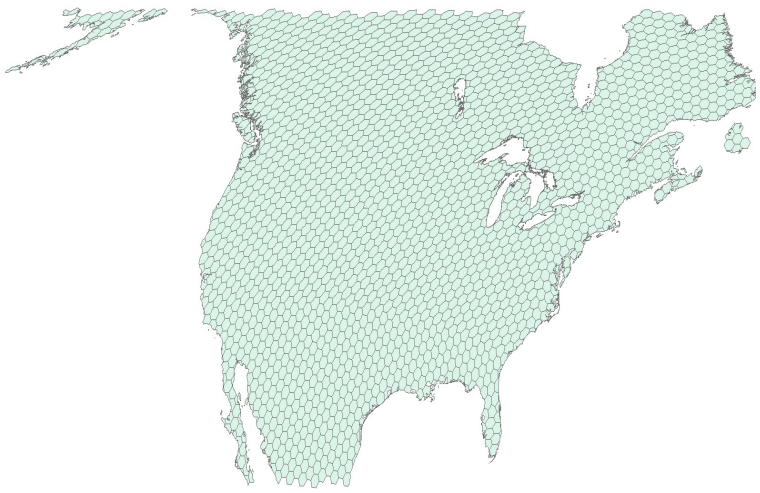


Figure 1: Map of grid cells included in the analysis

These data were overlaid on a ISEA3H regular grid with hexagonal cells with an area of about 7774 km^2 (for complete cells), which constituted the analysis units of the study. If a species' range included a given cell, the species was then counted as occurring within that cell. These data were used to construct a presence-absence matrix, containing the range of each species, as well as the species composition of every cell in the network. This presence-absence matrix was used as a basis for calculating the following biodiversity indices:

- observed species richness; quantifies the number of different species within an assemblage, the most "basic" measure of biodiversity
- generalized species richness; an index based on observed species richness that weighs species based on how widespread they are in the network. This index takes into account the pattern of species occurrences across assemblages throughout the study area.

- the anomaly, which we define (following [21]) as the discrepancy between normalized observed and generalized species richness; indicates assemblages that are more or less diverse (in terms of species richness) than would be expected based on their constituent species
- assemblage completeness; indicates how complete an assemblage is, i.e., how many of the species that could conceivably occur in it, actually do. This index is calculated using observed species richness and dark diversity

Additionally, ArcGis' zonal statistics tool was used in order to calculate the average value of each of the aforementioned environmental variables for each individual cell, except for elevation, for which the difference between the lowest and highest value (i.e. elevation range) was calculated.

Cells that lacked data on any of the environmental factors were excluded, as well as those that contained fewer than 10 species. This resulted in a total of 2093 cells.

Beal's Probability Index

Beal's Probability index[20] uses co-occurrences within a presence-absence matrix to calculate whether a species can be reasonably expected to belong to a cell's dark diversity or not. It was calculated using the following formula[20]:

$$P_{ij} = \frac{1}{S_i - I_{ij}} \sum_{k \neq j} \frac{N_{jk} I_{ik}}{N_k}$$

where: P_{ij} = Beal's probability index for species j at community i
S_i = number of species at community i
I_{ij} = incidence (0,1) of species j within community i
N_{jk} = number of joint occurrences of species j and k throughout the dataset
I_{ik} = incidence (0,1) of species k at community i
N_k = number of occurrences of species k throughout the dataset

Because the Beal's probability index for each cell/species combination depends strongly on how common the species is in the entire network (i.e. frequent species are systematically assigned higher Beal index values than uncommon species), probability attribution was based on the cumulative normal distribution of Beal's indices for all the cells in which a certain species *j* occurs. Unoccupied cells (i.e. cells in which species *j* was absent) were assigned a probability of occurrence for species *j* based on the cumulative percentile of the Beal's index value of the cell relative to the Beal's values of the cells in which species *j* does occur. The lowest Beal index value among the cells in which species *j* occurs is set as the absolute minimum; unoccupied cells that scored a Beal's index lower than the lowest value observed in the occupied cells, were assigned a probability of occurrence of 0%. Otherwise, they were assigned a probability of occurrence based on the cumulative percentile of Beal's values of the occupied cells. For example; if an unoccupied cell had a Beal's index for species *j* that was higher than 10% of the occupied cells, it was assigned a probability of occurrence of 0.1. If its Beal's index was higher than 20% of the occupied cells, it was assigned a probability of 0.2, etc.

In summary, Beal's index was calculated for every species/cell-combination. Then, for any given species, the cumulative normal distribution of Beal's indices among the cells in which the species

occurred, was used to assign a probability of occurrence for that species in any of the cells in which it did not. This probability of occurrence was then compared to a threshold value of 0.5. Any absent species that had a higher probability of occurrence in a cell than 0.5, was considered as part of that cell's dark diversity. The total number of species compromising a cell's dark diversity was used to calculate that cell's completeness, using the following relationship:
Community completeness = ln (observed diversity / dark diversity (+eps))
where eps is a small number (0.1 in this case) that prevents infinite values for completeness in cells with zero dark diversity.

Method of Reflections

While the Beal's probability index depends on estimating the composition of dark diversity for every cell, a different method that can be used to estimate how close a community (or in this case, a cell's mammal assemblage) is to its potential richness, the method of reflections[21], does not. Rather, the method of reflections yields a proxy for the *potential* richness of each cell by assigning weights to species based on how 'sociable' they are; that is, how many other species they on average occur with throughout the network (or rather: how diverse the cells within which they occur are on average).

The method of reflections is a network method that takes as its input a presence-absence matrix, and produces two sets of values, or reflections, which relate to species richness of the cells and species' range of occurrence. The zeroth order reflections were produced by summing the number of species for every cell in the network ($k_{s,0}$; species richness) and the number of cells inhabited for every species ($k_{p,0}$; species range). Species richness ($k_{s,0}$) is itself a commonly used measurement of biodiversity[31], and was also included as a biodiversity index. These zeroth order reflections were used to generate higher order reflections of the opposite sign through averaging. The first order cell reflections were calculated by averaging the range (i.e. the total number of cells inhabited) of all the species occurring within each particular cell. The first order species reflections are calculated by averaging the richness of all the cells in which a particular species occurs. This process is repeated iteratively to produce higher-order reflections of site diversity and species range. The repeated averaging removes environmental noise and results in a value of *generalized species richness* for every cell and a value of *generalized range of occurrence* for every species (the latter was not used in this project). Here, the 18th order reflection, $k_{s,18}$, was used, as higher-order reflections do not yield significantly different outcomes in terms of the generalized species richness ranking of the cells[21]. It should be noted that *generalized* or *potential* richness is not an estimate of how many species might be expected to inhabit the cell under optimal circumstances in an absolute sense. Instead, the higher-order reflections of all cells cluster around an average value. The small differences in values between the cells reveal the differences in generalized species richness between these cells when environmental noise has been largely removed through

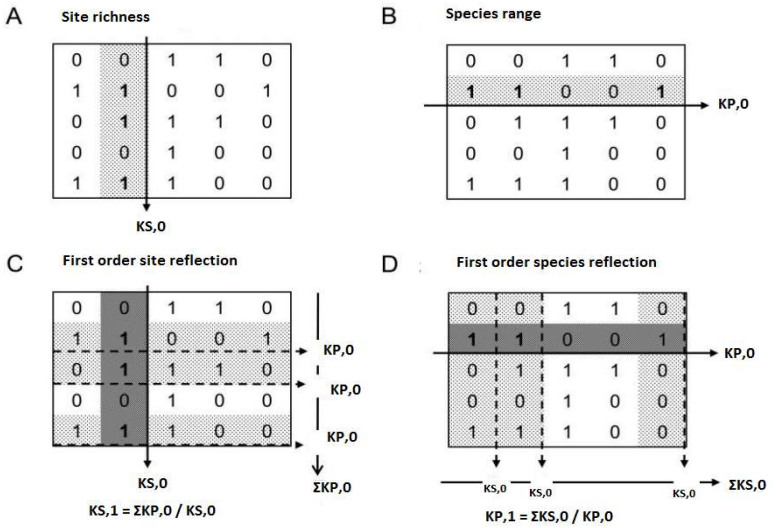


Figure 2: a schematic representation of the method of reflections. Rows represent species, columns represent sites.

averaging. Thus, cells that have a low observed species richness can nevertheless have a high generalized richness if the species occurring within them normally tend to occur together with many other species. Conversely, cells that are very diverse will score low on generalized richness, if the species they contain tend to live in low-diversity cells.

After applying the method of reflections, the observed and generalized species richness of the cells were normalized, such that each cell was assigned a score between 0 and 1 for both of these indices. This was done by subtracting the minimum value and dividing by the difference between maximum and minimum values. The normalized observed species richness was subtracted from the normalized generalized species richness to quantify the discrepancy between the two. This derived variable is referred to as the *anomaly.* High values indicate sites that are anomalous, in the sense that observed species richness is lower than generalized species richness would suggest. Conversely, negative values indicate sites that have a higher species richness than expected. Thus, the anomaly may be used as an indicator for how far removed a cell is from its potential; ostensibly comparable to calculating assemblage completeness from dark diversity and species richness[21]

Stepwise generalized linear models

The next step was to quantify to what extent observed/generalized species richness, anomaly and community completeness could be explained by the selected explanatory variables. In order to do so, stepwise linear regressions were used to formulate models, which were then evaluated. The Akaike Information Criterion (AIC) was used in order to rank the quality of each model. The AIC assigns score models based on their predictive power (i.e. coefficient of determination; R^2) and simplicity/parsimony, penalizing overly complex models that use many variables or higher order relationships[25,] .

For each model containing only a single predictor (either an environmental variable, or another biodiversity index), three models were considered: linear, quadratic or cubic. For multi-variable models, only linear (additive) relationships were considered, for the sake of simplicity, thus excluding interactions, i.e. the products of independent variables. Independent variables were only included in models together if they had a coefficient of variation (R) lower than 0.5 between them.

For every biodiversity index, models were assigned weights based on their distance from the model with the lowest AIC-score (Δ), using the following formula[33]

$$w_i = \frac{\exp(-\frac{1}{2}\Delta_i)}{\sum_{n=1}^{N} \exp(-\frac{1}{2}\Delta_n)}$$

where w_i = Akaike weight of model i

Δ = distance to the lowest AIC-score

N = the number of different models considered

The weights are normalized between 0 and 1 and together sum up to 1. An Akaike weight indicates the probability that a given model is the best considered. The higher the score, the higher the probability that the model is the best one.

Results

Correlations between biodiversity indices

The following section discusses how the various biodiversity indices correlate with one another. Observed species richness increased with generalized species richness ($R^2 = 0.69$; figure 3, upper graph). Strong deviations of this relationship indicate anomalous cells, which are either less or more species-rich than might be expected from the species that occur there. However, while cells containing many species on average tended to have lower anomaly scores, anomaly was not strongly dependent on observed species richness ($R^2 = 0.03$; graph not shown).

Community completeness, on the other hand, strongly correlated with species richness ($R^2 = 0.39$; figure 3, lower graph). The distinctive "bands" in the graph represent cells that have the same number of missing (dark diversity) species. Assemblage completeness responded more strongly to changes in dark diversity than it did to changes in observed species richness. In theory, an increase in species richness will decrease the number of missing species by an equal amount.

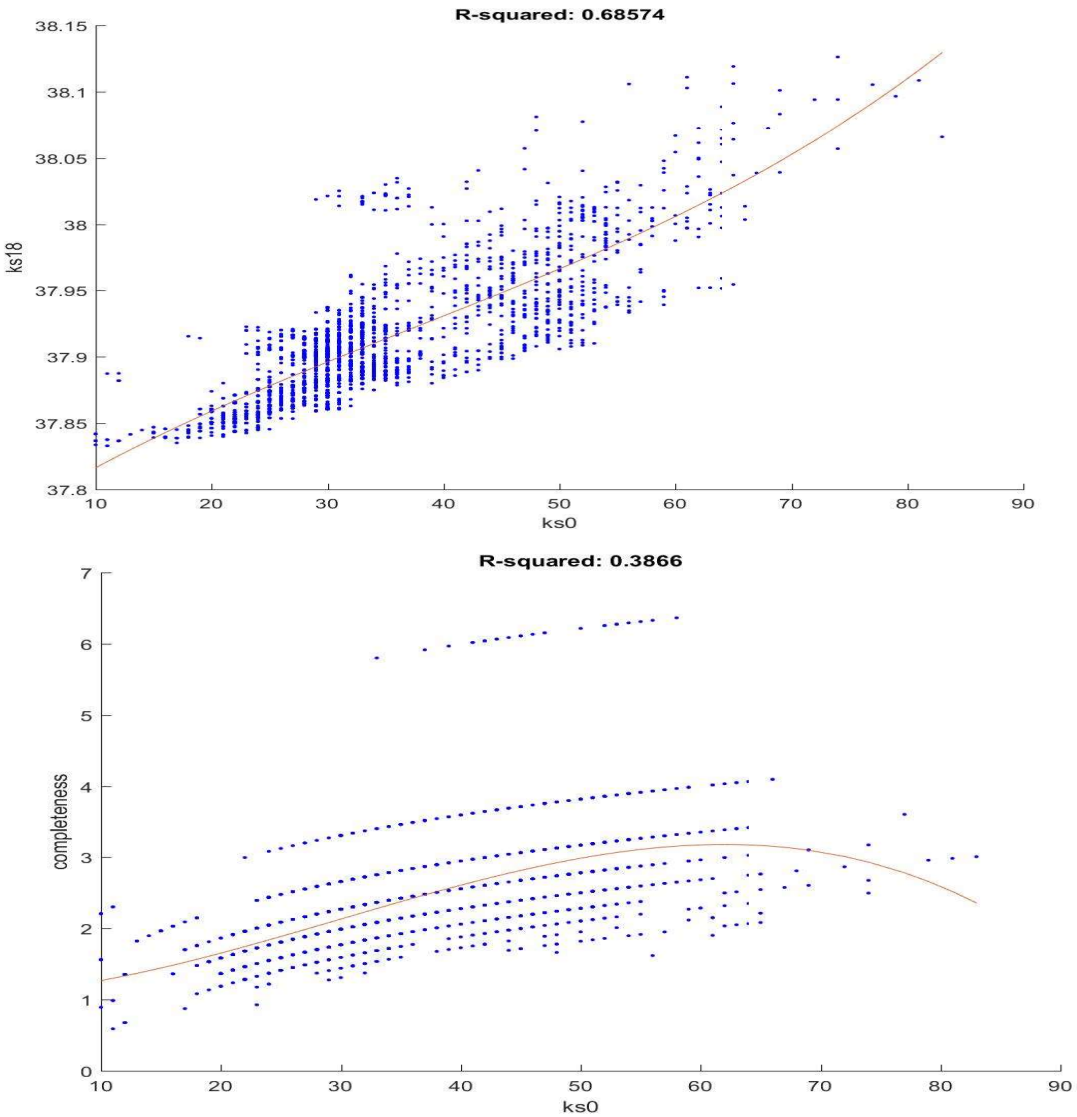


Figure 3: These graphs display how generalized species richness ($k_{s,18}$, top) and community completeness (bottom) respond to observed species richness ($k_{s,0}$).

The anomaly and assemblage completeness showed a cubic relationship (R^2 = 0.21; figure 4, upper graph). While completeness does, on average, tend to decrease with large anomaly-values, the variance explained is quite low. Thus, the latter cannot be used to accurately predict the former.

Completeness increased linearly with generalized species richness, albeit with a lower coefficient of determination (R^2 = 0.11; figure 4, lower graph)

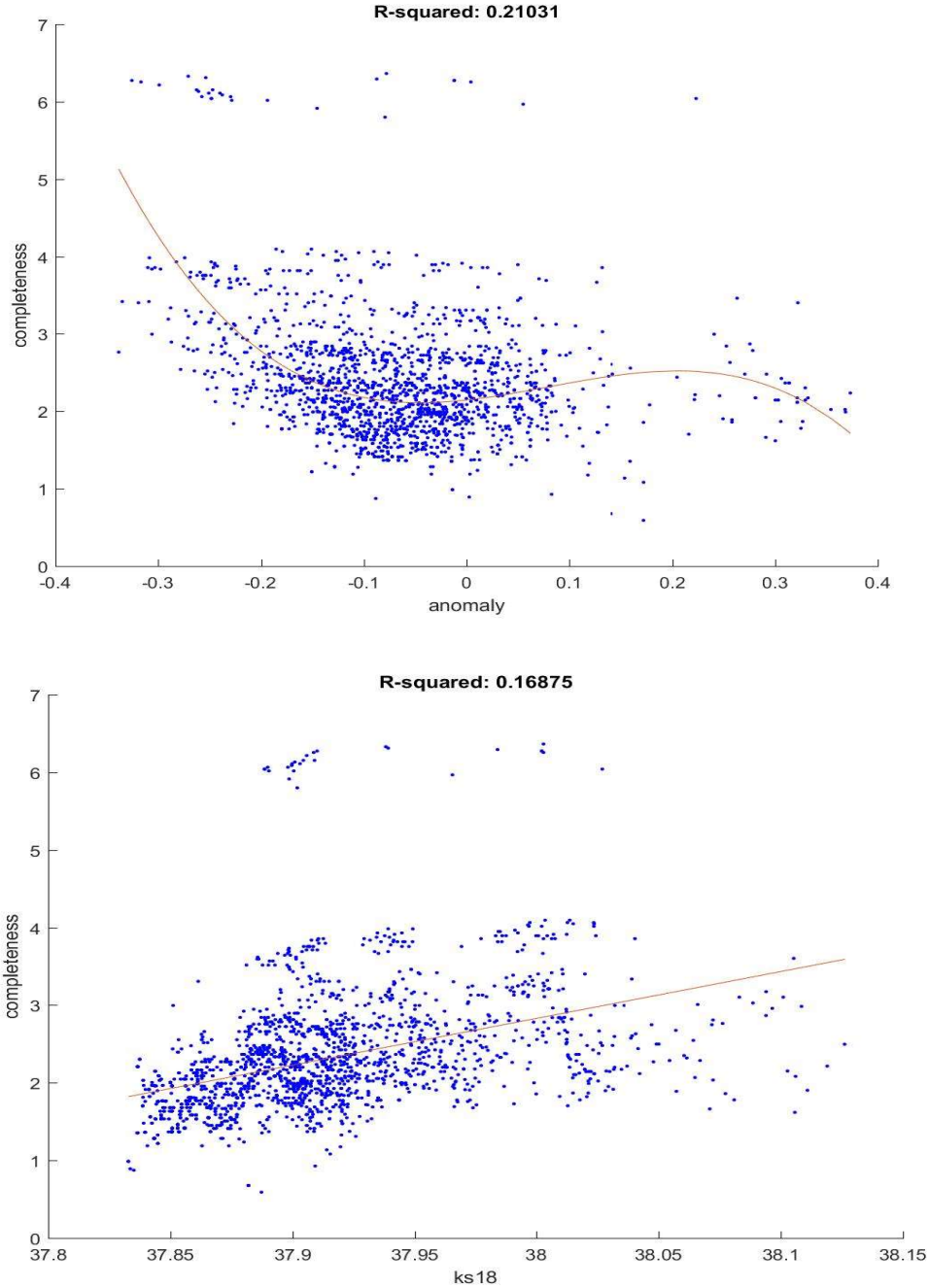


Figure 4: These graphs display how community completeness relate with *anomaly* (top) and $k_{s,18}$ (bottom)

Biodiversity indices across extratropical north America

Strong latitudinal and longitudinal gradients were found for both observed and generalized species richness, especially the latter (figure 5). Specifically, both measures of species richness increased from north to south, but also from east to west. To an extent, the latter took precedence over the former; most of the northwest (e.g. Oregon) was richer in species than the southeast (e.g. Florida). For observed species richness, there was a clear and sudden divide separating east and west, coinciding roughly with the frontier between the Rocky Mountains and the Great Plains. The anomaly and community completeness showed patterns which differentiate them from observed and generalized species richness. While the anomaly and community completeness "disagreed" on many cells, the cells with the lowest anomalies tended to be the same as those with the highest completeness, and vice versa. This pattern was most apparent in major mountain ranges (i.e. the Sierra Nevada, Rocky Mountains and Appalachian mountains), and, to a lesser extent, along the Gulf Coast and the Great Plains.

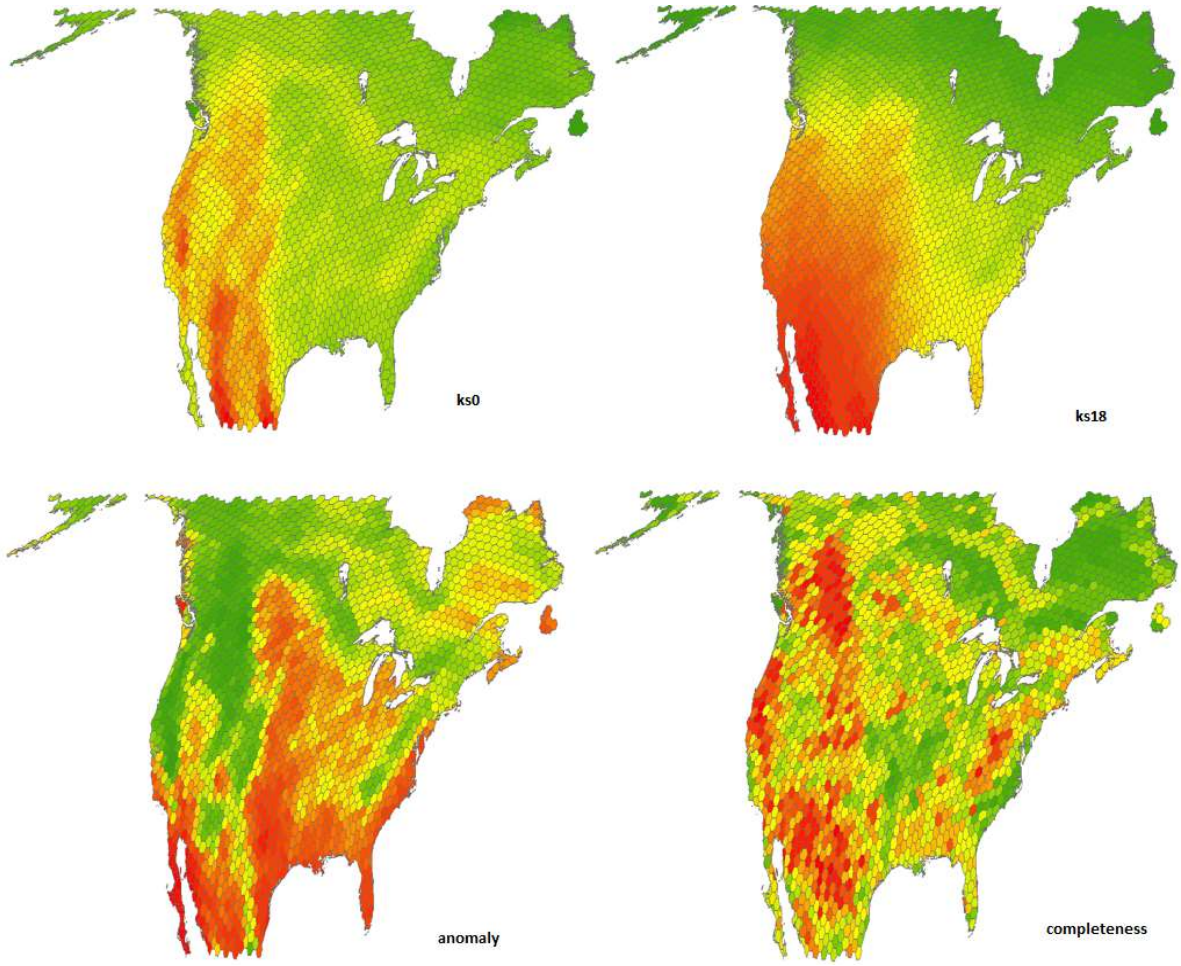


Figure 4: These maps indicate the values of the aforementioned biodiversity indices of mammals across North America. From top-left going clockwise: species richness (ks0), generalized species richness (ks18), community completeness and anomaly between ks18 and ks0. Warmer colors indicate higher values. Note the strong north-east to south-west gradient in observed and especially generalized species richness

Correlations between environmental variables

Linear regressions were performed on the explanatory variables in order to ascertain their independence from one another, as shown in figure 6. Those variables that had a coefficient of correlation above 0.5 (or below -0.5) were not included in models together (following ref [34]), namely:

- Temperature and human influence index
- Precipitation and net primary productivity
- Precipitation and habitat homogeneity
- Net primary productivity and human influence index

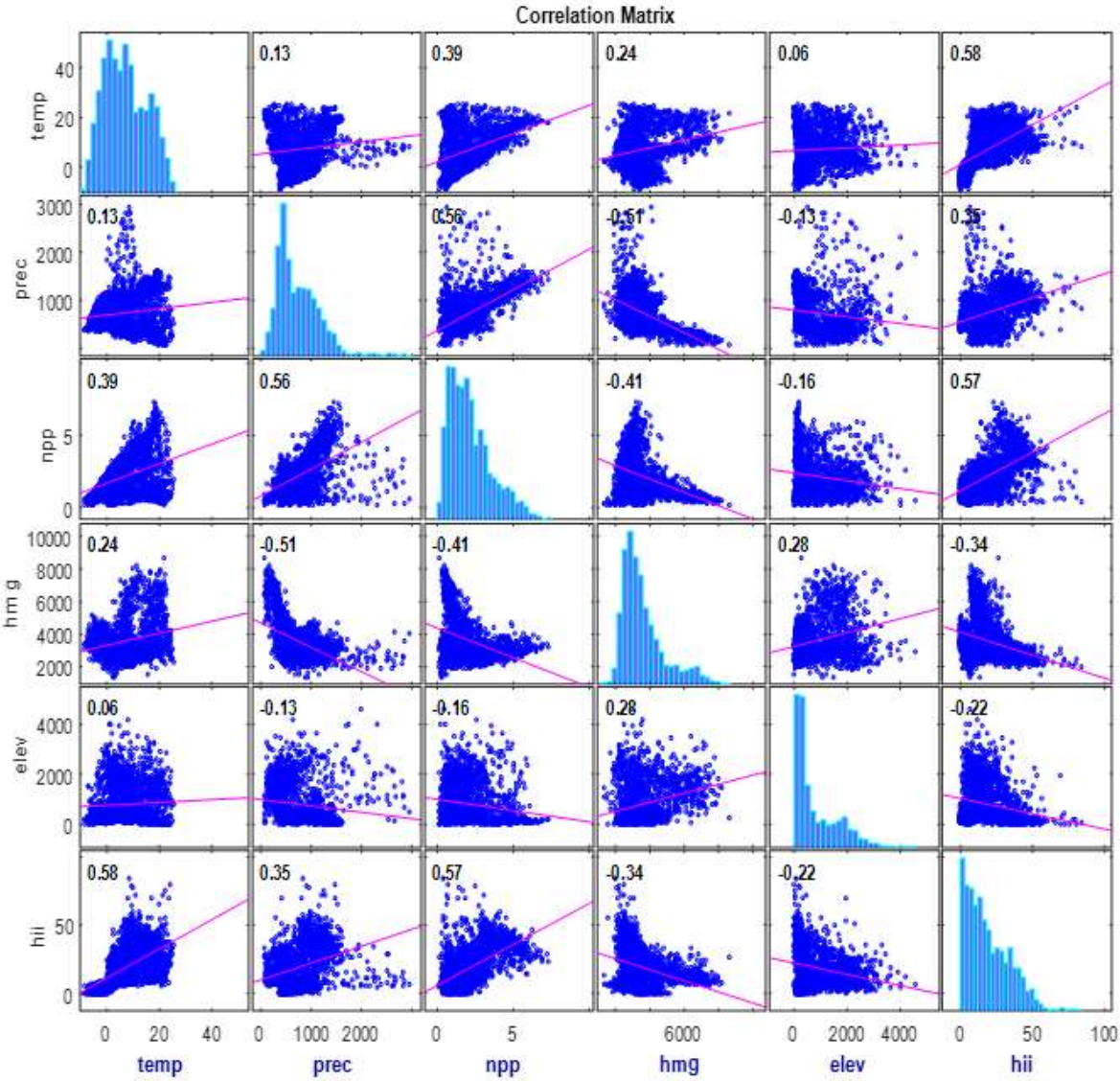


Figure 5: A correlation matrix showing the relationships between the six environmental variables considered. The diagonal shows the distribution of each variable. Temp = mean annual Temperature (°C), prec = mean annual precipitation (mm/y), npp = net primary productivity (C/g/year), hmg = habitat homogeneity, elev = elevation range (m), hii = human influence index

Correlations between environmental variables and biodiversity indices

The following sections discuss how the various environmental variables relate to each of the biodiversity indices.

Temperature

Each of the biodiversity indices increased with temperature (figure 7). Both observed species richness and completeness increase steeply with temperature in colder areas, but this effect is less marked in temperate zones. Anomaly also generally increases with increasing temperature.

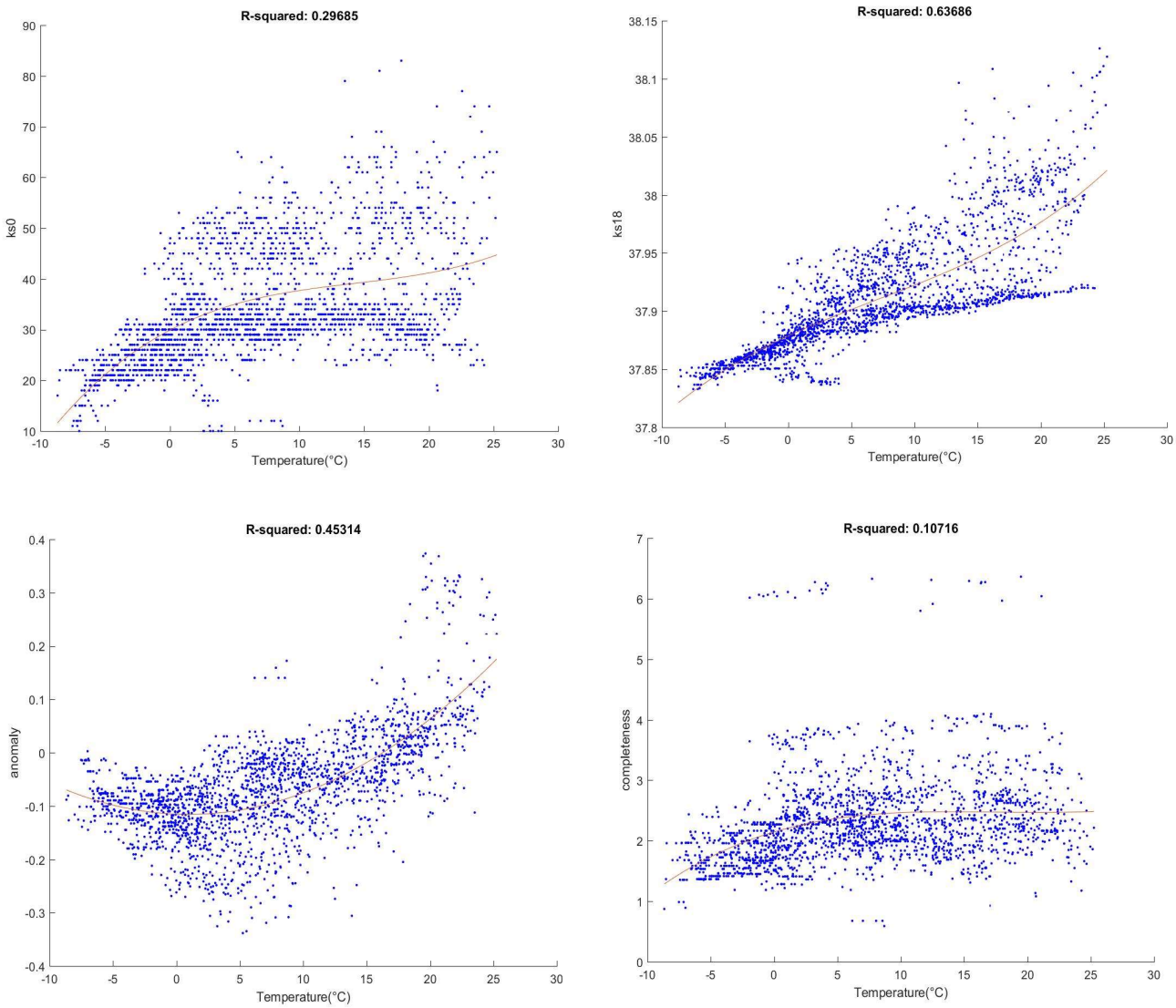


Figure 6: biodiversity variables as a function of temperature. From top left going clockwise: species richness ($k_{s,0}$), generalized species richness ($k_{s,18}$), completeness and anomaly

Precipitation

Precipitation shows a negative relationship with observed and potential species richness as well as completeness throughout most of the dataset, and this relationship is strongest in areas that receive little precipitation (figure 8). Above 500 mm/year, however, diversity increases slightly with precipitation.

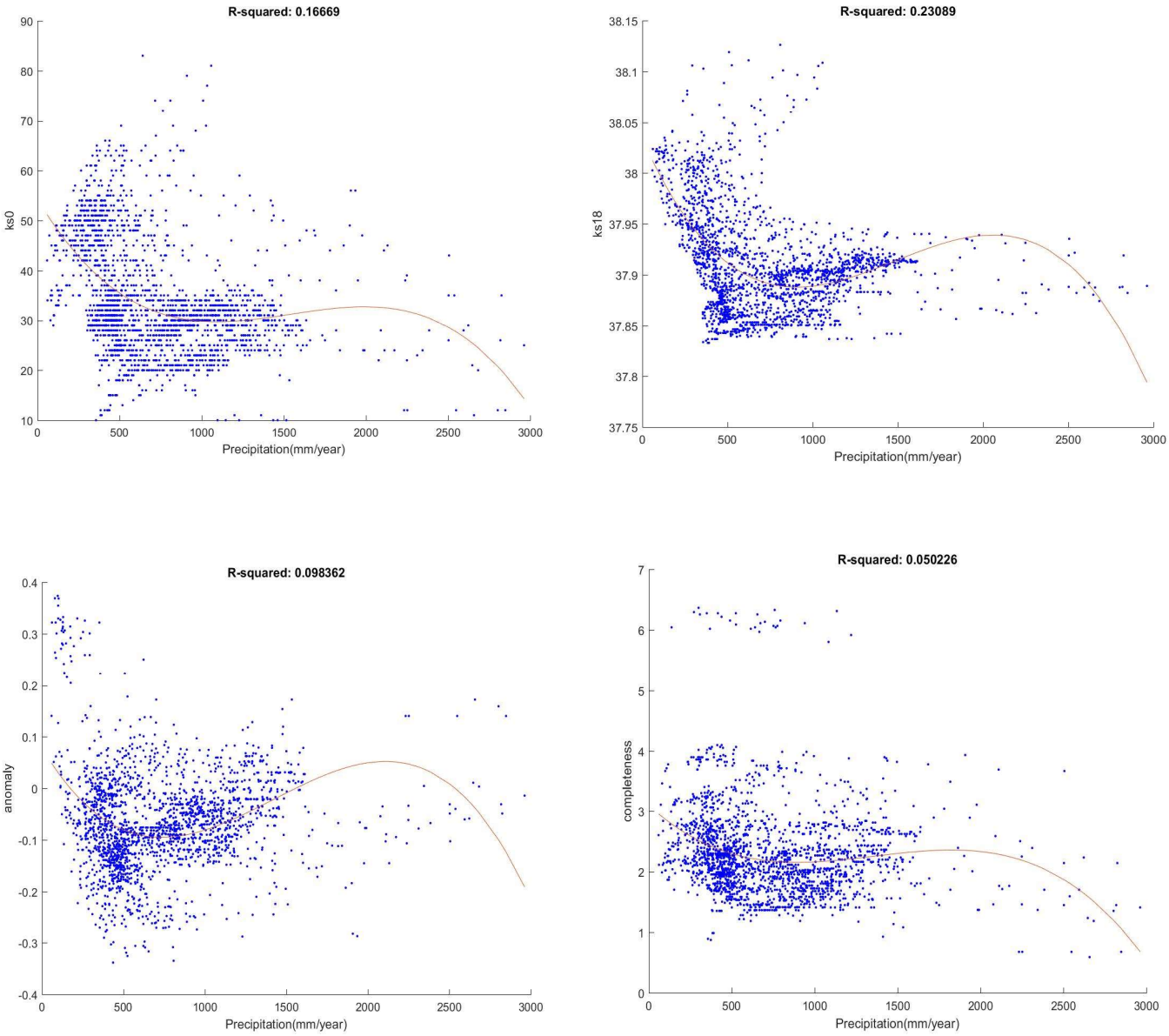


Figure 7: biodiversity variables as a function of precipitation. From top left going clockwise: species richness ($k_{s,0}$), generalized species richness ($k_{s,18}$), completeness and anomaly

Net Primary Productivity

No correlation was found between Net Primary Productivity and any of the biodiversity indices (not shown).

Habitat Homogeneity

Both observed and potential species richness show a positive overall correlation with habitat homogeneity, but, especially for potential species richness, parts of the dataset seem to show the exact opposite relationship, with diversity decreasing with homogeneity (figure 9). Anomaly and completeness were not explained by habitat homogeneity to a large degree ($R^2 < 0.05$; not shown).

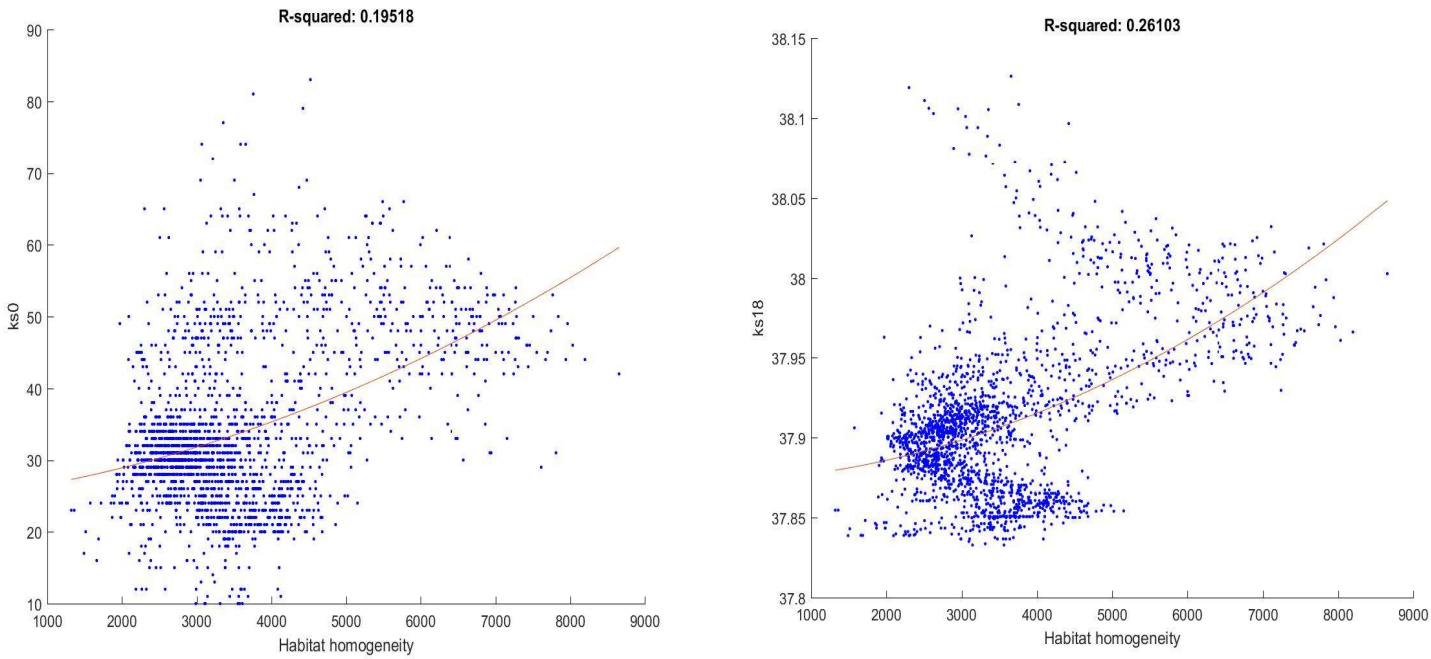


Figure 8: biodiversity variables as a function of habitat homogeneity. Left: observed species richness ($_{ks,0}$), Right: generalized species richness ($_{ks,18}$)

Elevation range

Elevation range showed a strong positive correlation with observed species richness, and to a lesser extent with potential species richness as well (figure 10). Cells with a wide range in elevation (i.e. variation in altitude), were less anomalous and scored higher on community completeness.

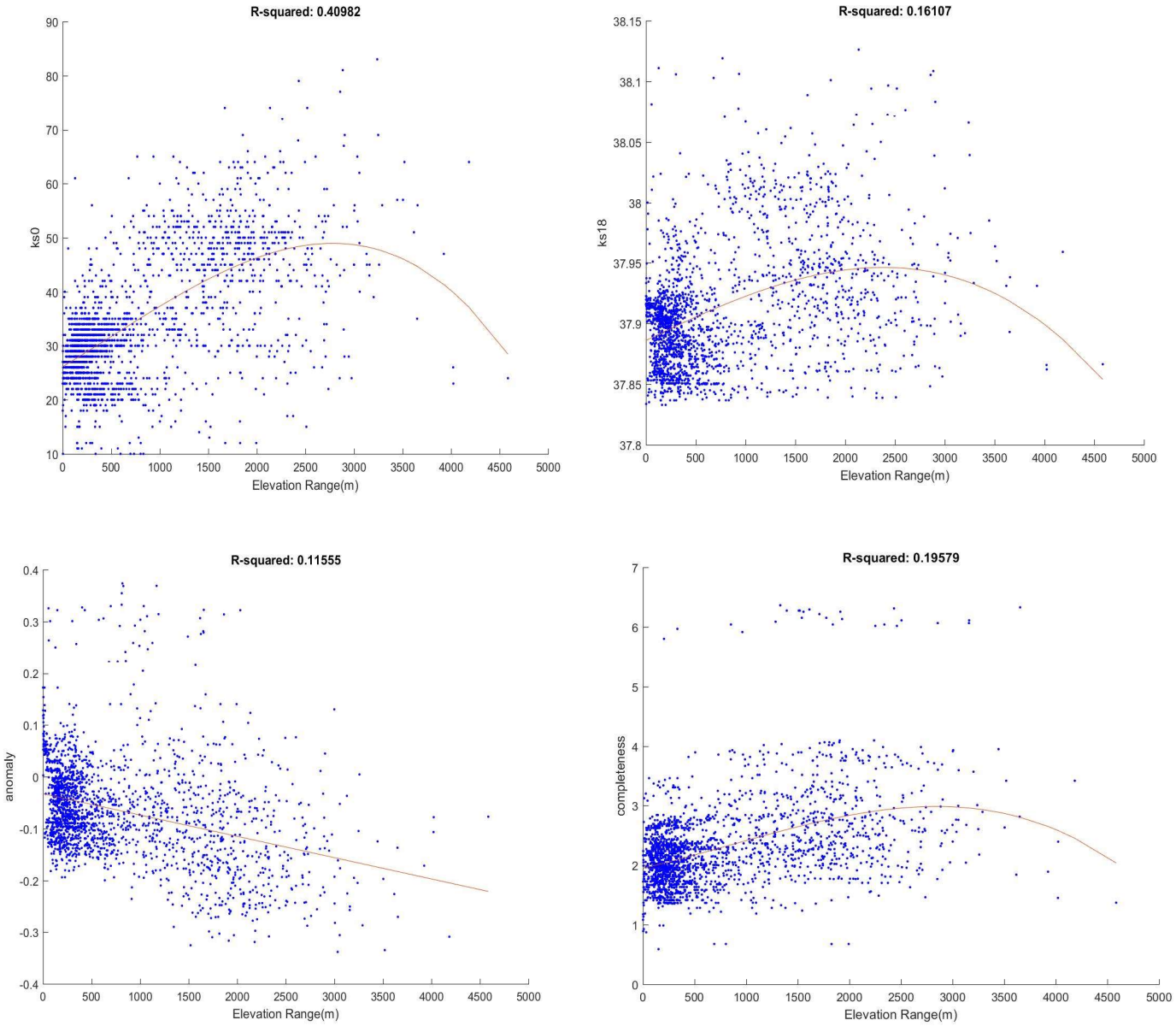


Figure 9: biodiversity variables as a function of elevation range. From top left going clockwise: species richness ($k_{s,0}$), generalized species richness ($k_{s,18}$), completeness and anomaly

Human Influence Index

Observed and potential species richness as well as completeness, showed a parabolic relationship with human influence (figure 11); the highest value for each of these indices was at intermediate HII-values. Moreover, the lowest species richness values were in cells with the lowest HII-scores. The anomaly slightly increased with increasing hii-values, but without much variance explained.

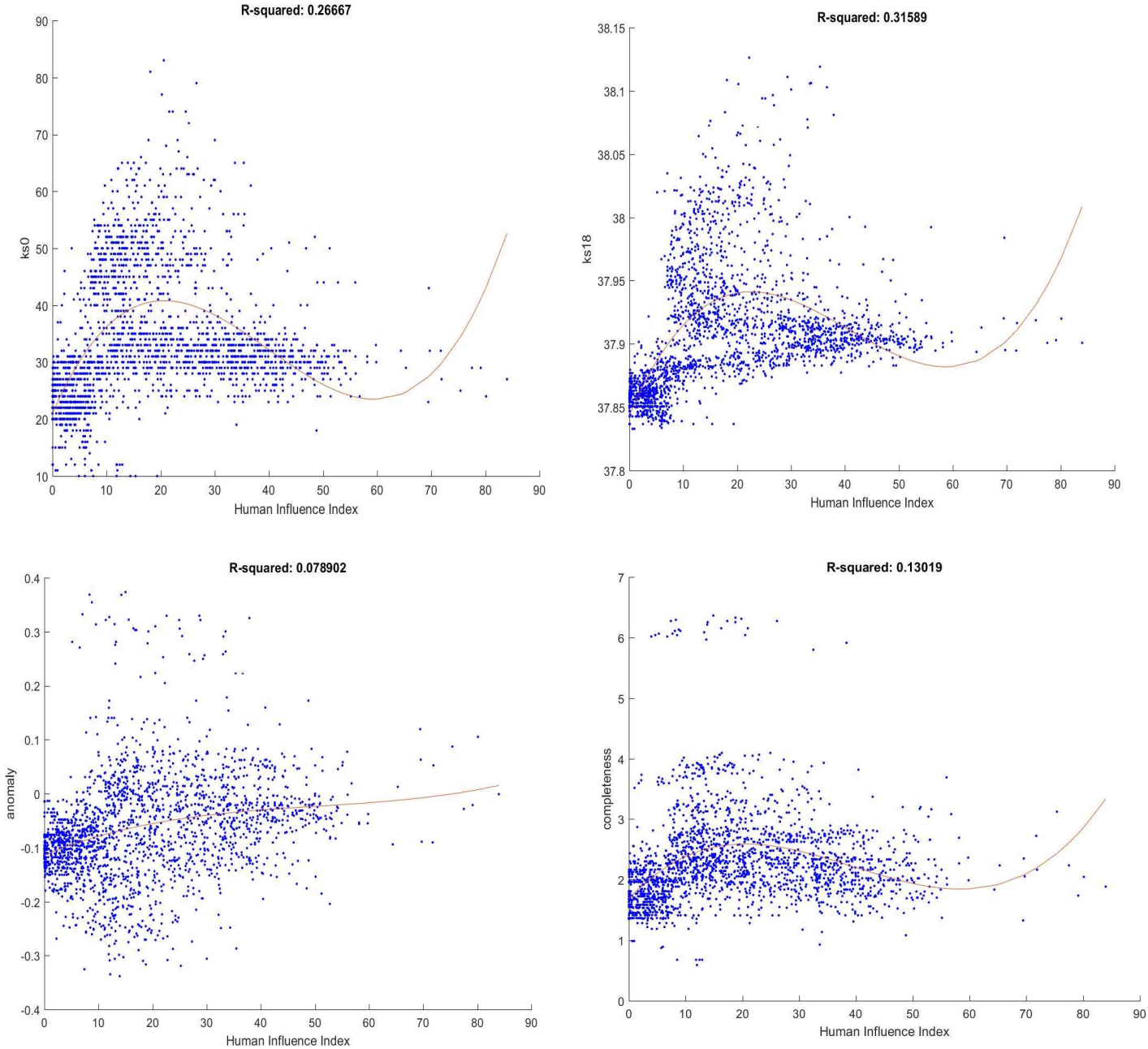


Figure 10: biodiversity variables as a function of human influence index. From top left going clockwise: species richness ($k_{s,0}$), generalized species richness ($k_{s,18}$), completeness and anomaly

Multivariate regressions

Table 1 displays the R^2-values as well as the AIC-values of generalized linear models for each of the explanatory variables. Also included were the 5 combined models with the lowest AIC-values. The delta of each model shows the difference in AIC-score with the lowest model (Δ), and the weight indicates the odds of it being the best models from amongst the total of 25 models that were considered. For model formulae, see appendix 2.

Table 1: The R^2-values, AIC-values, deltas and Akaike weights of all single-variable models and the 5 best multi-variable models for the 4 biodiversity indices. Temp = temperature, prec = precipitation, npp = net primary productivity, hmg = habitat homogeneity, elev = elevation range, hii = human influence index.

	R2	AIC	Δ	weight
Ks0				
Temp	0.30	15341	1803	0
Precipitation	0.17	15696	2158	0
NPP	0.00	16073	2534	0
Homogen	0.20	15623	2084	0
Elevation	0.41	14974	1436	0
Hii	0.27	15429	1891	0
Temp + prec + elev	0.70	13538	0	1
Temp + npp + hmg + elev	0.64	13955	416	0
Temp + hmg + elev	0.64	13966	428	0
Temp + npp + elev	0.61	14116	578	0
Temp + elev	0.61	14129	591	0
Ks18				
Temp	0.64	-8742	1799	0
Precipitation	0.23	-7171	3369	0
NPP	0.00	-6626	3915	0
Homogen	0.26	-7256	3285	0
Elevation	0.16	-6990	3551	0
Hii	0.32	-7417	3124	0
Temp + prec + elev	0.85	-10541	0	1
Temp + npp + hmg + elev	0.81	-10130	411	0
Temp + hmg + elev	0.80	-9964	576	0
Temp + npp + elev	0.79	-9929	611	0
Temp + prec	0.77	-9708	833	0
Anomaly				
Temp	0.45	-5120	628	0
Precipitation	0.10	-4073	1675	0
NPP	0.00	-3859	1889	0
Homogen	0.04	-3935	1812	0
Elevation	0.12	-4115	1633	0
Hii	0.08	-4028	1720	0
Temp + npp + hmg + elev	0.60	-5748	0	0.85
Temp + npp + elev	0.59	-5744	3	0.15
Temp + hmg + elev	0.51	-5335	413	0

Temp + prec + elev	0.49	-5263	484	0
Temp + elev	0.48	-5239	509	0
Completeness				
Temp	0.11	4382	404	0
Precipitation	0.05	4512	533	0
NPP	0.01	4601	622	0
Homogen	0.05	4520	541	0
Elevation	0.20	4163	185	0
Hii	0.13	4328	349	0
Temp + prec + elev	0.26	3979	0	0.998
Temp + npp + hmg + elev	0.26	3996	18	0.002
Temp + npp + elev	0.25	4016	38	0
Temp + hmg + elev	0.25	4024	45	0
Temp + elev	0.25	4026	47	0

A combination of temperature, elevation range and precipitation appears to yield the model which, among the models considered, best explains most of the variation in biodiversity for observed and generalized species richness, as well as completeness with very high probability (Akaike weight ≈ 1). For the anomaly, this combination of explanatory variables ranked fourth (Akaike weight ≈ 0) and was outdone by models that likewise included temperature and elevation, but combined with net primary productivity and/or habitat homogeneity. The model combining temperature and elevation was the only 2-variable model that was amongst the best 5 models for each biodiversity index except for generalized species richness. None of the five multi-variable models with the lowest AIC-values included the human influence index.

Discussion

Temperature had a strong influence on each of the biodiversity indices. It is a well-documented observation that biodiversity tends to decrease from the equator towards the poles (as does temperature). This general pattern seems to hold for mammals as well[35]. However, the mechanics driving this process are still up for debate. One theory is that temperature affects biodiversity mostly through its effect on plant productivity[36]. Our results, however, showed that net primary productivity explains only a very small percentage of variance in mammal diversity in North America. This decoupling of productivity and species richness may be a result of the relatively recent extinction event that occurred in the Americas around 13000 years ago. With the advent of the Holocene, global average temperatures suddenly rose steeply, which resulted in the extinction of many terrestrial mammal species that were unable to accommodate the rapid rate of change. The decline in species richness was probably exacerbated by the arrival of early humans into the Americas[37]. This widespread and relatively sudden dying out of species may have resulted in an under-saturation of species from which the Americas have yet to recover[36], leaving the current assortment of species relatively impoverished. As such, the relationship between plant productivity and mammal species richness is less pronounced in the Americas than in other parts of the world, and the strong effect of temperature on biodiversity cannot be attributed to plant productivity alone. Another factor, strongly correlated with temperature, that may influence mammal diversity, is seasonality. Higher latitudes have lower mean annual temperatures, as well as more profound seasonal differences. Such conditions require specific adaptations that make higher latitudes increasingly inhospitable for generalist species. In addition to this, species inhabiting temperate and polar regions tend to have/need larger geographical ranges, such that species turnover (beta diversity) is lower than in tropical regions[38]. Finally, low (winter) temperatures may to a large extent exclude small mammal species, which, due to their high surface-to-volume ratio, need a lot of energy to maintain their body heat[39]. Since small species constitute the majority of mammalian diversity[40], colder regions host a more modest range of mammal species than warmer ones.

Like temperature, elevation range also scored high R^2-values and low AIC-values when regressed to biodiversity indices. Part of the reason that a larger elevation range corresponds to higher species richness (and also higher completeness and lower anomaly), may be that cells with a large range in elevation on average also included a wider variety of habitat types (e.g. valleys, montane forest, mountain peaks), and therefore a larger number of species corresponding to those habitats. Moreover, many groups of small non-volant mammals (e.g. rodents, shrews, etc.) actually tend to peak in species richness at intermediate altitudes in mountains[41,42]. For these groups the highest levels of species richness consequently correspond to cells with high elevation ranges (because these cells tend to contain mountains). In addition to hosting a variety of habitats, the presence of mountains also influences the climate in complex ways, which in turn affects patterns of biodiversity. Most obviously, average temperatures are lower in mountainous regions as temperature decreases with altitude. High mountains also affect the movement of clouds which can result in more diverse precipitation patterns than in flat environments. Models that combined elevation range with climate seemed to be most suited for predicting biodiversity (between the models considered). For observed species richness, anomaly and community completeness, a combination of temperature and elevation yielded the best two-variable model. For generalized species richness, it was a combination of temperature and precipitation. This may indicate that of these indices, generalized species richness is most sensitive to purely climatic (rather than topographical) factors. Likely, this is because the method of reflection detects large-scale patterns in species richness, and climate is the most global factor influencing biodiversity.

Surprisingly, there did not appear to be a strong negative effect of human influence on biodiversity, except that anomaly tended to increase slightly ($R^2 = 0.08$) with increasing hii-values. Rather, observed/generalized species richness and community completeness tended to peak at hii-values between 10 and 20, which corresponds mostly with cells in the western mountain ranges. The cells with the lowest hii-values (i.e. lower than 10), were primarily situated in northern cells with low average annual temperatures. Since humans tend to favor warm temperatures over cold ones, it is difficult to separate actual human impact from temperature. Another issue in ascertaining the effect of human influence is the size of the grid cells. As most of the cells were around 7500 km^2 in size, they are large enough to include areas heavily altered by humans alongside relatively pristine ones. For instance, a cell that contains a large city and an adjacent national park, will have a high hii-value as well as high species richness. More importantly, the data used in this project was binary: either a species occurred in a given cell, or it didn't. Whether a species was thriving or in steep decline was not considered. This means that only the most drastic outcome of human influence, extirpation (from a 7500 km^{-2} area), would be registered as human influence affecting species richness. More informative indicators of biodiversity, such as evenness or Simpson's/Shannon's index, require population data. Unfortunately, this kind of data was not available. In order to more accurately assess the effect that humans have on their environment, much smaller study units are required.

The large size of the cells may also explain some other unexpected findings. For instance; the spatial patterns found for observed species richness and generalized species richness were roughly similar to one another. This was not the case when the same method was applied on a smaller scale in a preliminary analysis on trees in the eastern United States (see Appendix 3), nor in the study introducing the method of reflection[21]. In both cases, species were coupled to very small local plots rather than upscaled to cells. These plots were only a few square meters in area; a scale at which the plants may directly influence one another. Such local assemblages are much more sensitive to random disturbances, and sampling issues, and as a result observed species richness will vary more widely due to a higher level of noise. This is exactly the idea behind the utility of the method of reflections; it removes stochasticity through repeated averaging and yields an indicator that shows how species-rich an area should be relative to other areas in the network, irrespective of the number of species it actually contains. This indicator better represents the biodiversity of the area relative to the rest of the network, because it is less subject to random events that temporarily alter species richness. However, stochasticity plays only a minor role when study units are very large, as was the case in the current study. This could in part explain the similar patterns for observed and generalized species richness that were found. Nevertheless, the differences were enough to yield a unique pattern for the anomaly, that was distinct from either generalized or observed species richness, but that seemed to correlate with topographical features. Peninsulas and islands, for instance, were consistently highly anomalous, corresponding to previous studies that found such environments to be less diverse than the mainland[43,44]. In contrast, mountain ranges were consistently the least anomalous. Even though the distinction between observed and generalized species richness is less informative at this grain of analysis, the anomaly between them still reveals patterns of biodiversity that either of these variables by itself cannot.

The relationship between anomaly and assemblage completeness, both variables that ostensibly quantify how close a cell's species richness is to its potential, was not as straightforward as might have been expected given the results of the preliminary analysis shown in Appendix 3. Generally, the most complete cells tended to be the least anomalous. This was most obvious in the major mountain ranges: the Sierra Nevada, the Rockies and (to a lesser extent) the Appalachians. The reverse pattern was also true, and was most evident in the cells along the Gulf Coast as well as on the Great Plains. It appears that while observed and especially generalized species richness show the overall general

pattern of biodiversity in North America (i.e. increasing from the north-east to the south-west), anomaly and completeness were both more sensitive to topographical factors. This seems especially true of the anomaly, which follows topographical features more accurately than does completeness thus suggesting that it may be more appropriate for use at this scale (although this cannot be assessed with the data available). Despite both responding to topography, the two indicators were often in disagreement.

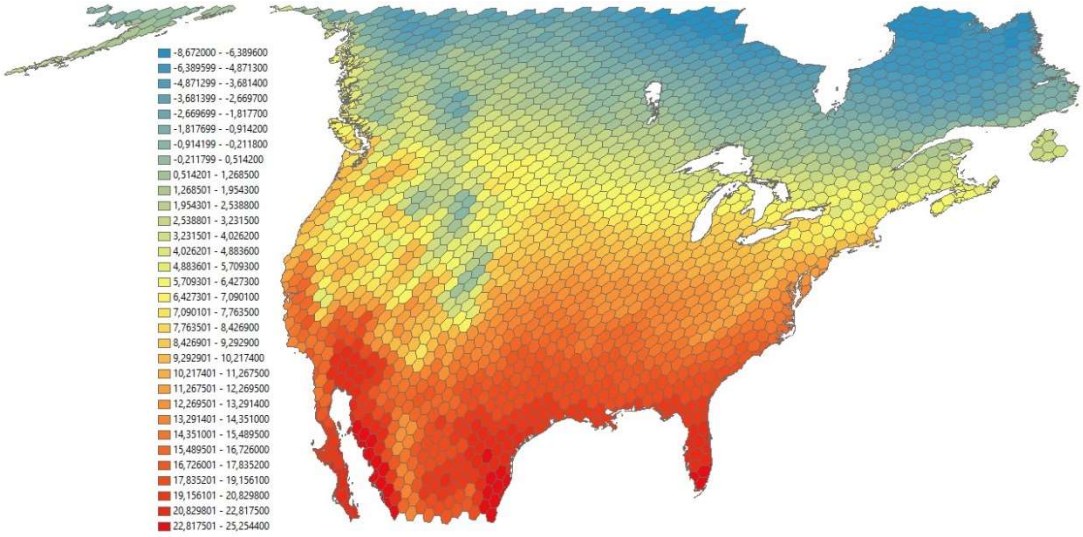
In the study of tree assemblages across the eastern United States discussed in Appendix 3, the assemblages' completeness was similarly regressed against their anomaly. Here, the relationship between the two indices was much more obvious; completeness decreased as the anomaly increased. Again, this could be a result of the fact that in this case sampling units were local plots, rather than very large grid cells. Moreover, since all the species sampled were trees, they belonged to the same trophic level of primary producers and were stationary. In contrast, our dataset included mammals of different trophic levels as well as of widely different home range sizes. Some of the species in the same cell might inhabit very different ecosystems and may never meet or interact with one another. As such, it might be erroneous to treat an assemblage of mammals in a hexagonal cell as if they constituted a single community. What has been treated as a community in this project might more accurately be considered gamma diversity for the different ecosystems *within* each cell. Calculating completeness depends on being able to ascertain an area's species' pool; the part of regional diversity that might be expected to inhabit the area given its environmental attributes. Beal's probability index infers an area's species pool based on co-occurrence patterns among the entire network considered, without having to model which species can inhabit it based on its ecological needs[18]. It is used to define the size and composition of the dark diversity in a cell, which is then used to calculate its completeness. However, the size of the cells means that species inhabiting dissimilar, separate habitats will nonetheless be marked as co-occurring, and on this bases inferences will be made regarding the species pool (and dark diversity, and thereby also community completeness). In a broad sense, this might be valid; the presence of desert-adapted species within a cell makes it more likely that other desert dwelling species would be present as well. But within regions with roughly the same climate, this may lead to spurious associations. For instance, while the presence of muskrats might be a realistic indicator of the presence of beavers (since they live in very similar environments), it might not be such a realistic predictor of the presence of red squirrels, despite often inhabiting the same cells. While the anomaly, through generalized species richness, is also based on species occurrence patterns in the network, it does not rely on explicitly defining which species occur together. Rather, generalized species richness is based on the 'sociability' of species occurring within a cell; which is to say whether a species is generally associated with high species richness or not. Because the anomaly does not need to explicitly define which species co-occur, it may be less prone to spurious associations than completeness, and may be more suitable to be applied on larger scales.

A next step would be to test and compare these methods on a smaller scale, utilizing local, unaggregated data. Also, instead of analyzing all mammals, only a subset occupying a similar trophic level/niche could be considered. This would be more appropriate in the context of applying the dark diversity concept for conservation purposes. However, due to their high mobility, it is difficult to "pin" mammal species to specific sites, and some kind of intermediate spatial scale, between local site data and upscaled cell data, may be required.

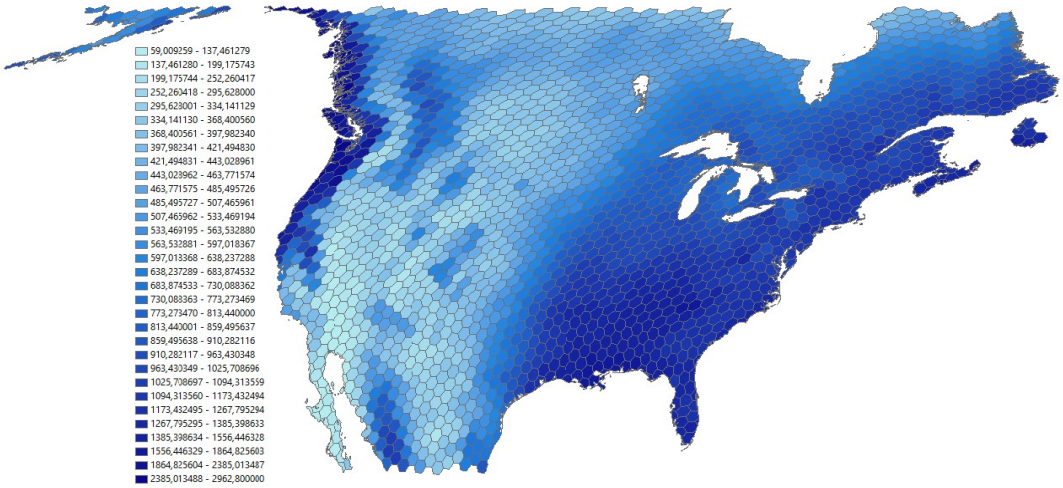
Conclusion

- At the scale considered in this project, a combination of elevation range, temperature and precipitation was best suited to explain the variance in different measures of biodiversity. The effect of human activity, while significant, was very limited.
- Anomaly and community completeness, the two biodiversity variables that were considered here to indicate how close a cell is to its potential (in terms of species richness), showed some agreement, particularly in cells that were on the extreme end of either of these variables. Variation in these variables corresponded mostly with topographical features. While these variables are not similar enough for the anomaly to substitute community completeness, they might be used as alternatives.

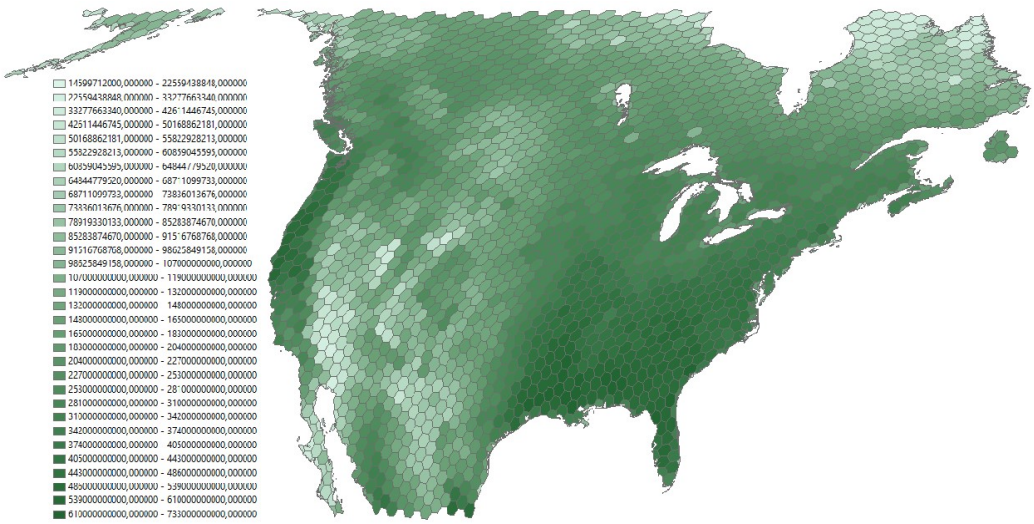
Appendix 1: map representation of explanatory variables



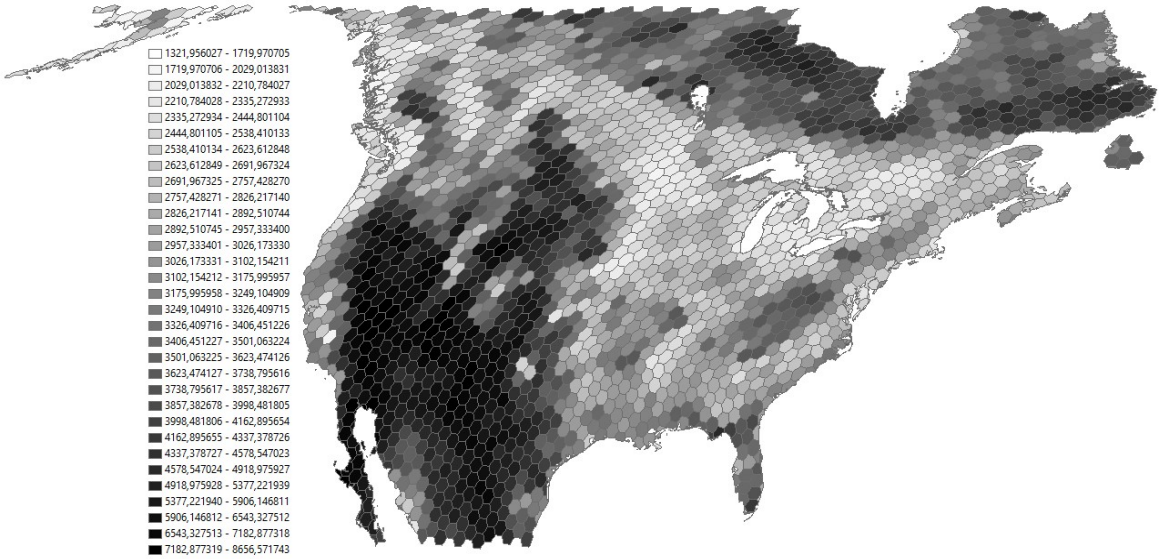
Appendix 1.1 Mean annual temperatures (°C) in North America



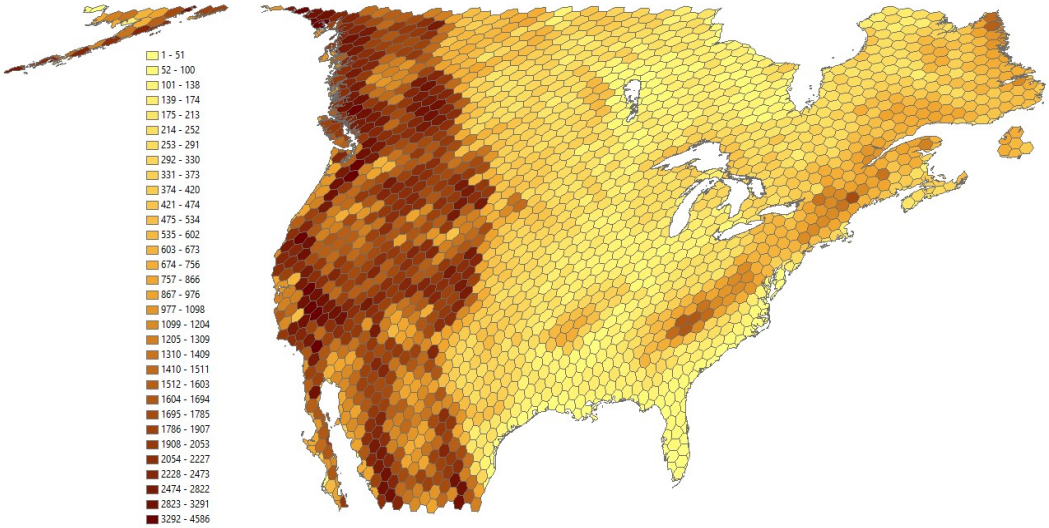
Appendix 1.2: mean annual precipitation (mm) in North America



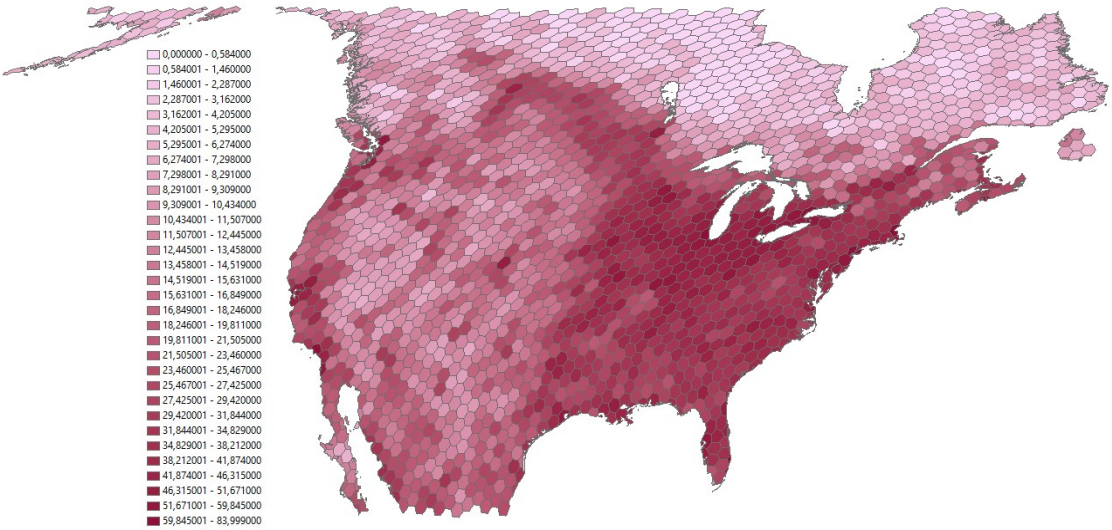
Appendix 1.3: net annual primary productivity (grams of carbon) in North America



Appendix 1.4: habitat homogeneity in North America



Appendix 1.5: elevation range (m) in North America



Appendix 1.6: Human influence (hii) in North America

Appendix 2: model formulae

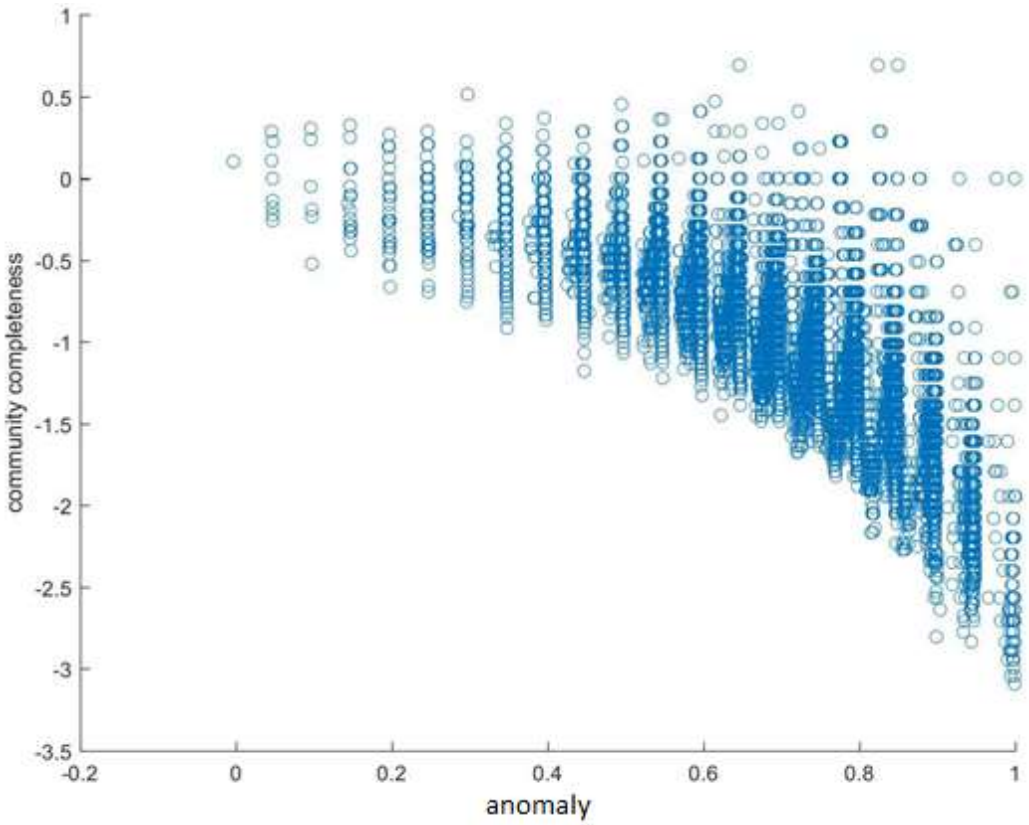
The following table shows the full model formulae of the models discussed in the result. For the multiple-variable models, the variable listed first is x1, the second x2, etc.

	Formula
Temp	$y = 29.8948 + 1.3419x + -0.073329x^2 + 0.001721x^3$
Precipitation	$y = 54.2754 + -0.054926x + 3.8877e\text{-}05x^2 + -8.4015e\text{-}09x^3$
NPP	$y = 33.4021 + 3.767e\text{-}12x$
Homogen	$y = 24.9478 + 0.0013722x + 3.0433e\text{-}07x^2$
Elevation	$y = 26.1414 + 0.010798x + 1.0876e\text{-}06x^2 + -7.2655e\text{-}10x^3$
Hii	$y = 21.0006 + 2.1689x - 0.070748x^2 + 0.00058816x^3$
Temp + prec + elev	y=29.226 +0.7294x1 -0.0085945x2 +0.0077186x3
Temp + npp + hmg + elev	y=15.6992 +0.56344x1 +5.2459e-12x2 +0.0020183x3 +0.0076944x4
Temp + homogen + elev	y=17.6514 +0.61044x1 +0.0017073x2 +0.007659x3
Temp + npp + elev	y=23.7215 +0.70229x1 -4.7808e-12x2 +0.0082053x3
Temp + elev **Ks0**	y=22.7966 +0.66921x1 +0.0083527x2
Temp	$y = 37.8785 + 0.0050519x + -0.00012692x^2 + 5.9645e\text{-}06x^3$
Precipitation	$y = 38.0337 + -0.00037717x + 2.9948e\text{-}07x^2 + -6.7334e\text{-}11x^3$
NPP	$y = 37.9078 + 1.4105e\text{-}14x$
Homogen	$y = 37.873 + 2.2905e\text{-}06x + 2.074e\text{-}09x^2$
Elevation	$y = 37.8861 + 4.0191e\text{-}05x - 1.8721e\text{-}09x^2 - 1.8354e\text{-}12x^3$
Hii	$y = 37.8437 + 0.0099194x - 0.00030445x^2 + 2.4966e\text{-}06x^3$
Temp + prec + elev	y=37.8904 +0.0051144x1 -4.0072e-05x2 +1.7329e-05x3
Temp + npp + hmg + elev	y=37.8503 +0.0049977x1 -5.9256e-14x2 +7.1476e-06x3 +1.5554e-05x4
Temp + hmg + elev	y=37.8283 +0.0044667x1 +1.0661e-05x2 +1.5954e-05x3
Temp + prec + elev	y=37.8787 +0.0054894x1 -9.4764e-14x2 +1.7364e-05x3
Temp + prec **Ks18**	y=37.9068 +0.005242x1 -4.4819e-05x2

Temp	y =-0.1153 -0.00094502x +0.00049338x^2
Precipitation	y =0.0785 -0.00053365x +4.8859e-07x^2 -1.145e-10x^3
NPP	y =-0.065649
Homogen	y =-0.067867 -1.0987e-05x +2.903e-09x^2
Elevation	y =-0.032757 -4.1155e-05x
Hii	y =-0.11382 +0.0041116x -6.8957e-05x^2 +4.5578e-07x^3
Temp + npp + hmg + elev	y=-0.018545 +0.0093227x1 +-2.7391e-13x2 +-3.2764e-06x3 -5.2365e-05x4
Temp + npp + elev	y=-0.031568 +0.0090973x1 -2.5763e-13x2 +-5.3195e-05x3
Temp + hmg + elev	y=-0.12048 +0.0068684x1 +1.2963e-05x2 -5.0519e-05x3
Temp + prec + elev	y=-0.06727 +0.007447x1 -1.8903e-05x2 - 4.6647e-05x3
Temp + elev **Anomaly**	y=-0.081411 +0.0073146x1 +-4.5252e-05x2
Temp	y =2.1458 +0.061753x -0.0036932x^2 +7.0598e-05x^3
Precipitation	y =3.0987 -0.0025112x +2.0883e-06x^2 -5.1173e-10x^3
NPP	y =2.1873 +4.8368e-13x
Homogen	y =2.1071 -1.0689e-05x +1.563e-08x^2
Elevation	y =1.9492 +0.00040347x +1.0051e-07x^2 -4.0197e-11x^3
Hii	y =1.6871 +0.099372x -0.0032645x^2 +2.7556e-05x^3
Temp + prec + elev	y=2.0035 +0.024487x1 -0.0002367x2 +0.00036321x3
Temp + npp + hmg + elev	y=1.4834 +0.015541x1 +7.1542e-13x2 +6.7903e-05x3 +0.00037514x4
Temp + npp + elev	y=1.7532 +0.020213x1 +3.7809e-13x2 +0.00039233x3
Temp + hmg + elev	y=1.7496 +0.021951x1 +2.5488e-05x2 +0.00037032x3
Temp + elev **Completeness**	y=1.8264 +0.022829x1 +0.00038067x2

Appendix 3

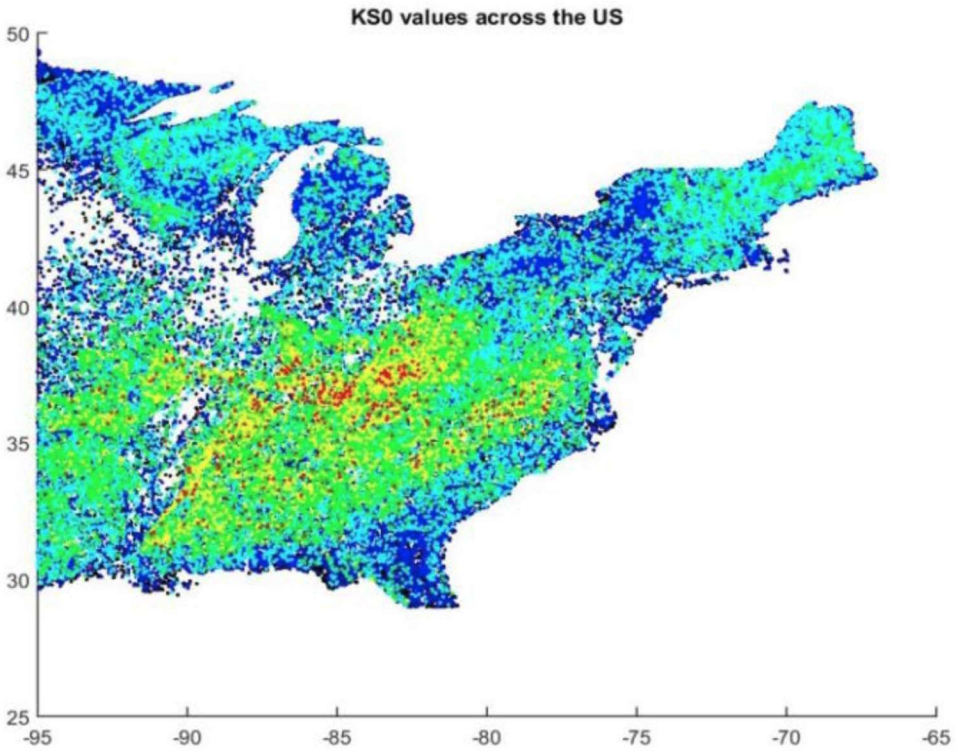
In the unpublished study mentioned in the text, a dataset of forest plots sampled by the US Forest Inventory and Analysis National Program[45,46] was used for a large-scale analysis of tree communities in the eastern United States. This dataset listed all the tree species present in the plot, as well as the abundance of seedlings and adults of said species. A total of 68290 forest plots of 168.22 m^2 were analyzed. The study area extended from -95°W to the east coast, and from the Canadian border to the south coast, although the southernmost part of Florida was excluded. The same methods as described in the Methods section were used to calculate both the anomaly, and community completeness. The results were as follows:

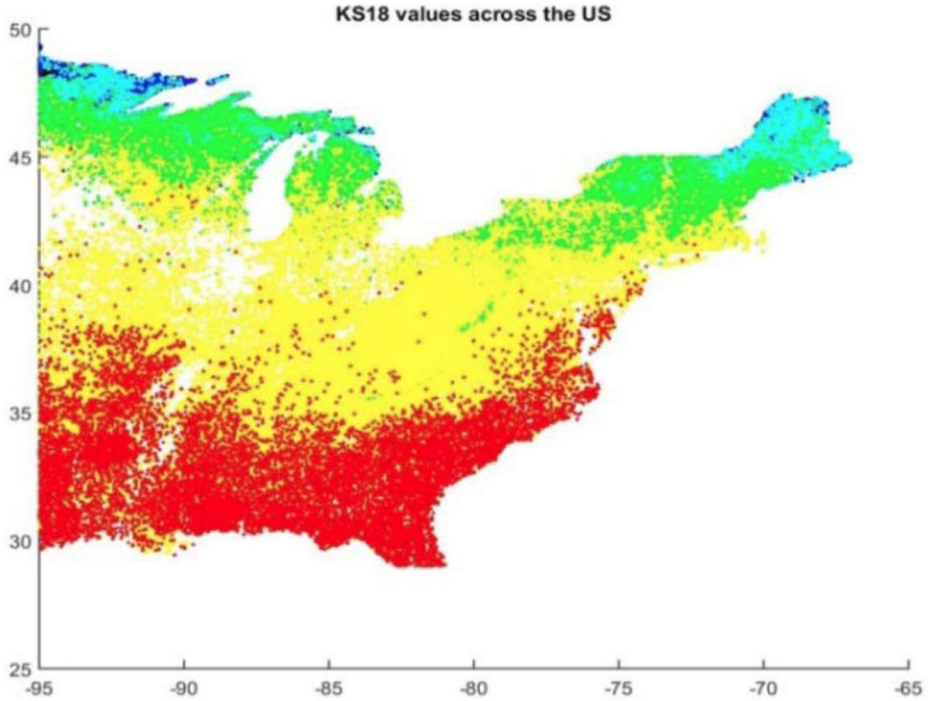


Appendix 3.1: relationship between anomaly and completeness of tree communities in the eastern U.S.

The relationship between completeness and anomaly is much more obvious here; sites that are highly anomalous (i.e. normalized $k_{s,18}$ >> normalized $k_{s,0}$) score lower for community completeness. As expected, a high anomaly means that there are fewer species in the assembly than would be expected based on the distribution of species across the network that do occur in the assembly. A low community completeness indicates a high ratio of missing species (dark diversity) versus observed species richness. So, while these indicators are different mathematically, they both indicate how many species are missing relative to what could be expected. The very different pattern found in the current study may be a result of the (much) larger sampling units, the fact that the species considered are motile and belong to different trophic levels, or all of the above.

In addition, the differences between the spatial patterns of observed and generalized species richness were much more pronounced than in the current study. A very clear gradual increase in generalized species richness was observed, but this was not the case for observed species richness (see maps).





Appendix 3.2: spatial patterns of observed species richness (above) and generalized species richness (below)

Bibliography

1. Kier, G. *et al.* A global assessment of endemism and species richness across island and mainland regions. *Proc. Natl. Acad. Sci.* **106,** 9322–9327 (2009).
2. Pärtel, M., Bennett, J. A. & Zobel, M. Macroecology of biodiversity: disentangling local and regional effects. *New Phytol.* **211,** 404–410 (2016).
3. Field, R. *et al.* Spatial species-richness gradients across scales: a meta-analysis. *J. Biogeogr.* **36,** 132–147 (2009).
4. Mittelbach, G. G. & Schemske, D. W. Ecological and evolutionary perspectives on community assembly. *Trends Ecol. Evol.* **30,** 241–247 (2015).
5. Zobel, M. The species pool concept as a framework for studying patterns of plant diversity. *J. Veg. Sci.* **27,** 8–18 (2016).
6. Pärtel, M. Community ecology of absent species: hidden and dark diversity. *J. Veg. Sci.* **25,** 1154–1159 (2014).
7. Cornell, H. V. & Harrison, S. P. What Are Species Pools and When Are They Important? *Annu. Rev. Ecol. Evol. Syst.* **45,** 45–67 (2014).
8. Pärtel, M., Szava-Kovats, R. & Zobel, M. Dark diversity: shedding light on absent species. *Trends Ecol. Evol.* **26,** 124–128 (2011).
9. Riibak, K. *et al.* Dark diversity in dry calcareous grasslands is determined by dispersal ability and stress-tolerance. *Ecography* **38,** 713–721 (2015).
10. Götzenberger, L. *et al.* Ecological assembly rules in plant communities-approaches, patterns and prospects. *Biol. Rev.* **87,** 111–127 (2012).
11. Haddad, N. M. *et al.* Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci. Adv.* **1,** e1500052 (2015).
12. Ronk, A., de Bello, F., Fibich, P. & Pärtel, M. Large-scale dark diversity estimates: new perspectives with combined methods. *Ecol. Evol.* **6,** 6266–6281 (2016).

13. Pärtel, M., Szava-Kovats, R. & Zobel, M. Community Completeness: Linking Local and Dark Diversity within the Species Pool Concept. *Folia Geobot.* **48,** 307–317 (2013).

14. Bennett, J. A. *et al.* Species pools, community completeness and invasion: disentangling diversity effects on the establishment of native and alien species. *Ecol. Lett.* **19,** 1496–1505 (2016).

15. Lewis, R. J. *et al.* Applying the dark diversity concept to nature conservation. *Conserv. Biol.* n/a-n/a (2016). doi:10.1111/cobi.12723

16. Collen, B. *et al.* Monitoring Change in Vertebrate Abundance: the Living Planet Index. *Conserv. Biol.* **23,** 317–327 (2009).

17. Ceballos, G. & Ehrlich, P. R. Mammal Population Losses and the Extinction Crisis. *Science* **296,** 904–907 (2002).

18. Lewis, R. J., Szava-Kovats, R. & Pärtel, M. Estimating dark diversity and species pools: an empirical assessment of two methods. *Methods Ecol. Evol.* **7,** 104–113 (2016).

19. Münzbergová, Z. & Herben, T. Identification of suitable unoccupied habitats in metapopulation studies using co-occurrence of species. *Oikos* **105,** 408–414 (2004).

20. De Cáceres, M. & Legendre, P. Beals smoothing revisited. *Oecologia* **156,** 657–669 (2008).

21. Baudena, M. *et al.* Revealing patterns of local species richness along environmental gradients with a novel network tool. *Sci. Rep.* **5,** 11561 (2015).

22. de Bello, F. *et al.* Measuring size and composition of species pools: a comparison of dark diversity estimates. *Ecol. Evol.* **6,** 4088–4101 (2016).

23. Spatial Data Download. Available at: http://www.iucnredlist.org/technical-documents/spatial-data. (Accessed: 19th June 2017)

24. Wildlife Conservation Society-WCS & Center For International Earth Science Information Network-CIESIN-Columbia University. Last of the Wild Project, Version 2, 2005 (LWP-2): Last of the Wild Dataset (Geographic). (2005). doi:10.7927/H4348H83

25. Robinson, N., Regetz, J. & Guralnick, R. P. EarthEnv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. *ISPRS J. Photogramm. Remote Sens.* **87,** 57–67 (2014).

26. Tuanmu, M.-N. & Jetz, W. A global 1-km consensus land-cover product for biodiversity and ecosystem modelling: Consensus land cover. *Glob. Ecol. Biogeogr.* **23,** 1031–1045 (2014).

27. WorldClim - Global Climate Data | Free climate data for ecological modeling and GIS. Available at: http://www.worldclim.org/. (Accessed: 20th June 2017)

28. Clarke, A. & Gaston, K. J. Climate, energy and diversity. *Proc. R. Soc. B Biol. Sci.* **273,** 2257–2266 (2006).

29. Allen, A. P., Brown, J. H. & Gillooly, J. F. Global Biodiversity, Biochemical Kinetics, and the Energetic-Equivalence Rule. *Science* **297,** 1545–1548 (2002).

30. Hawkins, B. A. *et al.* Energy, Water, and Broad-Scale Geographic Patterns of Species Richness. *Ecology* **84,** 3105–3117 (2003).

31. Gotelli, N. J. & Colwell, R. K. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **4,** 379–391 (2001).

32. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19,** 716–723 (1974).

33. Burnham, K. P., Anderson, D. R. & Burnham, K. P. *Model selection and multimodel inference: a practical information-theoretic approach*. (Springer, 2002).

34. D'Onofrio, D., von Hardenberg, J. & Baudena, M. The distribution of grasslands, savannas and forests in Africa: a new look at the relationships between vegetation, fire and climate at continental scale. in *EGU General Assembly Conference Abstracts* **19,** 17154 (2017).

35. Buckley, L. B. *et al.* Phylogeny, niche conservatism and the latitudinal diversity gradient in mammals. *Proc. R. Soc. Lond. B Biol. Sci.* **277,** 2131–2138 (2010).

36. Davies, T. J., Buckley, L. B., Grenyer, R. & Gittleman, J. L. The influence of past and present climate on the biogeography of modern mammal diversity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366,** 2526–2535 (2011).

37. Blois, J. L., McGuire, J. L. & Hadly, E. A. Small mammal diversity loss in response to late-Pleistocene climatic change. *Nature* **465,** 771–774 (2010).

38. Qian, H., Badgley, C. & Fox, D. L. The latitudinal gradient of beta diversity in relation to climate and topography for mammals in North America. *Glob. Ecol. Biogeogr.* **18,** 111–122 (2009).

39. Badgley, C. & Fox, D. L. Ecological biogeography of North American mammals: species density and ecological structure in relation to environmental gradients. *J. Biogeogr.* **27,** 1437–1467 (2000).

40. Brown, J. H. & Nicoletto, P. F. Spatial Scaling of Species Composition: Body Masses of North American Land Mammals. *Am. Nat.* **138,** 1478–1512 (1991).

41. McCain, C. M. Elevational Gradients in Diversity of Small Mammals. *Ecology* **86,** 366–372 (2005).

42. McCain, C. M. & Grytnes, J.-A. Elevational Gradients in Species Richness. in *Encyclopedia of Life Sciences* (ed. John Wiley & Sons, Ltd) (John Wiley & Sons, Ltd, 2010). doi:10.1002/9780470015902.a0022548

43. Olivier, P. I., Rolo, V. & Aarde, R. J. van. Pattern or process? Evaluating the peninsula effect as a determinant of species richness in coastal dune forests. *PLOS ONE* **12,** e0173694 (2017).

44. Whittaker, R. J. & Fern©Łndez-Palacios, J. M. *Island Biogeography: Ecology, Evolution, and Conservation*. (OUP Oxford, 2007).

45. Johnson, D. J. *et al.* Conspecific negative density-dependent mortality and the structure of temperate forests. *Ecology* **95,** 2493–2503 (2014).

46. Forest Inventory and Analysis National Program. Available at: https://www.fia.fs.fed.us/. (Accessed: 21st June 2017)