

Quantitative Analyse des Zusammenhangs zwischen der globalen Sprachendiversität und geographischen Faktoren

Masterarbeit GEO 511

Abgabedatum: 30. September 2013

Geographisches Institut der Universität Zürich
Abteilung Geographische Informationssysteme

Martina Köhli
06-111-488

Fakultätsvertreter & Betreuer:
Prof. Dr. Robert Weibel

Externer Betreuer:
Prof. Dr. Balthasar Bickel
Seminar für Allgemeine Sprachwissenschaft

Kontakt:

Martina Köhli
martinakoehli@hotmail.com

Prof. Dr. Robert Weibel
Universität Zürich
Geographisches Institut
Abteilung Geographische Informationssysteme
Winterthurerstrasse 190
CH-8057 Zürich

Prof. Dr. Balthasar Bickel
Universität Zürich
Seminar für Allgemeine Sprachwissenschaft
Plattenstrasse 54
CH-8032 Zürich

ABSTRACT

English

There are over 7100 known languages that are unevenly distributed all over the world. The linguistic diversity is strikingly higher near the equator than it is towards the poles. Likewise, a lot of climatic factors are displaying a latitudinal pattern. Hence, it seems reasonable to consider a connection between climate and language diversity. This work analyses in detail to what extent geographical factors are able to explain the global distribution of languages. For this purpose an extensive language database and high-resolution data concerning the geographical factors were used in combination with the methods and instruments of geographic information science. Climatic factors such as precipitation as well as situational factors such as height above sea level were included in the analysis. The correlations between these factors and language diversity, language family diversity as well as genealogical diversity were examined.

According to the analyses made in this work one third of the variance of language diversity could be explained by geographical factors. When the subsistence strategy is taken into account, this amount gets even higher. Similarly, nearly one third of the variance of language family diversity can be defined by geographical factors. This amount was noticeably lower before large language families were removed from the model. They were excluded, since their expansion was often shaped through cultural and historical rather than geographical factors. Lastly, one fourth is the amount of variance of genealogical diversity that is explicable by the chosen climatic and situational factors. It also needs to be noted that all of the results were considerably influenced by the language family groupings; the levels of diversity were *per se* affected by the random effects specific to particular language families.

Deutsch

Die weltweit über 7'100 Sprachen sind sehr unregelmässig auf der Welt verteilt. In der Nähe des Äquators ist die Sprachendiversität deutlich höher als zu den Polen hin. Da auch viele Klimafaktoren mit der geographischen Breite korrelieren, liegt die Vermutung nahe, dass hier ein Zusammenhang bestehen könnte. In dieser Arbeit wurde untersucht inwiefern geographische Faktoren die weltweite Sprachenverteilung erklären können. Dazu wurden eine grosse Sprachdatenbank und hochaufgelöste Daten zu geographischen Faktoren in Kombination mit den Methoden und Instrumenten der Geographischen Informationswissenschaft verwendet. Es wurden sowohl Klimafaktoren, wie zum Beispiel der Niederschlag, als auch Lagefaktoren, wie beispielsweise die Höhe über Meeresspiegel, einbezogen. Der Zusammenhang zu diesen Faktoren wurde für die Sprachendiversität, die Sprachfamiliendiversität, wie auch die genealogische Diversität untersucht.

Die Varianz der Sprachendiversität konnte zu gut einem Drittel mit geographischen Faktoren erklärt werden. Wird noch die Subsistenzstrategie der Sprecher der verschiedenen Sprachen berücksichtigt, steigt dieser Anteil deutlich an. Ebenfalls kann circa ein Drittel der Sprachfamiliendiversität durch geographische Faktoren erklärt werden. Dies war jedoch erst der Fall, nachdem grosse Sprachfamilien, deren Expansion stark durch kulturelle und historische Faktoren geprägt ist, nicht berücksichtigt wurden. Gut ein Viertel beträgt die durch die Klima- und Lagefaktoren erklärte Varianz der genealogischen Diversität. Bei allen berechneten Diversitäten spielt die Zugehörigkeit zu einer Sprachfamilie eine entscheidende Rolle, welche offenbar *per se* darüber entscheidet, auf welchem Niveau sich diese bewegen.

DANKSAGUNG

Ich möchte mich bei meinen Betreuern Prof. Dr. Robert Weibel und Prof. Dr. Balthasar Bickel für die hilfreichen Inputs, die interessanten Gespräche, die zeitintensiven Treffen und dafür, dass ich von ihrem fundierten Wissen profitieren durfte, bedanken.

Franziska Appelt möchte ich danken für ihre Unterstützung und ihren Korrekturstift, der zielstrebig Jagd auf meine grammatikalischen Missgeschicke gemacht hat.

Ein grosser Dank gilt meinen Eltern, welche mich immer unterstützt haben.

Inhaltsverzeichnis

Abstract	I
Danksagung	II
Abbildungsverzeichnis	IX
Tabellenverzeichnis	XIII
I. EINLEITUNG	1
1 Motivation	1
2 Problemstellung	1
2.1 Ziele.....	1
2.2 Forschungsfragen.....	2
II. HINTERGRUND	3
3 Grundlagen Linguistik	3
3.1 Einheit und Diversität der Sprachen.....	3
3.2 Sprachen verändern sich.....	3
3.2.1 Gründe für die Sprachveränderung.....	3
3.3 Verwandtschaft von Sprachen.....	4
3.4 Problematik der genealogischen Verwandtschaft.....	4
4 Forschungsstand	5
4.1 Geographische Faktoren und Sprachendiversität.....	5
4.1.1 Verschiedene Erklärungsansätze.....	5
4.1.2 Methoden und Erkenntnisse der verschiedenen Erklärungsansätze.....	5
4.2 Geographische Faktoren und Sprachfamiliendiversität.....	7
4.3 Geographische Faktoren und genealogische Diversität.....	8
4.4 Kritische Beurteilung der bisherigen Literatur.....	8
4.4.1 Indirektes Mass für Sprachendiversität.....	8
4.4.2 Unterschiedlich aggregierte Sprachendiversität und geographische Faktoren.....	9
4.4.3 Räumlich begrenzte Analysen.....	9
5 Forschungslücke	10
III. FORSCHUNGSANSATZ	11
6 Linguistische Datengrundlagen	11
6.1 Hammarström - Daten.....	11
6.1.1 Herkunft / Aufbau.....	11
6.1.2 Datenaufbereitung.....	11
6.2 AUTOTYP - Datenbank.....	13
6.2.1 Herkunft / Aufbau.....	13
6.2.2 Datenaufbereitung.....	13

7	Grundlagen geographische Faktoren	15
7.1	Klimatische Faktoren	15
7.1.1	Mittlere jährliche Temperatur	15
7.1.2	Jährlicher Temperaturschwankungsbereich	16
7.1.3	Jährlicher Niederschlag	16
7.1.4	Saisonalität des Niederschlags	17
7.2	Nettoprimärproduktion	17
7.3	Länge der Wachstumsphase	18
7.4	Küstendistanz	19
7.5	Höhenmodell	20
7.6	Höhendifferenz	20
7.7	Korrelationen der geographischen Faktoren untereinander	21
IV.	METHODIK.....	22
8	Vorgehen.....	22
9	Sprachendiversität.....	23
9.1	Lokale Sprachendichte.....	23
9.1.1	Berechnung	23
9.1.2	Häufigkeitsverteilung lokale Sprachendichte	24
9.2	Geographisch Nächster Nachbar (GNN)	25
9.2.1	Berechnung	25
9.2.2	Häufigkeitsverteilung GNN-Distanz	26
10	Sprachfamiliendiversität.....	27
10.1	Lokale Sprachfamiliendichte.....	27
10.1.1	Berechnung.....	27
10.1.2	Häufigkeitsverteilung lokale Sprachfamiliendichte	27
10.2	Geographisch Nächster Nachbar, nicht-verwandt (GNNnv).....	29
10.2.1	Berechnung.....	29
10.2.2	Häufigkeitsverteilung der GNNnv-Distanz	29
11	Genealogische Diversität	31
11.1	Mittlere Verwandtschaftsdistanz	31
11.1.1	Berechnung.....	31
11.1.2	Häufigkeitsverteilung der mittleren Verwandtschaftsdistanz	32
11.2	Phylogenetisch Nächster Nachbar (PNN)	32
11.2.1	Berechnung.....	32
11.2.2	Häufigkeitsverteilung PNN-Distanz.....	33
12	Berechnung geographische Faktoren.....	35
13	Software.....	37

V. RESULTATE	38
14 Überblick	38
14.1 Aufbau	38
14.2 Statistische Grundlagen.....	38
15 Sprachdiversität	40
15.1 Globale Verteilung.....	40
15.1.1 Lokale Sprachendichte.....	40
15.1.2 Geographisch Nächster Nachbar	42
15.2 Mittlere jährliche Temperatur.....	43
15.2.1 Lokale Sprachendichte.....	43
15.2.2 Geographisch Nächster Nachbar	43
15.3 Jährlicher Temperaturschwankungsbereich	44
15.3.1 Lokale Sprachendichte.....	44
15.3.2 Geographisch nächster Nachbar	44
15.4 Jährlicher Niederschlag	45
15.4.1 Lokale Sprachendichte.....	45
15.4.2 Geographisch Nächster Nachbar	46
15.5 Saisonalität des Niederschlags	46
15.5.1 Lokale Sprachendichte.....	46
15.5.2 Geographisch Nächster Nachbar	47
15.6 Nettoprimärproduktion.....	47
15.6.1 Lokale Sprachendichte.....	47
15.6.2 Geographisch Nächster Nachbar	48
15.7 Länge der Wachstumsphase	48
15.7.1 Lokale Sprachendichte.....	48
15.7.2 Geographisch Nächster Nachbar	49
15.8 Küstendistanz	50
15.8.1 Lokale Sprachendichte.....	50
15.8.2 Geographisch Nächster Nachbar	50
15.9 Höhenmodell.....	51
15.9.1 Lokale Sprachendichte.....	51
15.9.2 Geographisch Nächster Nachbar	52
15.10 Höhendifferenz.....	52
15.10.1 Lokale Sprachendichte.....	52
15.10.2 Geographisch nächster Nachbar	53
15.11 Zusammenfassung einzelne Prädiktoren	54
15.12 Multiple Regressionsanalyse	55

15.12.1	Lokale Sprachendichte	55
15.12.2	Geographisch Nächster Nachbar	57
15.13	Mehrebenenanalyse	59
15.13.1	Lokale Sprachendichte	59
15.13.2	Geographisch Nächster Nachbar	60
15.14	Modellvergleich	61
15.14.1	Lokale Sprachendichte	61
15.14.2	Geographisch Nächster Nachbar	62
15.15	Vergleich der Sprachendiversitäts-Masse	63
15.16	Interpretation	65
15.16.1	Masse für Sprachendiversität	65
15.16.2	Einordnung der Resultate	65
16	Sprachfamiliendiversität.....	69
16.1	Globale Verteilung	69
16.1.1	Lokale Sprachfamiliendichte	69
16.1.2	Geographisch Nächster nicht-verwandter Nachbar	70
16.2	Mittlere jährliche Temperatur	72
16.2.1	Lokale Sprachfamiliendichte	72
16.2.2	Geographisch Nächster nicht-verwandter Nachbar	72
16.3	Jährlicher Temperaturschwankungsbereich.....	73
16.3.1	Lokale Sprachfamiliendichte	73
16.3.2	Geographisch Nächster nicht-verwandter Nachbar	73
16.4	Jährlicher Niederschlag.....	74
16.4.1	Lokale Sprachfamiliendichte	74
16.4.2	Geographisch Nächster nicht-verwandter Nachbar	74
16.5	Saisonalität des Niederschlags	75
16.5.1	Lokale Sprachfamiliendichte	75
16.5.2	Geographisch Nächster nicht-verwandter Nachbar	75
16.6	Nettoprimärproduktion	76
16.6.1	Lokale Sprachfamiliendichte	76
16.6.2	Geographisch Nächster nicht-verwandter Nachbar	76
16.7	Länge der Wachstumsphase.....	77
16.7.1	Lokale Sprachfamiliendichte	77
16.7.2	Geographisch Nächster nicht-verwandter Nachbar	77
16.8	Küstendistanz	78
16.8.1	Lokale Sprachfamiliendichte	78
16.8.2	Geographisch Nächster nicht-verwandter Nachbar	78

16.9	Höhenmodell	79
16.9.1	Lokale Sprachfamiliendichte	79
16.9.2	Geographisch Nächster nicht-verwandter Nachbar	79
16.10	Höhendifferenz	80
16.10.1	Lokale Sprachfamiliendichte	80
16.10.2	Geographisch Nächster nicht-verwandter Nachbar	80
16.11	Zusammenfassung einzelne Prädiktoren	82
16.12	Multiple Regressionsanalyse	83
16.12.1	Lokale Sprachfamiliendichte	83
16.12.2	Geographisch Nächster nicht-verwandter Nachbar	84
16.13	Mehrebenenanalyse	85
16.13.1	Lokale Sprachfamiliendichte	86
16.13.2	Geographisch Nächster nicht-verwandter Nachbar	87
16.14	Modellvergleich	88
16.14.1	Lokale Sprachfamiliendichte	88
16.14.2	Geographisch Nächster nicht-verwandter Nachbar	88
16.15	Vergleich der Sprachfamiliendiversitäts-Masse	90
16.16	Interpretation	92
16.16.1	Masse für Sprachfamiliendiversität	92
16.16.2	Einordnung der Resultate	93
17	Genealogische Diversität	95
17.1	Globale Verteilung	95
17.2	Mittlere jährliche Temperatur	96
17.3	Jährlicher Temperaturschwankungsbereich	97
17.4	Jährlicher Niederschlag	97
17.5	Saisonalität des Niederschlags	98
17.6	Nettoprimärproduktion	98
17.7	Länge der Wachstumsphase	98
17.8	Küstendistanz	99
17.9	Höhenmodell	99
17.10	Höhendifferenz	100
17.11	Zusammenfassung einzelne Prädiktoren	101
17.12	Multiple Regressionsanalyse	102
17.12.1	Mittlere Verwandtschaftsdistanz	102
17.12.2	PNN-Distanz – Hammerström-Daten	103
17.12.3	PNN-Distanz – AUTOTYP-Daten	104
17.13	Mehrebenenanalyse	106

17.13.1	Mittlere Verwandtschaftsdistanz	106
17.13.2	PNN-Distanz – Hammarström-Daten	107
17.13.3	PNN-Distanz – AUTOTYP-Daten	108
17.14	Modellvergleich	109
17.14.1	Mittlere Verwandtschaftsdistanz	109
17.14.2	PNN-Distanz – Hammarström-Daten	109
17.14.3	PNN-Distanz – AUTOTYP-Daten	109
17.15	Vergleich der genealogischen Diversitäts-Masse	110
17.16	Interpretation	112
17.16.1	Masse für genealogische Diversität	112
17.16.2	Einordnung der Resultate	113
18	Vergleich der Resultate untereinander	114
18.1	Vergleich der Sprachenmasse	114
18.2	Fazit zu den Sprachenmassen	115
VI.	FAZIT	117
19	Forschungsfragen und Antworten	117
20	Grenzen der gewählten Methodik	118
21	Ausblick	118
22	Literaturverzeichnis	119
Anhang	122
Anhang A:	Forschungsansatz – Zusätze	122
	Berechnung der lokalen Sprachendichte 100 km	122
	Berechnung Distanz zum Geographisch Nächsten Nachbar (GNN)	122
	Berechnung der lokalen Sprachfamiliendichte 400 km	123
	Berechnung mittlere Verwandtschaftsdistanz und PNN-Distanz	124
Anhang B:	Resultate – Zusätze	125
	Sprachendiversität – Mehrebenenanalyse, restliche Subsistenztypen.....	125
	Sprachfamiliendiversität – Mehrebenenanalyse, restliche Subsistenztypen.....	126
	Genealogische Diversität – Mehrebenenanalyse, restliche Subsistenztypen.....	127
Persönliche Erklärung	129

ABBILDUNGSVERZEICHNIS

Abb. 1: Verteilung der Sprachen eingeteilt nach Subsistenzstrategien.....	12
Abb. 2: Verteilung der verwendeten Hammarström-Sprachdaten.	13
Abb. 3: Verteilung der verwendeten AUTOTYP-Sprachdaten.....	14
Abb. 4: Häufigkeitsverteilung der mittleren jährlichen Temperatur jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.	16
Abb. 5: Häufigkeitsverteilung des jährlichen Temperaturschwankungs-bereichs jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.....	16
Abb. 6: Häufigkeitsverteilung des jährlichen Niederschlags jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.	17
Abb. 7: Häufigkeitsverteilung der Saisonalität des Niederschlags jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.....	17
Abb. 8: Häufigkeitsverteilung der mittleren Nettoprimärproduktion pro Tag jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.....	18
Abb. 9: Häufigkeitsverteilung der Länge der Wachstumsphase auf die Sprachen. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.....	19
Abb. 10: Häufigkeitsverteilung der Distanz zur Küste jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.....	19
Abb. 11: Häufigkeitsverteilung der Höhe über Meer jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.....	20
Abb. 12: Häufigkeitsverteilung der maximalen Höhendifferenz in einem Umkreis von 50 km um eine Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.....	20
Abb. 13: Die Sprachen als Punkte mit einem Umkreis mit Radius 100 km. Die Sprache mit dem grün eingefärbten Umkreis zählt drei Nachbarsprachen.	23
Abb. 14: Häufigkeitsverteilung der Anzahl Nachbarsprachen im Umkreis von 100 km. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.	24
Abb. 15: Häufigkeitsverteilung der Anzahl Nachbarsprachen im Umkreis von 400 km. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.	24
Abb. 16: Häufigkeitsverteilung der Anzahl Nachbarsprachen im Umkreis von 800 km. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.	25
Abb. 17: Vier Sprachen mit der GNN-Distanz [km]. Die Pfeile geben die Richtung zum nächsten Nachbarn an.	25
Abb. 18: Häufigkeitsverteilung der Distanz zum geographisch nächsten Nachbarn (GNN) in logarithmierter Darstellung. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.....	26
Abb. 19: Die Sprachen als Punkte eingefärbt nach Sprachfamilie mit einem Umkreis mit Radius 400 km. Die Sprache mit dem grün eingefärbten Umkreis zählt eine Nachbarsprachfamilie.....	27
Abb. 20: Häufigkeitsverteilung der Anzahl Nachbarsprachfamilien im Umkreis von 100 km.	28
Abb. 21: Häufigkeitsverteilung der Anzahl Nachbarsprachfamilien im Umkreis von 400 km. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.....	28
Abb. 22: Häufigkeitsverteilung der Anzahl Nachbarsprachfamilien im Umkreis von 800 km. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.....	28
Abb. 23: Vier Sprachen von zwei verschiedenen Sprachfamilien (grün, rot) und ihre GNN-Distanz.....	29
Abb. 24: Häufigkeitsverteilung der Distanz zum geographisch nächsten, nicht-verwandten Nachbar (PNN) in logarithmierter Darstellung. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.	30
Abb. 25: Drei verwandte Sprachen (grün) mit Angaben zur mittleren Verwandtschaftsdistanz [km].	31
Abb. 26: Häufigkeitsverteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.....	32
Abb. 27: Drei verwandte Sprachen (grün). Hervorgehobene Sprache mit Angabe zur PNN Distanz [km].	32

Abb. 28: Häufigkeitsverteilung der Distanz zum phylogenetisch nächsten Nachbar (PNN) in logarithmierter Darstellung. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.	33
Abb. 29: Häufigkeitsverteilung der Distanz zum phylogenetisch nächsten Nachbar (PNN) der AUTOTYP-Datenbank in logarithmierter Darstellung. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.....	33
Abb. 30: Verteilung der mittleren nächste-Nachbar Distanz der einzelnen Sprachfamilien in logarithmierter Darstellung. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.	34
Abb. 31: Die x-Achse zeigt den jährlichen Niederschlag am Ort der Sprachkoordinate und die y-Achse den durchschnittlichen Niederschlag 50 km rund um jede Sprache.	35
Abb. 32: Kerndichteschätzung der Sprachen. Rot bedeutet eine hohe Dichte, blau eine niedrige Dichte.	40
Abb. 33: Verteilung der lokalen Sprachendichte über die Breitengrade. Farblich hervorgehoben sind die Sprachen der grössten Sprachfamilien.	41
Abb. 34: Verteilung der lokalen Sprachendichte über die Breitengrade. In blau die Werte der Südhalbkugel, in schwarze jene der Nordhalbkugel.	41
Abb. 35: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Breitengrade. Farblich hervorgehoben sind die Sprachen der grössten Sprachfamilien.	42
Abb. 36: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Breitengrade. In blau die Werte der Südhalbkugel, in schwarze jene der Nordhalbkugel.	42
Abb. 37: Verteilung der lokalen Sprachendichte über die mittlere jährliche Temperatur.	43
Abb. 38: Verteilung der GNN-Distanz in logarithmierter Darstellung über die mittlere jährliche Temperatur.	43
Abb. 39: Verteilung der lokalen Sprachendichte über den jährlichen Temperaturschwankungsbereich.	44
Abb. 40: Verteilung der GNN-Distanz in logarithmierter Darstellung über den jährlichen Temperaturschwankungsbereich in logarithmierter Darstellung.	45
Abb. 41: Verteilung der lokalen Sprachendichte über den jährlichen Niederschlag.	45
Abb. 42: Verteilung der GNN-Distanz in logarithmierter Darstellung über den jährlichen Niederschlag.	46
Abb. 43: Verteilung der lokalen Sprachendichte über die Saisonalität des Niederschlags.	46
Abb. 44: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Saisonalität des Niederschlags.	47
Abb. 45: Verteilung der lokalen Sprachendichte über die mittlere Nettoprimärproduktion.	48
Abb. 46: Verteilung der GNN-Distanz in logarithmierter Darstellung über die mittlere Nettoprimärproduktion.	48
Abb. 47: Verteilung der lokalen Sprachendichte über die Länge der Wachstumsphase.	49
Abb. 48: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Länge der Wachstumsphase.	49
Abb. 49: Verteilung der lokalen Sprachendichte über die Küstendistanz.	50
Abb. 50: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Küstendistanz.	51
Abb. 51: Verteilung der lokalen Sprachendichte über die Höhe über Meer.	51
Abb. 52: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Höhe über Meer.	52
Abb. 53: Verteilung der lokalen Sprachendichte über die maximale Höhendifferenz 50 km um die jeweilige Sprache.	52
Abb. 54: Verteilung der GNN-Distanz in logarithmierter Darstellung über die maximale Höhendifferenz 50 km um die jeweilige Sprache.	53
Abb. 55: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Anzahl an Nachbarsprachen in logarithmierter Darstellung.	63
Abb. 56: Umkreise mit Radius 400 km um jede Sprache, eingefärbt nach Anzahl Nachbarsprachfamilien. Rot bedeutet eine hohe Dichte, blau eine niedrige Dichte.	69
Abb. 57: Verteilung der lokalen Sprachfamiliendichten über die Breitengrade. Farblich hervorgehoben sind Sprachen der grössten Sprachfamilien.	70
Abb. 58: Verteilung der lokalen Sprachfamiliendichten über die Breitengrade. In blau die Werte der Südhalbkugel, in schwarze jene der Nordhalbkugel.	70

Abb. 59: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die Breitengrade. Farblich hervorgehoben sind Sprachen der grössten Sprachfamilien.	71
Abb. 60: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die Breitengrade. In blau die Werte der Südhalbkugel, in schwarze jene der Nordhalbkugel.	71
Abb. 61: Verteilung der lokalen Sprachfamiliendichten über die mittlere jährliche Temperatur.	72
Abb. 62: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die mittlere jährliche Temperatur.	72
Abb. 63: Verteilung der lokalen Sprachfamiliendichte über den jährlichen Temperaturschwankungsbereich.	73
Abb. 64: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über den jährlichen Temperaturschwankungsbereich in logarithmierter Darstellung.	74
Abb. 65: Verteilung der lokalen Sprachfamiliendichten über den jährlichen Niederschlag.	74
Abb. 66: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über den jährlichen Niederschlag.	75
Abb. 67: Verteilung der lokalen Sprachfamiliendichten über die Saisonalität des Niederschlags.	75
Abb. 68: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die Saisonalität des Niederschlags.	76
Abb. 69: Verteilung der lokalen Sprachfamiliendichte über die mittlere Nettoprimärproduktion.	76
Abb. 70: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die mittlere Nettoprimärproduktion.	77
Abb. 71: Verteilung der lokalen Sprachfamiliendichten über die Länge der Wachstumsphase.	77
Abb. 72: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die Länge der Wachstumsphase.	78
Abb. 73: Verteilung der lokalen Sprachfamiliendichten über die Küstendistanz.	78
Abb. 74: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die Küstendistanz.	79
Abb. 75: Verteilung der lokalen Sprachfamiliendichten über die Höhe über Meer.	79
Abb. 76: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Höhe über Meer.	80
Abb. 77: Verteilung der lokalen Sprachfamiliendichten über die maximale Höhendifferenz 50 km um die jeweilige Sprache.	80
Abb. 78: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die maximale Höhendifferenz 50 km um die jeweilige Sprache.	81
Abb. 79: Vergleich zwischen der GNNnv-Distanz in logarithmierter Darstellung und der Anzahl an Nachbarsprachfamilien in logarithmierter Darstellung.	90
Abb. 80: Vergleich zwischen der GNNnv-Distanz in logarithmierter Darstellung und der Anzahl an Nachbarsprachfamilien in logarithmierter Darstellung ohne grosse Sprachfamilien.	90
Abb. 81: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die Breitengrade. Farblich hervorgehoben sind die grössten Sprachfamilien.	95
Abb. 82: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die Breitengrade. In blau die Werte der Südhalbkugel, in schwarze jene der Nordhalbkugel.	96
Abb. 83: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die mittlere jährliche Temperatur.	96
Abb. 84: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über den jährlichen Temperaturschwankungsbereich in logarithmierter Darstellung.	97
Abb. 85: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über den jährlichen Niederschlag.	97
Abb. 86: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die Saisonalität des Niederschlags.	98
Abb. 87: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die mittlere Nettoprimärproduktion.	98
Abb. 88: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die Länge der Wachstumsphase.	99

Abb. 89: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die Küstendistanz.....	99
Abb. 90: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die Höhe über Meer.	100
Abb. 91: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die maximale Höhendifferenz 50 km um die jeweilige Sprache.....	100
Abb. 92: Vergleich zwischen der PNN-Distanz und mittleren Verwandtschaftsdistanz.	110

TABELLENVERZEICHNIS

Tab. 1: Anzahl Sprachen der Hammarström-Daten pro Subsistenzstrategie.	12
Tab. 2: Korrelationen zwischen den verschiedenen geographischen Faktoren inkl. geographischer Breite. Werte ≥ 0.5 resp. ≤ -0.5 sind fett hervorgehoben.	21
Tab. 3: Aufteilung der Methoden zur Beschreibung der Sprachendiversität in drei Teile.....	22
Tab. 4: Korrelationskoeffizienten des Zusammenhangs zwischen den Werten der geographischen Faktoren am Ort der Sprachen vs. des Durchschnitts in 50 km Umkreis. Alle Werte sind hoch signifikant.	36
Tab. 5: Übersicht über die Rangkorrelationskoeffizienten rho der lokalen Sprachendichte und der GNN-Distanz für jeden einzelnen geographischen Faktor. Die grössten Werte sind fett hervorgehoben. Kursive Werte sind nicht signifikant. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	54
Tab. 6: Multiple Regressionsanalyse mit der logarithmierten lokalen Sprachendichte als abhängige Variable und den geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	55
Tab. 7: Multiple Regressionsanalyse mit der logarithmierten lokalen Sprachendichte als abhängige Variable und sechs geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	56
Tab. 8: Multiple Regressionsanalyse mit der GNN-Distanz als abhängige Variable und den geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	57
Tab. 9: Multiple Regressionsanalyse mit der logarithmierten GNN-Distanz als abhängige Variable und sieben geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	58
Tab. 10: Mehrebenenanalyse der lokalen Sprachendichte mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	59
Tab. 11: Mehrebenenanalyse der lokalen Sprachendichte aufgeteilt nach Subsistenztyp Landwirtschaft und Jäger und Sammler. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	60
Tab. 12: Mehrebenenanalyse der GNN-Distanz mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	60
Tab. 13: Mehrebenenanalyse der lokalen Sprachendichte aufgeteilt nach Subsistenztyp. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	61
Tab. 14: Vergleich verschiedener Modelle zur lokalen Sprachendichte: Zwei multiple Regressionsmodelle und zwei Mehrebenenanalysen jeweils mit und ohne Subsistenzfaktor als Interaktion.	61
Tab. 15: Vergleich verschiedener Modelle zur lokalen Sprachendichte: Zwei multiple Regressionsmodelle und zwei Mehrebenenanalysen jeweils mit und ohne Subsistenzfaktor als Interaktion.	62
Tab. 16: Vergleich der Mehrebenenanalysen der lokalen Sprachendichte und GNN-Distanz. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	64
Tab. 17: Übersicht über die Rangkorrelationskoeffizienten rho der lokalen Sprachfamiliendichte und der GNNv-Distanz (mit und ohne grosse Sprachfamilien), für jeden einzelnen geographischen Faktor. Die grössten Werte sind fett hervorgehoben. Kursive Werte sind nicht signifikant. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	82
Tab. 18: Multiple Regressionsanalyse mit der lokalen Sprachfamiliendichte (ohne grosse Sprachfamilien) als abhängige Variable und den geographischen Faktoren als unabhängigen Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	83
Tab. 19: Multiple Regressionsanalyse mit der logarithmierten lokalen Sprachfamiliendichte als abhängige Variable und sechs geographischen Faktoren als unabhängige Variablen Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1	84

Tab. 20: Multiple Regressionsanalyse mit der GNNv-Distanz (ohne grosse Sprachfamilien) als abhängige Variable und den geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	85
Tab. 21: Multiple Regressionsanalyse mit der logarithmierten GNNv-Distanz als abhängige Variable und drei geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	85
Tab. 22: Mehrebenenanalyse der lokalen Sprachfamiliendichte mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	86
Tab. 23: Mehrebenenanalyse der Sprachfamiliendichte aufgeteilt nach Subsistenztyp Landwirtschaft und Jäger und Sammler. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	86
Tab. 24: Mehrebenenanalyse der GNNv-Distanz mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	87
Tab. 25: Mehrebenenanalyse der GNNv-Distanz aufgeteilt nach Subsistenztyp Landwirtschaft und Jäger und Sammler. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1. ...	87
Tab. 26: Vergleich verschiedener Modelle zur lokalen Sprachfamiliendichte: Zwei multiple Regressionsmodelle und zwei Mehrebenenanalysen jeweils mit und ohne Subsistenzfaktor als Interaktion.	88
Tab. 27: Vergleich verschiedener Modelle zur GNNv-Distanz: Zwei multiple Regressionsmodelle und zwei Mehrebenenanalysen jeweils mit und ohne Subsistenzfaktor als Interaktion.	89
Tab. 28: Vergleich der Mehrebenenanalyse der lokalen Sprachfamiliendichte und der GNNv-Distanz. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	91
Tab. 29: Übersicht über die Rangkorrelationskoeffizienten rho und deren Signifikanz der mittleren Verwandtschaftsdistanz und der PNN-Distanz für jeden einzelnen geographischen Faktor (sowohl für Hammarström- als auch AUTOTYP-Daten). Die grössten Werte sind fett hervorgehoben. Kursive Werte sind nicht signifikant. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	101
Tab. 30: Multiple Regressionsanalyse mit der mittleren Verwandtschaftsdistanz als abhängige Variable und den geographischen Faktoren als unabhängigen Variablen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	102
Tab. 31: Multiple Regressionsanalyse mit der logarithmierten mittleren Verwandtschaftsdistanz als abhängige Variable und fünf geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	103
Tab. 32: Multiple Regressionsanalyse mit der logarithmierten PNN-Distanz als abhängige Variable und den geographischen Faktoren als unabhängigen Variablen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	104
Tab. 33: Multiple Regressionsanalyse mit der logarithmierten PNN-Distanz als abhängige Variable und sechs geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	104
Tab. 34: Multiple Regressionsanalyse mit der logarithmierten PNN-Distanz der AUTOTYP-Daten als abhängige Variable und den geographischen Faktoren als unabhängigen Variablen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	105
Tab. 35: Multiple Regressionsanalyse mit der logarithmierten PNN-Distanz der AUTOTYP-Daten als abhängige Variable und sechs geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	105
Tab. 36: Mehrebenenanalyse der mittleren Verwandtschaftsdistanz mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	106
Tab. 37: Mehrebenenanalyse der mittleren Verwandtschaftsdistanz aufgeteilt nach Subsistenztyp Landwirtschaft und Jäger und Sammler. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 . <0.1 ' <1.	107

Tab. 38: Mehrebenenanalyse der PNN-Distanz mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 * <0.1 " <1.	107
Tab. 39: Mehrebenenanalyse der PNN-Distanz aufgeteilt nach Subsistenztyp Landwirtschaft und Jäger und Sammler. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 * <0.1 " <1.	108
Tab. 40: Mehrebenenanalyse der PNN-Distanz der AUTOTYP-Daten mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 * <0.1 " <1.	108
Tab. 41: Vergleich verschiedener Modelle zur mittleren Verwandtschaftsdistanz: Zwei multiple Regressionsmodelle und zwei Mehrebenenanalysen jeweils mit und ohne Subsistenzfaktor als Interaktion.	109
Tab. 42: Vergleich verschiedener Modelle zur PNN-Distanz: Zwei multiple Regressionsmodelle und zwei Mehrebenenanalysen jeweils mit und ohne Subsistenzfaktor als Interaktion.	109
Tab. 43: Vergleich zwischen dem multiplen Regressionsmodell und der Mehrebenenanalyse der PNN-Distanzen der AUTOTYP-Daten.	109
Tab. 44: Vergleich der Mehrebenenanalyse der mittleren Verwandtschaftsdistanz und der PNN-Distanz. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 * <0.1 " <1.	111
Tab. 45: Korrelation zwischen den verschiedenen Sprachenmasse. Für die lokale Sprachfamiliendichte und die GNNv-Distanz ist jeweils der erste Wert mit und der zweite ohne die grossen Sprachfamilien. Werte ≥ 0.5 resp. ≤ -0.5 sind fett hervorgehoben.	114
Tab. 46: Vergleich der standardisierten Koeffizienten der Mehrebenenanalyse zwischen den verschiedenen Sprachenmasse. Kursive Werte erreichen das Signifikanzniveau von 0.001 nicht.	115
Tab. XLVII: Mehrebenenanalyse der lokalen Sprachendichte aufgeteilt nach Subsistenztyp sesshafte Jäger und Sammler und Hirten. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 * <0.1 " <1.	125
Tab. XLVIII: Mehrebenenanalyse der GNN-Distanz aufgeteilt nach Subsistenztyp sesshafte Jäger und Sammler und Hirten. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 * <0.1 " <1.	126
Tab. XLIX: Mehrebenenanalyse der lokalen Sprachfamiliendichte aufgeteilt nach Subsistenztyp sesshafte Jäger und Sammler und Hirten. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 * <0.1 " <1.	126
Tab. L: Mehrebenenanalyse der GNNv-Distanz aufgeteilt nach Subsistenztyp sesshafte Jäger und Sammler und Hirten. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 * <0.1 " <1.	127
Tab. LI: Mehrebenenanalyse der mittleren Verwandtschaftsdistanz aufgeteilt nach Subsistenztyp sesshafte Jäger und Sammler und Hirten. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 * <0.1 " <1.	127
Tab. LII: Mehrebenenanalyse der PNN-Distanz aufgeteilt nach Subsistenztyp sesshafte Jäger und Sammler und Hirten. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 * <0.1 " <1.	128

I. Einleitung

1 MOTIVATION

Wer eine Weltkarte betrachtet, auf der alle rund 7'100 bekannten Sprachen (Lewis et al. 2013) abgebildet sind, stellt sofort fest, dass diese überhaupt nicht regelmässig auf der Welt verteilt sind. Je weiter entfernt vom Äquator, so scheint es, desto weniger dicht liegen die Sprachen beieinander. Zudem sind mehrere Gebiete zu erkennen, wo sich besonders viele Sprachen konzentrieren. So werden auf der Insel Neuguinea über 1'000 Sprachen gesprochen, während es in Europa, das mehr als zehn Mal so gross ist, weniger als 100 sind. Die Verteilung kann auch nicht einfach damit begründet werden, dass dort wo die Bevölkerungsdichte hoch ist, es die Sprachendichte auch ist. So werden beispielsweise im dünn besiedelten Australien mehr Sprachen gesprochen als im bevölkerungsreichsten Land der Welt in China (Lewis et al. 2013).

Es gibt verschiedenste Theorien und Hypothesen, die die unregelmässige Verteilung der Sprachen auf der Welt zu erklären versuchen. Da viele Klimafaktoren mit der geographischen Breite korrelieren, liegt die Vermutung nahe, dass hier ein Zusammenhang bestehen könnte. So gibt es viele Untersuchungen, die in diese Richtung zielen. Auch werden oftmals Faktoren herangezogen, die zur Erklärung der ebenfalls unregelmässigen Verteilung der Biodiversität gebraucht werden. Dies sind nebst Klimafaktoren, auch Faktoren, die wie beispielsweise die topographische Heterogenität, zu einer Isolierung über die Zeit führen (Moore et al. 2002). Andererseits wird auch versucht mittels historischer und kultureller Faktoren die Verteilung der Sprachendiversität auf der Welt zu erklären. In dieser Arbeit wird der Fokus auf die Umweltfaktoren gelegt.

In Zusammenarbeit mit dem Seminar für Allgemeine Sprachwissenschaft der Universität Zürich wird in einer interdisziplinären Arbeit versucht, die dargelegte Thematik mithilfe der Methoden und Instrumente der Geographischen Informationswissenschaft zu analysieren und neue Erkenntnisse zu gewinnen.

2 PROBLEMSTELLUNG

2.1 ZIELE

Der räumlichen Verteilung von Sprachen liegen verschiedenste Faktoren zugrunde. Dazu zählen auch geographische Faktoren, die offenbar mehr oder weniger dazu führen, dass Sprecher einer gemeinsamen Sprache sich voneinander isolieren und so neue Sprachen entstehen (Nettle 1998).

In dieser Arbeit geht es darum, verschiedene geographische Faktoren zusammenzustellen, zu modellieren und den Zusammenhang zur weltweiten Verteilung von Sprachen zu analysieren. Unter geographischen Faktoren werden Umweltbedingungen (z.B. humides Klima) wie auch Lagebedingungen (z.B. Küstennähe) verstanden. **Ziel** ist es, aufzuzeigen, ob und in welchem Masse die **Verteilungsmuster von Sprachen mit diesen geographischen Faktoren erklärt** werden können. Dabei werden die Verteilung der Sprachendiversität, der Sprachfamiliendiversität und die genealogische Diversität untersucht. Bei letzterem werden die Sprachendiversifikation und die Distanzen zwischen Verwandten unter verschiedenen geographischen Faktoren erforscht.

In der Linguistik gibt es bereits einige quantitative Analysen zum Zusammenhang zwischen geographischen Faktoren und der Sprachendiversität (Nettle 1998, Mace & Pagel 1995, Holman et al.

2007, Nichols 1990, Currie & Mace 2009, 2012, Moore et al. 2002). Viele der bisherigen Forschungen beschränkten sich allerdings auf einen Kontinent oder eine Region. In dieser Arbeit soll deshalb bewusst die ganze Welt als Betrachtungsebene dienen. Neu an dieser Arbeit wird sein, dass ein grosses Datenvolumen (7'035 Sprachdatensätze), Informationen zur genealogischen Struktur der Daten und verschiedene geographische Faktoren in Kombination mit den neusten GIS-Technologien eingesetzt werden, um quantitativ aussagekräftige Resultate zum Zusammenhang zwischen dem Verteilungsmuster von Sprachen und geographischen Faktoren machen zu können.

2.2 FORSCHUNGSFRAGEN

Die Ziele dieser Masterarbeit können in verschiedenen Forschungsfragen formuliert werden. Die erste Forschungsfrage ist dabei sehr weit gefasst. Sie ist jedoch zentral, da sie alle weiteren Fragen umrahmt.

- Gibt es einen Zusammenhang zwischen der weltweiten Verteilung von Sprachen und geographischen Faktoren?

Zur Erklärung der Verteilung von Sprachen können verschiedenste Faktoren wie beispielsweise historische, wirtschaftliche oder eben geographische herbeigezogen werden. In dieser Arbeit wird der Fokus bewusst auf nur eine Faktorenkategorie gelegt. Damit wird aber keineswegs der Einfluss anderer Faktoren auf die Verteilung von Sprachen abgesprochen.

Die folgenden Forschungsfragen formulieren die erste weiter aus:

- Bei welchen geographischen Faktoren oder Konstellationen von Faktoren kann eine hohe Sprachendiversität respektive Sprachfamiliendiversität erwartet werden?

Mit anderen Worten: Gibt es einen Zusammenhang zwischen der Verteilung der Sprachendichten und bestimmten Umweltbedingungen? Und bestehen dieselben Zusammenhänge, wenn anstatt der Sprachendichte, die Sprachfamiliendichte untersucht wird? Wo existieren allenfalls Unterschiede? Wie bereits erwähnt, gibt es schon einige quantitative Untersuchungen in diese Richtung. In dieser Arbeit soll mithilfe von GIS-Methoden und möglichst hochauflösenden räumlichen Daten eine quantitative Antwort auf die gestellte Forschungsfrage gefunden werden. Auf dieselbe Art wird eine Antwort auf folgende Forschungsfrage gesucht:

- Gibt es einen Zusammenhang zwischen den geographischen Umweltbedingungen und der genealogischen Diversität?

Unter welchen Umweltbedingungen findet eine höhere Sprachendiversifikation statt und hat es deshalb viele eng verwandte Sprachen geographisch nahe beieinander? Welche Faktoren begünstigen eine hohe, welche eine tiefe Dichte? Sind die Distanzen zu den nächsten verwandten Sprachen beim Vorkommen von bestimmten Umweltbedingungen deutlich länger respektive kürzer? Dies soll Rückschlüsse darauf geben, ob sich Sprachen unter gewissen Umweltbedingungen möglicherweise auf kürzere Distanz separieren und ob eine hohe Differenzierungsrate vorherrscht.

II. Hintergrund

Im ersten Kapitel dieses Teils folgt eine kurze Einführung in die Grundlagen der Linguistik (*Kapitel 3*), die nötig sind um die Arbeit verstehen zu können. Der zweite Teil *Kapitel 4 Forschungsstand* arbeitet die bisher erschienene Literatur zum Thema auf und der dritte Teil *Kapitel 5 Forschungslücke* zeigt auf, wo noch Forschungslücken bestehen.

3 GRUNDLAGEN LINGUISTIK

Wie bereits in der Problemstellung erwähnt, gibt es heute rund 7'100 Sprachen (Lewis et al. 2013). Es dürften jedoch noch mehr sein, da in schwer zugänglichen Regionen noch unbekannte Sprachen vermutet werden (McGregor 2009). Die Anzahl Sprecher dieser vielen verschiedenen Sprachen ist dabei sehr unterschiedlich auf die einzelnen Sprachen verteilt. So sprechen mehr als 40 % der Erdbevölkerung eine von neun Sprachen mit mehr als 100 Millionen Sprechern. Hingegen haben über 3'000 Sprachen weniger als 10'000 Sprecher und machen zusammen lediglich 0.3 % der Weltbevölkerung aus (McGregor 2009).

3.1 EINHEIT UND DIVERSITÄT DER SPRACHEN

Wann spricht man überhaupt von einer eigenständigen Sprache und wann sind es nur Dialekte einer gemeinsamen Sprache? Eine einfache Einteilung erfolgt darüber, ob die Sprachen das Kriterium der gegenseitigen Verständlichkeit (*mutual intelligibility*) erfüllen. Wenn zwei Sprecher sich gegenseitig verstehen ohne die Sprache des anderen gelernt haben zu müssen, sind die Sprachen der Sprecher *gegenseitig verständlich* und sind somit Dialekte einer einzigen Sprache. Verstehen sich die beiden Sprecher nicht (*mutually unintelligibility*), handelt es sich um zwei verschiedene Sprachen (McGregor 2009).

3.2 SPRACHEN VERÄNDERN SICH

Sprachen sind nicht statisch, sondern verändern sich über die Zeit. Das Tempo mit dem sie dies tun, kann dabei sehr unterschiedlich sein. Schon innerhalb weniger Jahrhunderte kann sich eine Sprache deutlich verändern. Im Verlauf der Zeit kommen beispielsweise neue Wörter in Gebrauch, andere werden nicht mehr verwendet, Wörter werden abgekürzt, zusammengesetzt oder neu erfunden. Es finden aber auch morphologische und syntaktische Veränderungen (z.B. Veränderung der Wortreihenfolge) sowie Lautveränderung statt (McGregor 2009).

3.2.1 Gründe für die Sprachveränderung

Weshalb verändern sich Sprachen überhaupt? Es gibt verschiedene psychologische, ethnische, soziale und geographische Erklärungsansätze. Meistens ist es eine Kombination von Faktoren, die zu einer Sprachveränderung führt. Beispielsweise tritt Sprachveränderung auf, wenn ein intensiver Kontakt zwischen Sprechern verschiedener Sprachen besteht. Eine Sprachveränderung kann auch durch die Sprecher motiviert sein. Eine unterscheidbare Sprache führt zu einer stärkeren Gruppenidentität und führt zu Abgrenzung gegenüber anderen (McGregor 2009).

Veränderungen innerhalb einer Sprache treffen oft nicht auf die ganze Gruppe der Sprecher zu. Oftmals führen soziale und/oder geographische Separationen einer Gruppe von Sprechern dazu, dass sich deren Sprache unabhängig verändert und weiter entwickelt. So entstehen aus einer gemeinsamen Sprache Dialekte und mit der Zeit zwei verschiedene Sprachen, die gegenseitig unverständlich sind. Die **soziale**

und geographische Separation von Sprechern gemeinsamer Sprachen hat zu der heutigen enormen Sprachendiversität beigetragen.

3.3 VERWANDTSCHAFT VON SPRACHEN

Um die Vielzahl an Sprachen zu gruppieren und zu kategorisieren gibt es nebst morphologischen, syntaktischen und phonologischen Merkmalsbeschreibungen auch die Möglichkeit Sprachen mittels Verwandtschaft zu differenzieren. Der Begriff *Sprachfamilie* ist dabei der zentrale Punkt um die Verwandtschaft von Sprachen zu erklären. Sprachen die von einer gemeinsamen Sprache abstammen, einer so genannten Protosprache, sind genealogisch (d.h. von der Entstehung her) verwandt und gehören zur selben Sprachfamilie. Oft werden in diesem Zusammenhang auch die Bezeichnungen *phylogenetisch* oder *genetisch verwandt* verwendet, die beide aussagen, dass diese Sprachen eine gemeinsame Vorgängersprache haben. Die Sprachen einer Sprachfamilie entstehen durch das Divergieren der Protosprachen und den wiederum daraus entstehenden Sprachen. Ungefähr bis 10'000 Jahre vor heute kann eine gemeinsame Protosprache nachgewiesen werden. Es gibt einige Sprachen, so genannte *Isolates*, die zu keiner anderen Sprache genealogisch verwandt sind und somit ganz allein eine Sprachfamilie bilden (z.B. Baskisch). Die verschiedenen Sprachfamilien werden meist noch weiter in Subfamilien und –gruppen unterteilt, welche die genealogische Verwandtschaft der Sprachen innerhalb einer Familie aufzeigen (McGregor 2009). Diese Struktur wird oft in einem sogenannten Familienbaum-Modell abgebildet, welches die Aufspaltung der Protosprache bis zu den heutigen Sprachen einer Familie darstellt. Es gibt über 300 Sprachfamilien (Nichols & Bickel 2009, Hammarström 2012). Jedoch sind sich nicht alle Forscher über diese Zahl einig, da sie stark davon abhängt, wie grosszügig man die Sprachen in einer Familie zusammenfasst (McGregor 2009). Die grössten Sprachfamilien mit über 300 Sprachen und insgesamt 80 % der Weltbevölkerung sind: Indogermanisch, Austronesisch, Afroasiatisch, die Niger-Kongo-Sprachen, Sinotibetisch und die Trans-Neuguinea-Sprachen (Lewis et al. 2013).

Zur Sprachfamilie Indogermanisch gehören beispielsweise über 400 verschiedene Sprachen, darunter Deutsch, Lateinisch, Russisch, Paschtu, Kurdisch und Persisch. Die Ursprungsregion der Protosprache liegt wahrscheinlich in Südost-Europa und wurde dort vor circa 6'000-8'000 Jahren gesprochen. Von dort aus verbreitet sie sich nach Osten und Westen und teilte sich in immer mehr gegenseitig unverständliche Sprachen auf (McGregor 2009).

3.4 PROBLEMATIK DER GENEALOGISCHEN VERWANDTSCHAFT

Die Darstellung genealogischer Verwandtschaft mithilfe von Sprachfamilien und Subgruppen vereinfacht die Realität, denn Sprachen interagieren stark mit anderen Sprachen in ihrer Umgebung. Es können auch Sprachen aus der Interaktion zweier oder mehrerer Sprachen entstehen. Pidgin-Sprachen, Kreolsprachen und gemischte Sprachen entstehen durch solche Prozesse. *Pidgin-Sprachen* sind vereinfachte Sprachen, die oft entstehen wenn Sprecher verschiedener Sprachen nur für bestimmte Interaktionen miteinander kommunizieren. Viele dieser Sprachen entstanden beispielsweise während der europäischen Kolonialisierung. Die Wörter dieser Sprachen entstammen oftmals dem Vokabular der Kolonialisten, während die Grammatik und die Betonung von der Sprache der Einheimischen übernommen wird. Gewinnen diese Sprachen an Wortschatz und werden sogar zu Muttersprachen, werden sie *Kreolsprachen* genannt (McGregor 2009).

Für viele Sprachfamilien ist der Aufbau des Familienbaum-Modells, welches die genealogische Verwandtschaft abbildet, umstritten und zum Teil unklar. Zudem sind die Protosprachen bei vielen Sprachfamilien nicht rekonstruiert worden oder rekonstruierbar. Das Familienbaum-Modell vermittelt einen Überblick darüber, wie Sprachen divergierten, stösst jedoch, wenn es um die Abbildung aller Typen von Sprachveränderung geht, an seine Grenzen (Nettle 1999).

4 FORSCHUNGSSTAND

4.1 GEOGRAPHISCHE FAKTOREN UND SPRACHENDIVERSITÄT

4.1.1 Verschiedene Erklärungsansätze

Wie weiter unten gezeigt wird, wurden schon viele Publikationen erstellt, in denen nach einer Erklärung für die unregelmässige Verteilung der Sprachendiversität oder auch der ethnischen Diversität auf der Welt gesucht wird. Bereits 1953 stellte Birdsell fest, dass die natürliche Umwelt einen Einfluss auf die Grösse von ethnischen Gruppen hat (Birdsell 1953). Bei seinen Untersuchungen zu den Aborigines Australiens entdeckte er, dass die Gruppengrösse von Stämmen negativ mit dem dortigen Niederschlag korrelierte. Zudem konnte er zeigen, dass bei geringerem Niederschlag auch das Gebiet, welches die Stämme besetzen, grösser ist. Was wiederum bedeutet, dass dort eine tiefere ethnische Diversität zu finden ist.

Auch bei den weiteren Publikationen, die auf Birdsell folgten, stand als Einflussgrösse meist ein geographischer Faktor wie beispielsweise Niederschlag, Temperatur oder Küstennähe im Fokus. Die Sprachendiversität, als davon abhängige Variable, wird meist in irgendeiner Form von Sprach- oder Sprachfamiliendichte erfasst. Dabei fokussieren alle Analysen auf die Frage was darüber entscheidet, wie die Sprachendiversität auf der Welt verteilt ist.

Verschiedene Autoren, die zu diesem Thema forschten, verfolgten unterschiedliche Erklärungsansätze und stellten deswegen unterschiedliche Faktoren, welche die Sprachendiversität oder allgemein die Diversität von ethnischen Gruppen beeinflussen, in den Fokus. Currie & Mace (2012) fassen dazu drei Typen von Hypothesen zusammen:

- Topographische Hypothese
- Umweltrisiko-Hypothese
- Umweltproduktivität-Hypothese

Topographische Hypothese: Anhänger dieser Hypothese sehen die Barrierewirkung von topographischen Elementen wie beispielsweise Berge verantwortlich dafür, dass sich Gesellschaften voneinander separieren (Currie & Mace 2012). Es wird deshalb in bergigen Regionen eine höhere Sprachendiversität erwartet.

Umweltrisiko-Hypothese: Diese Hypothese besagt, dass die Sprachendiversität vom Risiko bestimmt wird nicht genügend Existenzgrundlagen zu haben. Deswegen sind in Regionen, in denen das Umweltrisiko hoch ist, die sozialen Netzwerke grösser. Dies ist notwendig um möglichen Verknappungen in der Produktion entgegen zu wirken. Durch den häufigen Kontakt separieren sich die Gesellschaften weniger (Currie & Mace 2012).

Umweltproduktivität-Hypothese: Diese Hypothese besagt, dass die ethnische Diversität von der Produktivität der Umwelt abhängt. Da die Mitgliedschaft in einer Gruppe sowohl Vor- als auch Nachteile mit sich bringt, sollte eine Gruppe gerade so gross sein, dass sie Nahrungssicherheit und Fortbestand gewährleisten kann. In einer produktiven Umwelt wird daher eine höhere Diversität an ethnischen Gruppen erwartet, da diese Umgebung genug hergibt um eine höhere Bevölkerungsdichte respektive viele kleine Gruppen zu tragen (Currie & Mace 2012).

4.1.2 Methoden und Erkenntnisse der verschiedenen Erklärungsansätze

Was sofort auffällt, wenn die Sprachenverteilung auf einer Weltkarte betrachtet wird, ist, dass es in Äquatornähe viel mehr Sprachen gibt als anderswo auf der Erde. Collard & Foley (2002) beschreiben in

ihrer Publikation denn auch den deutlichen Breitengrad-Gradienten in der weltweiten kulturellen Diversität. Mace & Pagel (1995) zeigten zudem, dass in Nordamerika die Sprachendichte mit der geographischen Breite sinkt, während die Fläche über die eine Sprache gesprochen wird, grösser wird. Der Breitengrad-Gradient in der Sprachendiversität führte zu vielen Erklärungsversuchen. Cashdan (2001) stellte dazu fest, dass es schwierig sei die wichtigsten klimatischen Faktoren zu isolieren, da viele davon mit der geographischen Breite korrelieren. Nachfolgend sind die verschiedenen überprüften Faktoren, Methoden und Erkenntnisse nach der Zugehörigkeit zur jeweiligen Hypothese genauer erläutert.

4.1.2.1 Topographie

Durch die Barrierewirkung von Bergen wird ein Einfluss auf die Sprachendiversität vermutet. Diese Annahme wurde schon mehrfach überprüft. Cashdan (2001) stellte fest, dass es einen kleinen Zusammenhang zwischen dem Gefälle von Landschaften und der ethnischen Diversität gibt. Currie & Mace (2009) prüften ob in bergigen Regionen die Ausdehnungsgebiete von Sprachen durch die Barrierewirkung von Bergen kleiner wird. Dabei ergab sich eine signifikante Korrelation, jedoch dahingehend, dass in Gebirgen die Flächen von Sprachen grösser werden. Also genau dem entgegengesetzt, was erwartet wurde (Currie & Mace 2009). Bei Moore et al. (2002) zeigte sich für Afrika ebenfalls kein Zusammenhang zwischen der Sprachendiversität und der topographischen Heterogenität.

4.1.2.2 Ökologisches Risiko

Wie erwähnt wird in Regionen mit grossem ökologischem Risiko eine tiefere Sprachendiversität erwartet, da die Gesellschaften dort auf soziale Netzwerke angewiesen sind und sich dementsprechend nicht voneinander isolieren. Nettle (1998) definiert das *ökologische Risiko*, wie er es nennt, als die Summe der Variation, die sich bei der Nahrungsversorgung stellt. Diese Variabilität kann saisonal sein oder auch mehrjährig. Er begründet die Umweltrisikohypothese damit, dass dort, wo es das Klima ermöglicht über das ganze Jahr hinweg zu produzieren, sich eher kleine Gruppen von Personen selbstversorgen können. Und so teilen sich die Menschen in diesen Gebieten in kleine soziale und ökonomische Gesellschaften mit verschiedenen Sprachen auf. In Regionen hingegen, in denen die Klimavariabilität gross ist, brauchen Gesellschaften ein grosses soziales Netzwerk um gegen das Risiko von schlechten Ernten abgesichert zu sein. In solchen Gebieten sind einzelne Sprachen deshalb über grössere Gebiete verteilt und haben mehr Mitglieder. Nettle belegt seine Theorie damit, dass die Länge der Anbauphase pro Jahr mit der Sprachendiversität steigt. Einige kritische Anmerkungen zu seiner und weiteren Methoden sind im Abschnitt 4.4 *Kritische Beurteilung der bisherigen Literatur* zu finden. Currie & Mace (2009) nahmen ebenfalls die Länge der Anbauphase pro Jahr als Prädiktor. Als abhängige Variable nahmen sie jedoch die Gebietsgrösse einer Sprache und nicht die Dichte. Sie fanden einen signifikanten, aber schwachen, positiven Zusammenhang. Auch Moore et al. (2002) konnten für Afrika eine positive Korrelation zwischen der Sprachendiversität und der Länge der Anbauphase nachweisen. Weiter werden auch die Variabilität des Niederschlags und des Temperaturschwankungsbereichs zur Überprüfung des Umweltrisikos herbeigezogen. Die Sprachendiversität sinkt in Afrika, wenn die Variabilität dieser beiden Faktoren grösser wird. Signifikant ist dieser Zusammenhang allerdings nur für die jährliche Niederschlagsvariabilität. Weiter zeigte Cashdan (2001), dass in Regionen mit tiefer ethnischer Diversität sehr variable klimatische Bedingungen vorherrschen.

4.1.2.3 Produktivität der Umwelt

Nettle (1998) begründet seine Vermutung zwar mit dem ökologischen Risiko, seine Überlegungen zur Möglichkeit der Selbstversorgung von Gruppen überschneiden sich jedoch mit der Umweltproduktivitätshypothese.

Cashdan (2001) konnte keinen oder nur einen sehr kleinen Zusammenhang zwischen der ethnischen Diversität und dem jährlichen Niederschlag einerseits und der Temperatur andererseits finden. Im Gegensatz dazu fanden Collard & Foley (2002), die 3'800 Kulturen für ihre Untersuchung verwendeten, heraus, dass die kulturelle Diversität mit der Temperatur und insbesondere mit dem Niederschlag ansteigt.

Currie & Mace (2009) verwendeten als Mass der Produktivität der Umwelt die Nettoprimärproduktion (NPP), also die Menge an Biomasse die während der Photosynthese aus Sonnenenergie entsteht. Der Zusammenhang zur Gebietsgrösse der Sprachen war allerdings nicht signifikant. Als sie jedoch die Sprachen anhand der Versorgungsstrategie (Jäger und Sammler, Hirten, Landwirte) unterteilten, erhielten sie signifikante Resultate (Currie & Mace 2012). Sie konnten aufzeigen, dass es vor allem in Jäger-und-Sammler-Gesellschaften einen Zusammenhang zwischen der Gebietsgrösse einer Sprache und der Nettoprimärproduktion, dem Niederschlag und der Temperatur gibt. Bei Gesellschaften die hauptsächlich von Landwirtschaft leben, konnten sie dies nicht nachweisen. Jäger und Sammler sowie Hirten sind also eher von der Produktivität der Umwelt abhängig. In der Landwirtschaft hingegen spielt diese nicht mehr eine so grosse Rolle, da dort für verschiedene klimatische Bedingungen verschiedene Pflanzen verwendet werden. Daraus schlussfolgerten sie unter anderem, dass die Faktoren, welche die ethnische Diversität beeinflussen, sich durch die Landwirtschaft stark verändert haben (Currie & Mace 2012). Für den Kontinent Afrika konnte gezeigt werden, dass die Sprachendiversität sowohl mit der Temperatur, dem Niederschlag und der NPP steigt. Signifikant war allerdings nur der Zusammenhang mit dem Niederschlag (Moore et al. 2002).

Nebst diesen klimatischen Faktoren wurden auch die Bio- und Habitatdiversität als Prädiktoren überprüft. So zeigte sich, dass die linguistische Diversität in Nordamerika in Regionen mit einer höheren Habitatdiversität grösser ist (Mace & Pagel 1995). Cashdan (2001) zeigte ebenfalls, dass die ethnische Diversität von nicht politisch komplexen ethnischen Gruppen in Regionen mit einer höheren Habitatdiversität grösser ist. Sie begründete das damit, dass dort eine grössere ökonomische Nischenspezialisierung möglich ist. Möglicherweise führen dieselben Faktoren, welche auch zu Habitatdiversität führen, zu höherer Sprachendiversität (Mace & Pagel 1995). Für Afrika konnte kein signifikanter Zusammenhang zwischen der Sprachendiversität und der Lebensraum-Heterogenität, dafür zwischen der Sprachen- und Biodiversität gefunden werden (Moore et al. 2002). Die Biodiversität in der Alten Welt, in diesem Fall geprüft mit der Pflanzendiversität, war nur ein schwacher bis gar kein Prädiktor für Sprachendiversität (Currie & Mace 2009).

4.1.2.4 Weitere Prädiktoren

Neben ökologischen und geographischen Faktoren, gibt es auch weitere Faktoren, welche als Prädiktoren für die Sprachendiversität geprüft wurden. Currie & Mace (2009) zeigten, dass der Grad der politischen Komplexität ein wichtiger Prädiktor für die Gebietsgrösse einer Sprache ist. In einem *Linear Mixed Model*, welches sie aufstellten, konnten die Umweltfaktoren im Gegensatz zur politischen Komplexität nur einen kleinen Teil der Varianz erklären. Ihr Fazit ist, dass kulturelle Faktoren die Gebietsgrösse von Sprachen besser erklären können als Umweltfaktoren. Ein Prozess von kultureller Gruppenselektion habe möglicherweise komplexere Gesellschaften bevorzugt und so zur heutigen Sprachverteilung geführt.

Hoher *Pathogen-Stress* (z.B. durch Malaria) ist bei Cashdan (2001) einer der besten Prädiktoren für hohe ethnische Diversität. Ihrer Meinung nach macht es das hohe Vorkommen an Pathogenen um den Äquator schwierig dort grosse Siedlungen zu bilden, was wiederum die Grösse von sozialen Gruppen einschränkt. Da auch der Pathogen-Stress zum Äquator hin ansteigt und dieser selbst stark abhängig ist von anderen Faktoren, wie beispielsweise Niederschlag, ist dieser Faktor, obwohl er immer wieder geprüft wird, äusserst umstritten (Currie & Mace 2009, Nettle 2009).

4.2 GEOGRAPHISCHE FAKTOREN UND SPRACHFAMILIENDIVERSITÄT

Anstatt der Sprachendiversität oder der ethnographischen Diversität wurde auch die Verteilung der Sprachfamilien im Zusammenhang mit geographischen Faktoren untersucht. So berechnete beispielsweise Nichols (1990) die Sprachfamiliendichte (Sprachfamilien pro Mio. Quadratmeile) für verschiedene Regionen und Kontinente. Weiter verglich sie die Anzahl an Ästen die von den verschiedenen Sprachfamilien ausgehen. Aufgrund von Beobachtungen schlug sie mehrere geographische Faktoren vor, welche hohe oder tiefe Sprachfamiliendichten begünstigen. Dazu gehören: Die

geographische Breite, Küsten, Niederschlag und Berge. An den Küsten ist die Dichte zwar höher, aber es herrscht nicht an allen Küsten eine hohe Dichte. Die Gründe für die hohen Dichten in Küstenregionen, sieht sie in den ganzjährig zur Verfügung stehenden Proteinquellen, in der Reduzierung von extremen Klimata durch den Ozean und in der Möglichkeit der Handelsverbindungen. Den Niederschlag sieht sie vor allem im Zusammenhang mit der Küste als Erklärungsfaktor. Dort, wo es wenig Niederschlag an Küsten gibt, ist die Sprachfamiliendichte tiefer. Weiter sieht sie Berge, durch ihre isolierende Wirkung, als Grund für hohe Sprachfamiliendichten. Die Einflüsse all dieser Faktoren können ihrer Meinung nach mit der ökonomischen Skala erklärt werden. Denn sämtliche Faktoren begünstigten kleinräumige Wirtschaftssysteme.

4.3 GEOGRAPHISCHE FAKTOREN UND GENEALOGISCHE DIVERSITÄT

Bisher wurde zur genealogischen Diversität und der Sprachendiversifikation unter bestimmten Umweltbedingungen nur sehr wenig geforscht. Von Currie & Mace (2012) wurde die Wichtigkeit Forschungen in diesem Bereich zu machen, damit begründet, dass man so in Erfahrung bringen könnte, wie sich die Diversifikation bei einer Migration in eine neue Umwelt verhält. Einzig Holman et al. (2007) führten Untersuchungen in diese Richtung durch. Sie interessierten sich für die Distanzen zwischen verwandten Sprachen und konnten eine leichte Tendenz nachweisen, dass in tropischen Regionen Sprachen auf kürzeren Distanzen weniger ähnlich sind als andernorts. Sprachen in tropischen Regionen entfernen sich also möglicherweise schneller voneinander als in nicht-tropischen Regionen.

4.4 KRITISCHE BEURTEILUNG DER BISHERIGEN LITERATUR

Die vorgestellten Methoden und Resultate zeigen, dass schon einiges zum Zusammenhang zwischen Sprachendiversität und geographischen Faktoren geforscht wurde, dabei jedoch oft widersprüchliche Resultate herauskamen. Das ist unter anderem auf unterschiedliche Datengrundlagen, Versuchsregionen und Analysemethoden zurückzuführen. Ähnlich wie es auch schon von Currie & Mace (2012) diskutiert wurde, folgen in diesem Abschnitt ein paar kritische Anmerkungen zu den Methoden und Grundlagen der vorgestellten Texte und auch Bemerkungen, weshalb diese, im Zusammenhang mit Sprachendiversität und geographischen Faktoren, nur bedingt aussagekräftig sind.

4.4.1 Indirektes Mass für Sprachendiversität

In vielen Publikationen wird entweder nicht direkt die Sprachendiversität untersucht, sondern ein anderes Phänomen welches indirekt mit Sprachen zusammenhängt, oder es wird nur ein indirektes Mass für die Sprachendiversität verwendet.

So interessierten sich Holman et al. (2007) nicht für die Sprachendiversität, sondern für die strukturelle Diversität von Sprachen im Zusammenhang mit geographischen Distanzen unter unterschiedlichen Umweltbedingungen. Zudem begrenzten sie ihre Untersuchung auf 2'488 Sprachen. Collard & Foley (2002) verwendeten für ihre Analysen über 3'800 Kulturen, die daraus resultierenden Ergebnisse können aber nicht direkt auf die Sprachendiversität übertragen werden, da Kulturen nach mehr Kriterien als nur der Sprache definiert werden. Zudem verwendeten Currie & Mace (2009, 2012) nur einen Indikator, nämlich die Gebietsgrösse einer Sprache, für die Sprachendiversität. Auf diese wird im nächsten Abschnitt (*Abschnitt 4.4.2*) noch genauer eingegangen.

4.4.2 Unterschiedlich aggregierte Sprachendiversität und geographische Faktoren

Wie schon Currie & Mace (2012) festgestellt haben, wird in einigen Analysen ein seltsames Vergleichslevel oder Mass für die Sprachendiversität verwendet. Nettle (1998) verglich mittels einer Regressionsgerade die Anzahl an Sprachen in einem Land mit der Länderfläche. Auch andere Autoren berechneten die Sprachendiversität aufgrund der Anzahl Sprachen in einem bestimmten Gebiet, wie beispielsweise Kontinente, Regionen etc. (Nichols 1990). Andere wiederum bildeten Breitengradbänder und berechneten pro Band die Dichte der ethnischen Gruppen (Collard & Foley 2002) oder zählten alle 2° wie viele Sprachen sich mit diesem Breitengrad schnitten (Mace & Pagel 1995). Moore et al. (2002) berechneten ein 2° × 2° Raster und zählten pro Zelle, die darin vorkommenden Sprachen. Das Problem das diese Methode birgt, ist, dass die Zellen vom Äquator weg flächenmassig kleiner werden. Für die geographischen Faktoren, die mit der Sprachendiversität verglichen werden sollten, wurden dann oft auch die Durchschnittswerte einer solchen Aggregationsstufe verwendet.

Das Problem beim Aggregieren von Daten ist, dass dabei die Methode nach welcher die Daten zusammengefasst werden (z.B. nach Kontinenten), einen entscheidenden Einfluss auf das Resultat haben kann. Dieses Phänomen ist auch unter dem Namen *The Modifiable Unit Problem* bekannt (Openshaw 1983, O'Sullivan & Unwin 2010). Zudem werden Korrelationen oft verstärkt, da das Aggregieren auch dazu führt, dass Extreme abgeflacht werden, da der Durchschnitt berechnet wird.

Currie & Mace (2009, 2012) konnten dieses Problem teilweise umgehen, indem sie als indirektes Mass für die Sprachendiversität die Grösse des Gebiets, über welches eine Sprache verteilt ist, verwendeten. Während die Sprachendiversität mit höherer geographischer Breite sinkt, wird die Fläche, über die eine Sprache gesprochen wird, grösser (Mace & Pagel 1995). Currie & Mace (2009) schlossen daraus, dass in Gegenden in denen die Sprachendiversität hoch ist, die Gebietsgrössen von Sprachen eher kleiner sind und dort wo die Diversität klein ist, die Gebiete eher grösser sind. Jedoch können ihre Resultate nicht direkt als Schlussfolgerungen für die Sprachendiversität gebraucht werden, da sie den direkten Zusammenhang zwischen Diversität und Gebietsgrösse nicht überprüften. Trotzdem schliessen sie von ihren Resultaten direkt darauf.

4.4.3 Räumlich begrenzte Analysen

Viele Analysen konzentrieren sich bisher auf eine bestimmte Region oder sie bezogen nur Gebiete ein, die bestimmte Bedingungen erfüllten. Nettle (1998) verwendete nur Länder die grösser als 50'000 km² sind und in tropischen Regionen liegen. Mace & Pagel (1995) konzentrierten ihre Analysen auf Nordamerika, Nichols (1990) hingegen auf ganz Amerika und Moore et al. (2002) auf Afrika. Currie & Mace (2009, 2012) berechneten die Korrelationen separat für die Neue Welt, die Alte Welt und Australien. Unter anderem weil sie noch die Versorgungsstrategie mit berücksichtigten und diese Information in den entsprechenden Regionen verschieden gut bekannt ist. Zusätzlich wurden Sprachen, die auf Inseln mit keiner weiteren Sprache gesprochen werden, ausgeschlossen.

5 FORSCHUNGSLÜCKE

Viele Untersuchungen zum Zusammenhang zwischen geographischen Faktoren und der Sprachendiversität sind entweder räumlich beschränkt, untersuchen nur gewisse Prädiktoren, verwenden ein indirektes Mass für die Sprachendiversität oder gehen nicht auf die Genealogie ein. Entsprechend kann hier angesetzt werden, indem eine Analyse auf globaler Ebene mit mehreren Prädiktoren, ein verfeinertes Mass für die Sprachendiversität und Informationen zur Verwandtschaft der Sprachen verwendet werden.

Insbesondere der letzte Punkt ist zentral, da bisherige Analysen kaum auf die genealogische Diversität von Sprachen eingegangen sind. Nichols (1990) bildet hier eine Ausnahme. Sie machte im Zusammenhang mit Sprachendiversifikation die wichtige Überlegung, dass beim Wechsel von Faktoren, die eine hohe Dichte begünstigen, zu solchen die eine tiefe Dichte begünstigen, ein Sinken der Differenzierungsrate einer Sprache zu erwarten ist (Nichols 1990). Currie & Mace (2012) messen Analysen in dieser Richtung grosses Potenzial bei. Diese ermöglichen es herauszufinden, welche geographischen Faktoren die kulturelle und linguistische Gruppendiversifikation beeinflussen. Zudem können solche Analysen zeigen, wie sich die Ethnogenesisrate bei einer Migration in eine neue Umwelt verändert (Currie & Mace 2012). In diesem Bereich kann somit neues Wissen beigetragen werden. Einerseits indem auch die Sprachfamiliendiversität analysiert wird, andererseits indem auf einer noch tieferen Verwandtschaftsebene die Diversität und die Sprachendiversifikation in Abhängigkeit von geographischen Faktoren untersucht wird. Im Bereich der kulturellen Phylogenetik gibt es einige Arbeiten mit phylogenetischen Ansätzen, jedoch nicht im Zusammenhang mit geographischen Faktoren und genealogischer Diversität (Currie & Mace 2012, Mace & Jordan 2011, Mace & Holden 2005). Holman et al. (2007) beziehen geographische Distanzen und geographische Faktoren in ihre Analysen mit ein, jedoch steht bei ihnen die typologische Differenz zwischen Sprachen im Vordergrund. Zudem überprüften sie den Zusammenhang zwischen der geographischen und typologischen Distanz nur unter den Faktoren tropisch/nicht-tropisch und Küste/Inland.

III. Forschungsansatz

In diesem Teil der Arbeit werden die Datengrundlagen für die Analysen vorgestellt und es wird erläutert, wie diese aufbereitet wurden. Kapitel 6 *Linguistische Datengrundlagen* stellt die zwei verwendeten Sprachdatensätze vor. Das folgende Kapitel (7 *Grundlagen geographische Faktoren*) führt in die verwendeten geographischen Faktoren ein.

6 LINGUISTISCHE DATENGRUNDLAGEN

In dieser Arbeit werden zwei unterschiedliche Sprachdatenbanken verwendet. Die zweite, die *AUTOTYP-Datenbank (Unterkapitel 6.2)*, wird allerdings erst bei den Auswertungen zur genealogischen Diversität gebraucht, um die verwendete Methode besser abzustützen. Da es dort um genealogische Verwandtschaften geht, hängt diese stark von der jeweilig getätigten Einteilung des Erstellers der Taxonomie ab. So soll gesichert werden, dass die Resultate nicht nur ein Produkt der Systematik des Autors sind.

6.1 HAMMARSTRÖM - DATEN

6.1.1 Herkunft / Aufbau

Die meisten Analysen in dieser Arbeit werden mit Daten, die von Harald Hammarström (Max-Planck-Institut für evolutionäre Anthropologie, Leipzig) zusammengestellt wurden, gemacht. Darin sind 7'109 Sprachen und ihre Koordinaten im WGS84 (World Geodetic System von 1984) enthalten. Die Koordinaten stammen dabei vom Ethnologue (Lewis et al. 2013), wurden jedoch von Hammarström korrigiert und ergänzt. Eine dazugehörige Taxonomie (Hammarström 2012, Nordhoff et al. 2013) teilt die Sprachen in verschiedene Sprachfamilien und weiter in mehrere Stufen von Subgruppen ein. Die Daten sind online zugänglich (<http://glottolog.org/glottolog>, Zugriff: 23.09.2013).

In dieser Taxonomie kommen unter anderem nur Sprachen vor, die:

- nicht gegenseitig verständlich sind,
- als Sprache einer menschlichen Gesellschaft dienen (Pidgins gehören demzufolge nicht dazu),
- gesprochen werden (somit keine Gebärdensprachen).

Zusätzlich ist für jede Sprachfamilie vermerkt, ob es sich bei deren Mitglieder um Jäger und Sammler, sesshafte Jäger und Sammler, Hirte oder Landwirte handelt (Hammarström 2012). Bevor mit den Sprachdaten von Hammarström gearbeitet werden konnte, wurden noch einige Anpassungen vorgenommen, die im nachfolgenden Abschnitt *6.1.2 Datenaufbereitung* beschrieben werden.

6.1.2 Datenaufbereitung

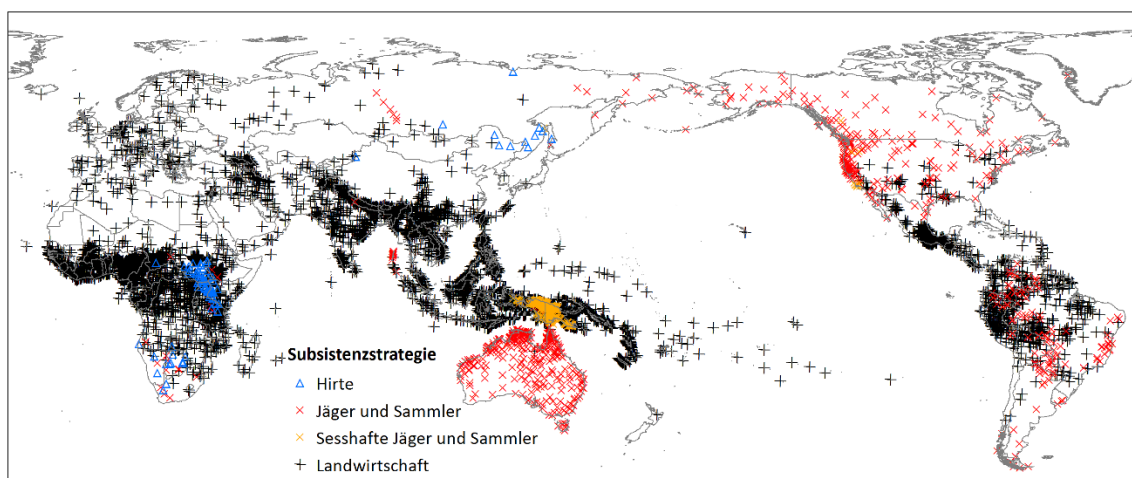
Die Daten von Hammarström liegen auf zwei Dateien verteilt vor. In der einen Datei sind die Punkt-Koordinaten und die ISO-Codes. In der anderen Datei ist die Taxonomie abgelegt. Diese enthält auf der obersten Ebene die Sprachfamilien, gefolgt von weiteren Aufteilungen in Subgruppen. Zuunterst in der Taxonomie ist wieder die Sprache mit ISO-Code aufgeführt. Über diesen Code wurden die beiden Dateien verbunden. Die Hammarström-Daten wurden so in einer Datenbank abgelegt, dass die Sprachbaumstruktur effizient wiedergegeben werden kann.

Als zusätzliches Attribut für jede Sprache wird die Subsistenzstrategie ihrer Sprecher vermerkt. Diese Information liegt jeweils pro Sprachfamilie vor. Es gibt verschiedene Typen von Subsistenzstrategien: Landwirtschaft, Hirte, Jäger und Sammler und sesshafte Jäger und Sammler. Bei drei Sprachfamilien

(Sepik, Angan, Ramanos) wird dieses Attribut nicht ergänzt, da diese Information entweder nicht vorhanden ist oder sich die Literatur uneinig ist (Hammarström 2012). In Tabelle 1 ist zu sehen, wie sich die Sprachen auf die verschiedenen Subsistenztypen verteilen. Abbildung 1 ist zudem deren weltweite Verbreitung, welche sehr unregelmässig ist, dargestellt.

Tabelle 1: Anzahl Sprachen der Hammarström-Daten pro Subsistenzstrategie.

Subsistenzstrategie	Anzahl Sprachen
Landwirtschaft	5'979
Jäger und Sammler	702
Sesshafte Jäger und Sammler	304
Hirte	77
Unbekannt / Uneinigkeit	47



Eigene Darstellung. Quellen: UNEP (2013), Hammarström (2012)

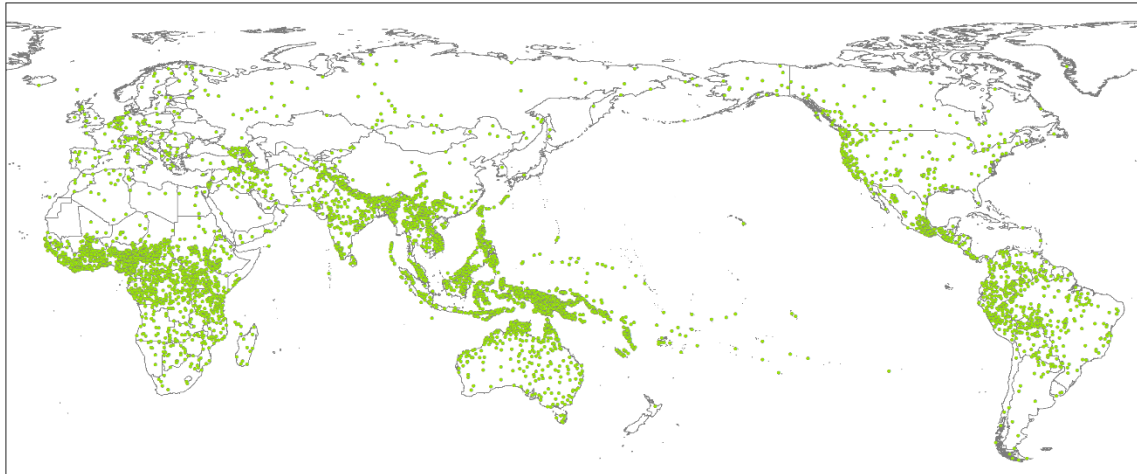
Abbildung 1: Verteilung der Sprachen eingeteilt nach Subsistenzstrategien.

6.1.2.1 Bearbeitungsschritte

Da in den Daten von Hammarström auch nach der Kolonialisierung entstandene Sprachen enthalten sind, wurden diese in der Datenbank vermerkt (insgesamt 74) und nicht mit in die Auswertungen einbezogen. In Abbildung 2 sind die Sprachen aus den Hammarström-Daten aufgrund ihrer Koordinaten abgebildet. 175 Sprachen liegen, wenn eine Verschneidung mit der Landfläche (FAO 2010) gemacht wird, im Meer. Mögliche Gründe dafür sind, dass die Koordinaten ungenau gesetzt wurden, dass sie zu grob (oft nur eine Kommastelle) sind, was bei kleinen Inseln dazu führen kann, dass die Sprachen im Meer zu liegen kommen oder, dass der Schwerpunkt der Sprache im Meer liegt. Die meisten dieser Sprachen befinden sich in Regionen wie Indonesien, auf den Philippinen oder auf Papua-Neuguinea, wo viele Inseln vorzufinden sind. Die Sprache Schwedisch ist ein Beispiel, wo das sprachgeographische Zentrum im Meer, genauer gesagt in der Ostsee, liegt. All diese Sprachen werden nicht weiterverwendet. Um sie weiterverwenden zu können, hätte man diese Daten korrigieren müssen, indem man für jede einzelne in deren Quellen nach dem Ausbreitungsgebiet gesucht hätte, um dann eine neue, auf eine Landmasse fallende Koordinate zu definieren.

Insgesamt 120 Sprachen haben die exakt gleiche Koordinate wie eine andere Sprache. Davon liegen 70 Sprachen sogar mit mindestens zwei anderen Sprachen am gleichen Ort. Dass Sprachen den genau gleichen Schwerpunkt haben, ist möglich. Da jedoch bei einer Koordinate sogar zehn Sprachen liegen, kann dort wohl davon ausgegangen werden, dass die Koordinaten nur sehr grob gesetzt wurden. Bei einigen Analysen haben diese Sprachen keinen grossen Einfluss und bei einigen können sie die Resultate

stark beeinflussen. Deshalb werden sie zum Teil weggelassen. Im *Methoden-Teil* dieser Arbeit ist dies jeweils vermerkt.



Eigene Darstellung. Quellen: UNEP (2013), Hammarström (2012)

Abbildung 2: Verteilung der verwendeten Hammarström-Sprachdaten.

6.2 AUTOTYP - DATENBANK

6.2.1 Herkunft / Aufbau

Die AUTOTYP-Datenbank (Nichols & Bickel 2009) enthält genealogische und geographische Daten zu 2'922 Sprachen und Sprachvarietäten weltweit. Die Klassifikation der Sprache, in diesem Fall Dialekte oder Varietäten, und des *Stocks*, der Sprachfamilie als höchste Einteilung, sind dabei über alle Sprachen vergleichbar. Zwischen diesen beiden Stufen sind die Sprachen zum Teil weiter gegliedert nach *major Branch*, *Subbranch*, *Sub-subbranch* und *lowest Branch* (Nichols & Bickel 2009).

Die Koordinaten für die einzelnen Sprachen sind in dieser Datenbank so gesetzt, dass sie sich im soziolinguistischen Zentrum der Sprache befinden. Beispielsweise liegt die Koordinate für Deutsch in der Nähe von Hannover. Die ganze Datenbank ist so erstellt, dass sie ungefähr die Situation im 14. Jahrhundert widerspiegelt. Also bevor sich mit der Kolonialisierung und Globalisierung einige Sprachen auf der ganzen Welt ausgebreitet haben, während andere verdrängt wurden oder ausgestorben sind. Die Datenbank enthält auch noch ein Attribut, welches Auskunft über die Subsistenzstrategie gibt. Es ist nur ergänzt, ob es sich bei den Sprechern einer Sprache um Jäger und Sammler handelt oder nicht. Diese Angabe ist jedoch nicht für alle Sprachen ergänzt, sondern nur für 712 von 2'922 Sprachen. Auch an dieser Datenbank wurden für die zugrundeliegende Arbeit noch einige Anpassungen vorgenommen, die im nachfolgenden Abschnitt 6.2.2 *Datenaufbereitung* aufgeführt sind.

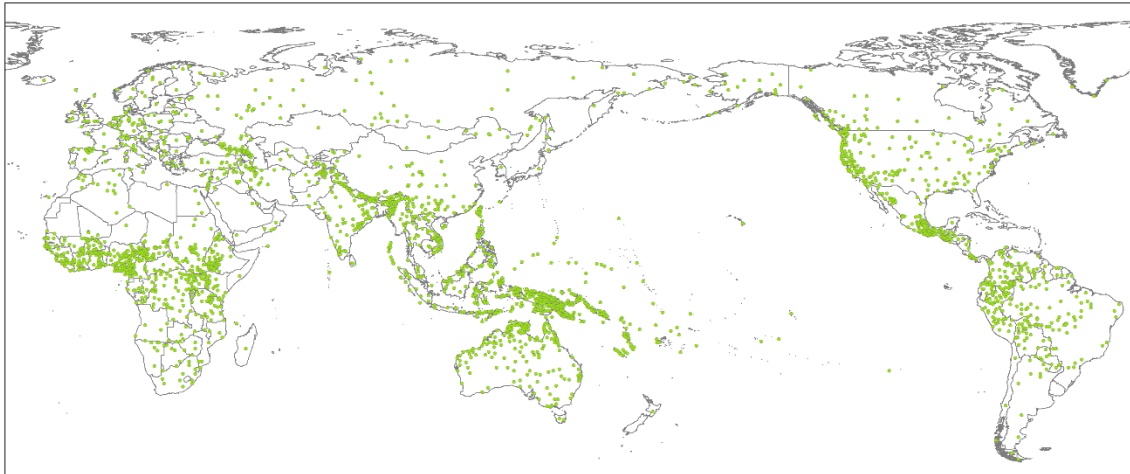
6.2.2 Datenaufbereitung

Die Daten der AUTOTYP Datenbank konnten direkt als eine Tabelle in die SQL-Datenbank importiert werden, da sie bereits entsprechend aufbereitet waren.

6.2.2.1 Bearbeitungsschritte

Auch in der Tabelle mit den AUTOTYP-Daten wurden einige Sprachen entfernt. Da diese Daten nur im dritten Teil der Arbeit verwendet wurden, bei dem die genealogische Verwandtschaft zentral ist, wurden die Gebärdensprachen (40) sowie Kreol- und Pidgin-Sprachen (32) entfernt. Diese Sprachen können nämlich nicht in einem Familienbaummodell abgebildet werden, wodurch auch keine Auswertungen diesbezüglich gemacht werden können. Zudem sind die AUTOTYP-Daten so besser vergleichbar mit den Hammarström-Daten, da dort weder Gebärdensprachen noch Kreol- oder Pidgin-Sprachen enthalten sind. Drei Sprachen wurden entfernt, da diese keine Angaben zu ihren Koordinaten hatten: Matanawí (LID 3107), Taruma (LID 3113) und Bunun (LID 3115). Auch wurden drei weitere Sprachen gelöscht: Atong (LID

3032), da diese eine fehlerhafte Koordinate hatte (907, 25.32) sowie French (Cajun) (LID 2898) und Afrikaans (LID 844), da diese erst durch die Kolonialisierung entstanden sind.



Eigene Darstellung. Quellen: UNEP (2013), Nichols & Bickel (2009)

Abbildung 3: Verteilung der verwendeten AUTOTYP-Sprachdaten.

In Abbildung 3 sind die überarbeiteten AUTOTYP-Sprachen, insgesamt noch 2'844, zu sehen. Auch hier liegen einige Sprachen (139) nicht auf einer Landfläche, sondern im Meer. Wie schon bei den Hammarström-Daten sind die meisten davon in Südostasien und Ozeanien zu finden, da es dort viele kleine Inseln gibt. Diese Sprachen werden nicht in die Auswertungen miteinbezogen.

Wie schon bei den Hammarström Daten, haben auch hier einige (48) Sprachen die exakt gleichen Koordinaten wie andere Sprachen. Einerseits ist es möglich, dass Sprachen das gleiche soziolinguistische Zentrum haben, andererseits kann es gut sein, dass diese Sprachen am gleichen Ort liegen, da sie sehr grobe Koordinaten haben. Denn die meisten der 48 Sprachen haben keine Dezimalstellen, wie zum Beispiel: N47°, E105°. Es wird jeweils in der Methode vermerkt, ob diese Sprachen dennoch verwendet wurden.

7 GRUNDLAGEN GEOGRAPHISCHE FAKTOREN

Die Daten zu den verschiedenen geographischen Faktoren stammen von unterschiedlichen Quellen. Damit ein Datensatz verwendet wurde, musste er neben der inhaltlichen Relevanz zwei weitere Kriterien erfüllen. Erstens musste er möglichst hochauflösend sein und zweitens musste nachvollzogen werden können, wie die Daten entstanden sind. Als relevant betrachtet wurden jene geographischen Faktoren, die selbst unabhängig von der Sprachendiversität sind, wie beispielsweise Niederschlag. Der Waldanteil ist ein Beispiel, bei dem nicht mit Bestimmtheit gesagt werden kann, ob er nicht auch durch die Sprachendiversität beeinflusst ist, da der Wald stark durch Menschen verändert werden kann. Solche nicht eindeutigen Faktoren wurden nicht verwendet. Dabei wurde versucht, nach Möglichkeit die untersten Faktoren von Wirkungsketten zu untersuchen. So wurden beispielsweise nicht die Tonnen Getreidepflanzen, die den Sprechern einer Sprache zu Verfügung stehen, berechnet, sondern lediglich die Nettoprimärproduktion, von der diese Pflanzen abhängen, in die Analysen miteinbezogen.

Die Faktoren lassen sich einteilen in klimatische Faktoren (*Unterkapitel 7.1*), davon abhängige Faktoren (*Unterkapitel 7.2 und 7.3*) und in Faktoren, die eine Lagebedingung beschreiben (*Unterkapitel 7.4-7.6*). Einige Daten mussten noch bearbeitet werden, bevor sie für die Analysen weiterverwendet werden konnten. Im Folgenden (*Unterkapitel 7.1 - 7.7*) sind die wichtigsten Eckdaten und die Bearbeitungsschritte zu allen geographischen Faktoren aufgeführt. Auf der beigelegten CD befinden sich entsprechende Karten, auf denen die verschiedenen Faktoren abgebildet sind.

7.1 KLIMATISCHE FAKTOREN

Die klimatischen Faktoren stammen alle von Hijmans et al. (2005) und sind alle als Raster-Daten online zugänglich (<http://worldclim.org/current>, Zugriff: 15.07.2013). Alle Daten wurden basierend auf dem monatlichen Niederschlag und der mittleren, minimalen und maximalen Temperatur der Periode 1950-2000 berechnet und haben eine Auflösung von 30 Bogensekunden (ca. 1 km² am Äquator) (Hijmans et al. 2005). In dieser Arbeit wurden vier der zur Verfügung stehenden bioklimatischen Variablen, welche durchschnittliche jährliche Werte repräsentieren, verwendet:

- Mittlere jährliche Temperatur,
- Jährlicher Temperaturschwankungsbereich,
- Jährlicher Niederschlag und
- Saisonalität des Niederschlags.

7.1.1 Mittlere jährliche Temperatur

Abbildung 4 zeigt, dass sich die meisten Sprachen an Orten befinden, wo die mittlere jährliche Temperatur zwischen 25-30 °C beträgt. Genauer gesagt, liegt über die Hälfte der Sprachen in Gebieten, wo diese über 24 °C beträgt. Mit -15.5 °C weist die Sprache Even in Sibirien die tiefste und Saho in Ostafrika mit 30.5 °C die höchste mittlere jährliche Temperatur auf.

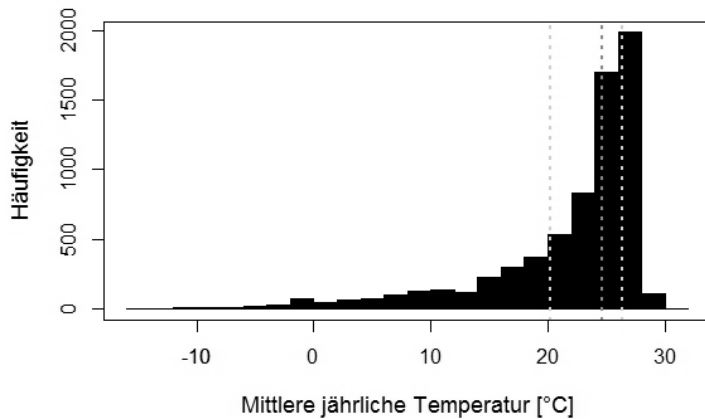


Abbildung 4: Häufigkeitsverteilung der mittleren jährlichen Temperatur jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

7.1.2 Jährlicher Temperaturschwankungsbereich

Der jährliche Temperaturschwankungsbereich wurde aus der Differenz zwischen der maximalen Temperatur des wärmsten Monats und der minimalen Temperatur des kältesten Monats berechnet (Hijmans et al. 2005). In Abbildung 5 ist die Häufigkeitsverteilung zu sehen. Über die Hälfte der Sprachen befindet sich an einem Ort, wo der jährliche Temperaturschwankungsbereich kleiner als 17 °C ist. Zum Vergleich: im Schweizer Mittelland beträgt dieser Wert ca. 25 °C. Den grössten Schwankungsbereich mit 71.7°C weist die Sprache Yakut in Ostsibirien auf und den kleinsten die Sprache Nauruan in Ozeanien mit 6.5°C.

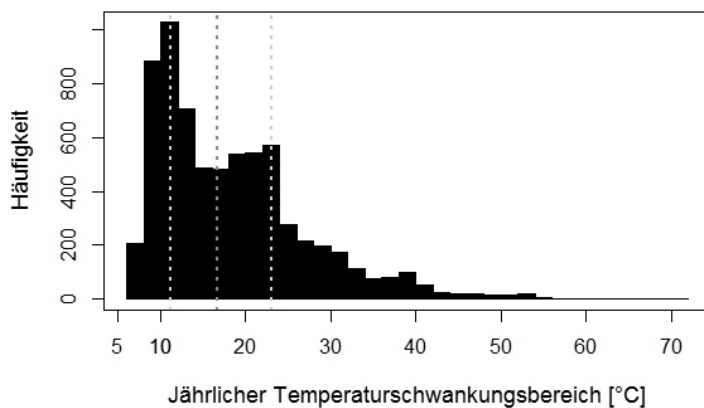


Abbildung 5: Häufigkeitsverteilung des jährlichen Temperaturschwankungsbereichs jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

7.1.3 Jährlicher Niederschlag

Der jährliche Niederschlag beschreibt die durchschnittliche Menge an Niederschlag in Millimeter (mm), die pro Jahr fällt (Hijmans et al. 2005). Vier Sprachen, die sich alle in der nordöstlichen Sahara befinden, weisen weniger als 10 mm Niederschlag auf. Den höchsten Wert hält die Sprache Epena an der Küste Kolumbiens mit 7'153 mm. Aus Abbildung 6 ist ersichtlich, dass sich mehr als die Hälfte der Sprachen in Gebieten mit mehr als 1'500 mm Niederschlag pro Jahr befinden.

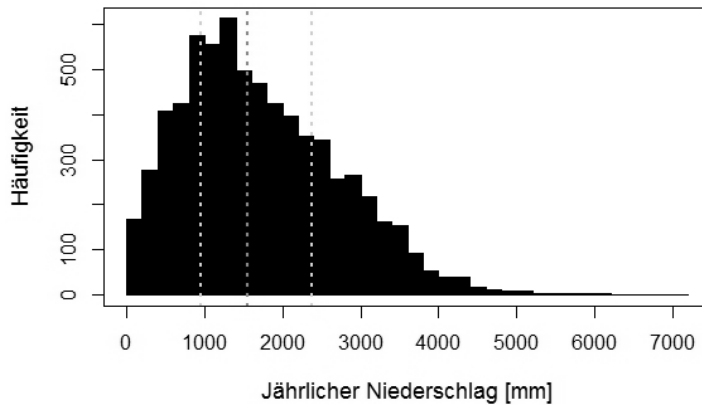


Abbildung 6: Häufigkeitsverteilung des jährlichen Niederschlags jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

7.1.4 Saisonalität des Niederschlags

Die Saisonalität des Niederschlags beschreibt den Variationskoeffizienten der monatlichen Niederschläge. Sie sagt also aus, ob die Niederschläge eher regelmässig verteilt übers Jahr fallen oder sehr verschieden sind. Die höchsten Werte erreichen Sprachen, welche in den Monsungebieten liegen.

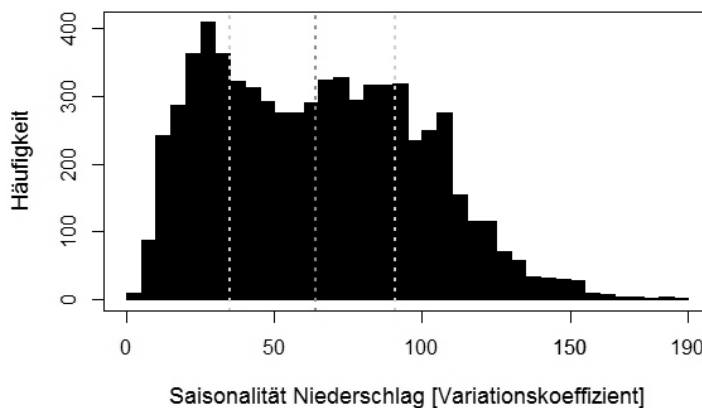


Abbildung 7: Häufigkeitsverteilung der Saisonalität des Niederschlags jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

7.2 NETTOPRIMÄRPRODUKTION

Ein weiterer Faktor, der selbst kein klimatischer Faktor ist, aber durch das Klima beeinflusst wird, ist die Nettoprimärproduktion (NPP). Pflanzen nehmen während der Photosynthese Kohlenstoffdioxid auf und geben bei der Atmung Kohlenstoffdioxid ab. Die Differenz zwischen der Aufnahme und der Abgabe beschreibt die effektive Menge an Kohlenstoffdioxid, die durch Pflanzen aufgenommen wurde. Dies wird Nettoprimärproduktion genannt. Photosynthese ist die Grundlage der Nahrungskette, da sie Solarenergie in den Pflanzen speichert, die dann von anderen Organismen konsumiert wird (NASA 2013).

Die Nettoprimärproduktion wurde aus Daten der *NASA Earth Observations* (NEO) (<http://neo.sci.gsfc.nasa.gov>, Zugriff 05.08.2013), die mithilfe von *Moderate Resolution Imaging Spectroradiometer* (MODIS) an Board von NASA-Satelliten entstanden sind, berechnet. Auf der Website von NEO werden seit dem Jahr 2000 jeweils die Durchschnittswerte der Nettoprimärproduktion von einem Monat in Gramm Kohlenstoffdioxid pro Quadratmeter pro Tag ($\text{gC}/\text{m}^2/\text{Tag}$) veröffentlicht (NASA 2013). Die Werte können auch negativ sein, wenn die Pflanze mehr abgibt, als sie aufnimmt.

Für die vorliegende Arbeit wurden alle Monate der Jahre 2000-2009 verwendet, um die durchschnittliche Aufnahme von $\text{gC}/\text{m}^2/\text{Tag}$ zu berechnen (NASA 2000-2009). Abbildung 8 verdeutlicht, dass sich über die

Hälfte aller Sprachen in Gebieten mit einem Wert über 2.2 gC/m²/Tag befinden. Zum Vergleich: im Schweizer Mittelland beträgt dieser Wert ca. 2 gC/m²/Tag. Die höchsten Werte treten in Regionen nahe des Äquators auf, wie zum Beispiel in Mexiko, Südostasien und im Kongo.

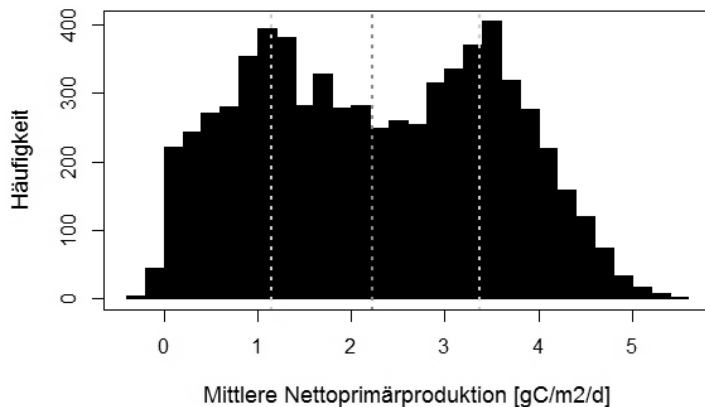


Abbildung 8: Häufigkeitsverteilung der mittleren Nettoprimärproduktion pro Tag jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

Es wurden die Daten der NASA (2000-2009) verwendet, da andere Daten zur Nettoprimärproduktion entweder ungenügend oder nicht ausführlich genau beschrieben waren. Die Daten, welche über das *NASA Socioeconomic Data and Applications Center (SEDAC)* (Imhoff et al. 2004) zugänglich sind und die auch schon von Currie & Mace (2009) verwendet wurden, sind leider in Küstennähe ungenügend. Die Werte sind dort oft viel tiefer als im unmittelbaren Inland. Da viele Sprachen in Küstennähe liegen, erhalten dann viele Sprachen einen tiefen Wert, obwohl sie in tropischen Regionen mit eigentlich hohen NPP-Werten lägen. In den Daten der NASA (2000-2009) unterscheiden sich die Werte an den Küsten kaum von denen im unmittelbaren Inland.

7.3 LÄNGE DER WACHSTUMSPHASE

Die Länge der Wachstumsphase (FAO 2000) beschreibt die Anzahl an Tagen pro Jahr, an welchen es die Temperatur und die vorhandene Feuchtigkeit erlauben, dass Getreide wachsen kann. Genauer gesagt beschreibt diese die Anzahl Tage an denen die Temperatur über 5 °C beträgt und die Feuchtigkeit dabei ausreichend ist, damit Nahrungsmittelproduktion möglich ist (FAO 2000).

Abbildung 9 veranschaulicht, wie die Länge der Wachstumsphase auf die Sprachen verteilt ist. Die meisten Sprachen befinden sich offensichtlich in Regionen, in denen fast das ganze Jahr über produziert werden kann. Für die Hälfte aller Sprachen zählt die Wachstumsphase 262 Tage im Jahr. Zum Vergleich: für die Sprache Deutsch wird ein Wert von 225 ausgegeben, für die Sprache Spanisch nur einer von 128. Dies begründet sich darin, dass es dort oft zu wenig Niederschlag hat, wodurch die Wachstumsphase weniger lang andauert.

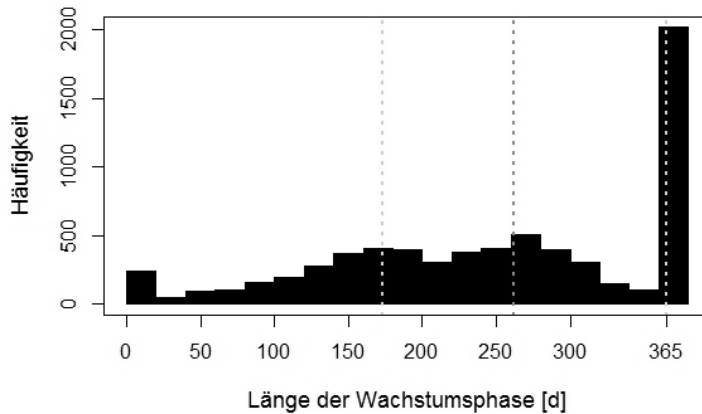


Abbildung 9: Häufigkeitsverteilung der Länge der Wachstumsphase auf die Sprachen. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

7.4 KÜSTENDISTANZ

Die Küstendistanz soll die kürzeste Distanz jeder Sprache zur nächsten Küste beschreiben. Die Küstenlinien (FAO 2010) dienen als Berechnungsgrundlage für die Küstendistanz. Für jede Sprache wurde die sphärische Distanz zur nächsten Küste berechnet. Die Küsten des Schwarzen und des Kaspischen Meeres wurden ebenfalls berücksichtigt. In Abbildung 10 ist das Histogramm zur Küstendistanz dargestellt. 25 % der Sprachen liegen näher als 30 km, 50 % näher als 161 km und 75 % näher als 560 km von einer Küste entfernt. Mit 2'255 km liegt die Sprache Tuwinisch im südlichen Sibirien am weitesten von der Küste entfernt.

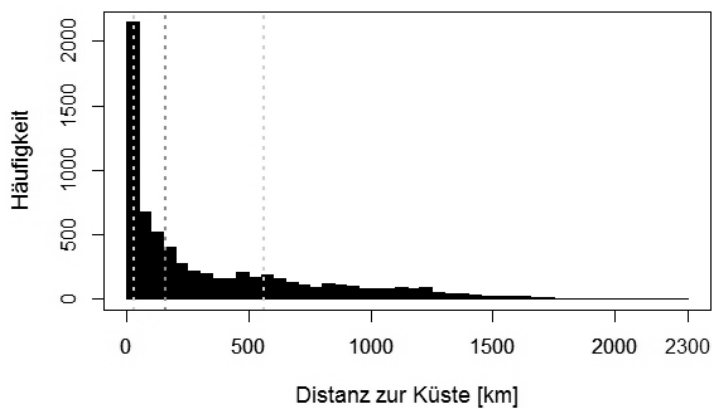


Abbildung 10: Häufigkeitsverteilung der Distanz zur Küste jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

7.5 HÖHENMODELL

Das Höhenmodell wurde von Fischer et al. (2008) aus den Daten der *NASA Shuttle Radar Topographic Mission* (SRTM) (CGIAR-CSI 2008) und bei Breitengraden über 60 aus dem *GTOPO30* (USGS 2012) zusammengestellt. Es hat eine Auflösung von 30 Bogensekunden (ca. 1 km² Rastergrösse). Abbildung 11 zeigt deutlich, dass die meisten Sprachen nur einige hundert Meter über dem Meeresspiegel liegen. Über die Hälfte der Sprachen liegt unterhalb von 370 m ü. M. Am höchsten liegt die Sprache Chantyal in Nepal auf 6'184 m.

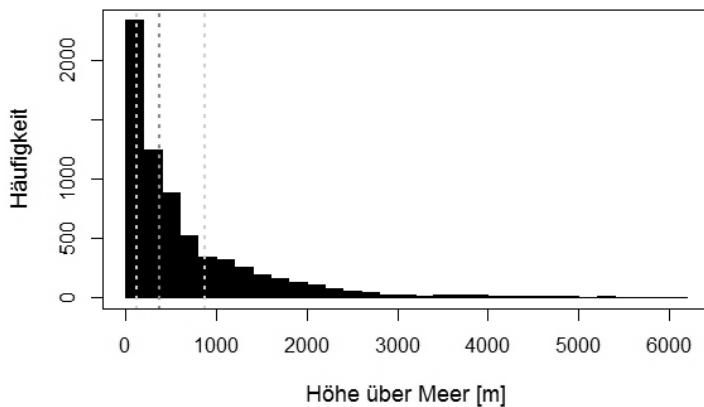


Abbildung 11: Häufigkeitsverteilung der Höhe über Meer jeder Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

7.6 HÖHENDIFFERENZ

Aus dem Höhenmodell wurde ein zusätzlicher geographischer Faktor, die Höhendifferenz, berechnet. In der physischen Geographie und Geologie ist dieses Mass auch unter dem Begriff *Reliefenergie* bekannt (Press & Siever 2003). Dabei wurde im Umkreis von 50 km um jede Sprache die Differenz zwischen der höchsten und tiefsten Stelle berechnet. Die maximale Höhendifferenz, die im Umkreis von 50 km um eine Sprache zu finden ist, beträgt 7'983 m. Die betreffende Sprache Yamphu liegt in Nepal, wie auch viele andere Sprachen mit grossen Höhendifferenzen. In Abbildung 12 ist klar erkennbar, dass mehr als 50 % der Sprachen im Umkreis von 50 km eine Höhendifferenz grösser als 1'100 m haben.

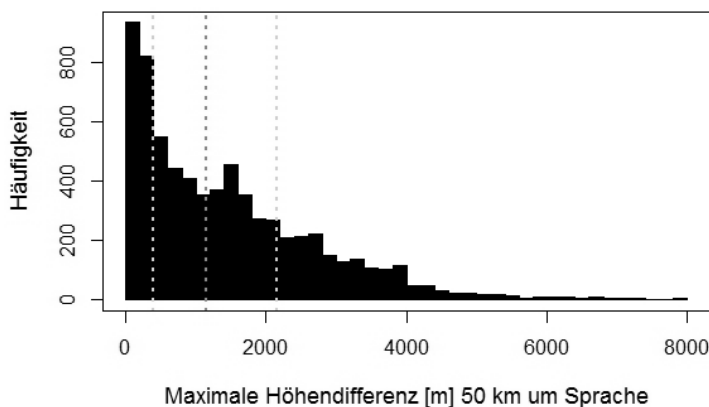


Abbildung 12: Häufigkeitsverteilung der maximalen Höhendifferenz in einem Umkreis von 50 km um eine Sprache. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

7.7 KORRELATIONEN DER GEOGRAPHISCHEN FAKTOREN UNTEREINANDER

Wie in Tabelle 2 zu sehen ist, sind die geographischen Faktoren nicht unabhängig voneinander. Viele beeinflussen sich gegenseitig oder sind teilweise von einem oder mehreren anderen Faktoren abhängig. Für die spätere Interpretation der statistischen Auswertungen ist es wichtig dies zu überprüfen. Neben den Korrelationen unter den geographischen Faktoren wurden auch die Korrelationen der Faktoren mit der geographischen Breite berechnet.

Mit der geographischen Breite (die Werte der Südhalbkugel wurden positiv gesetzt), korrelieren am stärksten die beiden Temperatur-Faktoren, gefolgt vom Niederschlag und der Länge der Wachstumsphase. Wie zu erwarten, besteht kaum eine Korrelation mit der Küstendistanz, dem Höhenmodell und der Höhendifferenz. Allgemein korrelieren keine Faktoren stark mit diesen drei nicht klimatischen Faktoren. Die klimatischen Faktoren korrelieren zum Teil mit den anderen klimatischen Faktoren und auch mit der Nettoprimärproduktion und der Wachstumsphase, da diese beide vom Klima abhängen.

Tabelle 2: Korrelationen zwischen den verschiedenen geographischen Faktoren inkl. geographischer Breite. Werte ≥ 0.5 resp. ≤ -0.5 sind fett hervorgehoben.

	Geographische Breite	Temperatur	Temperatur- schwankungsbereich	Niederschlag	Saisonalität Niederschlag	NPP	Wachstumsphase	Küstendistanz	Höhenmodell	Höhendifferenz
Geographische Breite		-0.68	0.80	-0.56	0.27	-0.46	-0.55	0.09	0.13	0.10
Temperatur	-0.68		-0.55	0.39	0.09	0.23	0.29	-0.11	-0.42	-0.37
Temperaturschwankungs- bereich	0.80	-0.55		-0.79	0.53	-0.70	-0.78	0.49	0.29	-0.05
Niederschlag	-0.56	0.39	-0.79		-0.51	0.73	0.84	-0.41	-0.22	0.17
Saisonalität Niederschlag	0.27	0.09	0.53	-0.51		-0.53	-0.64	0.34	0.17	-0.13
NPP	-0.46	0.23	-0.70	0.73	-0.53		0.80	-0.42	-0.06	0.29
Wachstumsphase	-0.55	0.29	-0.78	0.84	-0.64	0.80		-0.37	-0.18	0.18
Küstendistanz	0.09	-0.11	0.49	-0.41	0.34	-0.42	-0.37		0.42	-0.17
Höhenmodell	0.13	-0.42	0.29	-0.22	0.17	-0.06	-0.18	0.42		0.51
Höhendifferenz	0.10	-0.37	-0.05	0.17	-0.13	0.29	0.18	-0.17	0.51	

IV. Methodik

8 VORGEHEN

In dieser Arbeit werden mehrere Methoden und Datensätze verwendet, um Antworten auf die gestellten Forschungsfragen zu finden. Die Datengrundlagen für die geographischen Faktoren sind jedoch für alle die gleichen. Das Mass für die Diversität von Sprachen ändert sich, je nachdem welche Forschungsfrage beantwortet werden soll. Die dafür verwendeten Methoden können in drei Abschnitte aufgeteilt werden mit jeweils anderen Forschungsfragen und einem anderen Mass (Tabelle 3).

Tabelle 3: Aufteilung der Methoden zur Beschreibung der Sprachendiversität in drei Teile.

	Was wird untersucht?	Was wird quantifiziert?	Wie wird das Mass bezeichnet?	Welche Forschungsfrage soll beantwortet werden?
Teil 1	Sprachendiversität	Anzahl Nachbarsprachen in einem bestimmten Umkreis jeder Sprache	Lokale Sprachendichte	Bei welchen geographischen Faktoren oder Konstellationen von Faktoren kann eine hohe Sprachendiversität erwartet werden?
		Distanz zum geographisch nächsten Nachbarn	Geographisch Nächster Nachbar (GNN)	
Teil 2	Sprachfamilien-diversität	Anzahl Nachbarsprachfamilien in einem bestimmten Umkreis jeder Sprache	Lokale Sprachfamiliendichte	Bei welchen geographischen Faktoren oder Konstellationen von Faktoren kann eine hohe Sprachfamiliendiversität erwartet werden?
		Distanz zum geographisch nächsten Nachbar einer anderen Sprachfamilie	Geographisch Nächster Nachbar, nicht verwandt (GNNnv)	
Teil 3	Genealogische Diversität / Sprachen-diversifikation	Mittlere Distanz zu den nächsten verwandten Sprachen	Mittlere Verwandtschaftsdistanz	Gibt es einen Zusammenhang zwischen den geographischen Umweltbedingungen und der genealogischen Diversität resp. der Sprachendiversifikation?
		Distanz zur geographisch nächsten, nahe verwandten Sprache	Phylogenetisch Nächster Nachbar (PNN)	

Wie in Tabelle 3 ersichtlich ist, wird in jedem Abschnitt einem anderen Gesichtspunkt der Sprachendiversität nachgegangen. Im ersten Teil interessiert nur die Sprachendiversität unter bestimmten geographischen Faktoren, egal ob die involvierten Sprachen verwandt sind oder nicht. Im zweiten Teil geht es um die Diversität auf der obersten genealogischen Verwandtschaftsstufe. Der Fokus liegt hierbei auf den verschiedenen Sprachfamilien und unter welchen Umständen diese besonders zahlreich sind. Im dritten Teil interessieren die noch näher verwandten Sprachen und unter welchen Umweltbedingungen diese besonders nahe oder weit auseinander liegen. Wie die Vergleichsgrösse, also die Werte der geographischen Faktoren für jede Sprache ermittelt werden, ist in Kapitel 12 *Berechnung geographische Faktoren* erläutert.

9 SPRACHENDIVERSITÄT

9.1 LOKALE SPRACHENDICHTE

9.1.1 Berechnung

In dieser Arbeit soll der Zusammenhang zwischen Sprachendiversität und geographischen Faktoren untersucht werden. Direkt aus den Sprachdaten, die als Punkte vorliegen, ist dies jedoch nicht möglich, da diese an sich noch kein Mass für die Sprachendiversität bilden. In einem ersten Schritt wird hier deshalb eine Sprachendichte berechnet, die als Mass für die Sprachendiversität dient. Wie im *Unterkapitel 2.2 Forschungsstand* im ersten Teil der Arbeit gezeigt wurde, haben einige Autoren eine Dichte als Mass für die Sprachendiversität verwendet (Cashdan 2001, Collard & Foley 2002, Nettle 1998, Nichols 1990). Dabei wurde meist die Anzahl Sprachen in einer bestimmten Region durch die entsprechende Fläche jener Region geteilt (Collard & Foley 2002, Nettle 1998, Nichols 1990). Dabei stellt sich, wie in *Unterkapitel 2.2 Forschungsstand* beschrieben, oft das Problem der *Modifiable Areal Unit* (Openshaw 1983).

In dieser Arbeit wird versucht, dieses Problem zu minimieren. Zudem wäre es sehr schwierig auf globaler Ebene Kriterien für eine Einteilung zu finden. Auch eine *Quadrat Count Method* kommt nicht in Frage, da diese die sehr unregelmässige Verteilung der Sprachen nur ungenügend erfassen würde (O'Sullivan & Unwin 2010). Deshalb wird ein ähnlicher Ansatz gewählt, wie ihn schon Cashdan (2001) verwendet hat. Sie berechnete die Anzahl ethnischer Gruppen in einem Umkreis von 250 Meilen jeder der 186 Gesellschaften des *Standard Cross-Cultural Sample (SCCS)* (Murdock & White 2006). Das SCCS ist eine repräsentative Auswahl von Kulturen, die historisch und kulturell möglichst unabhängig voneinander sind (Murdock & White 2006). Das Problem bei Cashdans Ansatz jedoch ist, dass sie nur 186 Werte erhält, die sie dann mit geographischen Faktoren vergleichen kann. So ist ihre Auswertung zwar kulturell repräsentativ, aber zu wenig signifikant, was den Zusammenhang zur Umwelt betrifft.

In dieser Arbeit wird die Dichte ebenfalls mit Nachbarschaften berechnet, jedoch wird dabei ein Wert für jede einzelne Sprache berechnet. Es wird die Anzahl an Sprachen in einem Umkreis von 100 km um jede Sprache berechnet. Der SQL-Code zur Berechnung der Anzahl Nachbarsprachen ist im Anhang zu finden. Abbildung 13 zeigt ein Beispiel, wie die Anzahl an Nachbarn berechnet wird. Die Ellipsen-Form ergibt sich, da es sich um eine WGS84-Projektion handelt und sich die Sprachen ca. beim 20-sten Breitengrad Nord befinden.

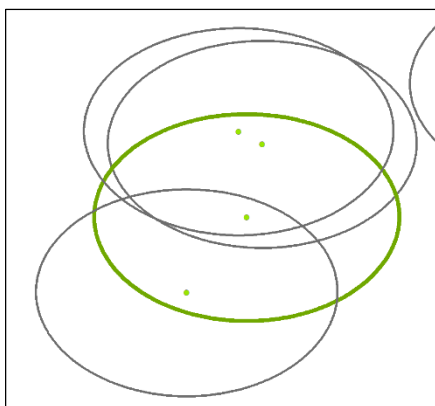


Abbildung 13: Die Sprachen als Punkte mit einem Umkreis mit Radius 100 km. Die Sprache mit dem grün eingefärbten Umkreis zählt drei Nachbarsprachen.

Der Wert der sich daraus ergibt, also die Anzahl Nachbarsprachen, ist im Prinzip eine Dichte, die sich aus der folgenden Formel ergibt:

$$\text{Sprachdichte} = \frac{\text{Anzahl Nachbarsprachen}}{\pi 100^2}$$

Da die Fläche aber überall gleich ist, wird darauf verzichtet und es wird als Mass für die Sprachendiversität die Anzahl an Nachbarsprachen verwendet. Die Anzahl wird in dieser Arbeit als *lokale Sprachendichte* bezeichnet. Als Radius wurde 100 km ausgewählt, weil dies in etwa einem Radius entspricht, in dem sich eine Person während ihres Lebens bewegt und sie darin somit potentiell Kontakt mit anderen Sprachen haben könnte. Wenn ein zu grosser Radius gewählt wird, besteht die Gefahr, dass nicht mehr untersucht wird, ob sich Sprachen aufgrund von bestimmten Umweltbedingungen ausgebildet haben,

sondern aufgrund von geographischer Isolation über die Distanz. Damit ist gemeint, dass sich die Sprecher einer Sprache voneinander isoliert haben, weil sie zu weit entfernt waren für einen intensiven Kontakt und nicht, weil ihnen beispielsweise so viel Nahrungsmittel zur Verfügung standen, dass sie auf keine sozialen Netzwerke angewiesen waren.

9.1.2 Häufigkeitsverteilung lokale Sprachendichte

In Abbildung 14 ist die Häufigkeitsverteilung der Anzahl Nachbarsprachen jeder Sprache dargestellt. Die grösste Anzahl an Nachbarn ist 132. Die Hälfte aller Sprachen hat mehr als sieben und 75 % haben mehr als 20 Nachbarsprachen. Wie die Häufigkeitsverteilung aussieht, wenn der Radius auf 400 km erhöht wird, ist in Abbildung 15 zu sehen. Abbildung 16 zeigt noch die Verteilung bei einem Radius von 800 km. In diesen beiden Abbildungen (Abbildung 15 und Abbildung 16) sind gewisse räumliche Muster erkennbar.

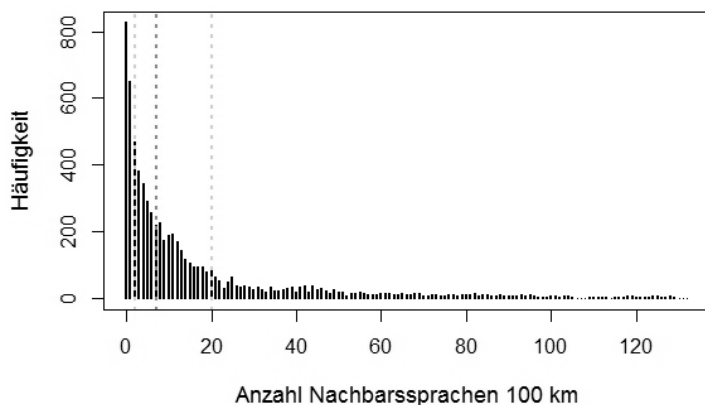


Abbildung 14: Häufigkeitsverteilung der Anzahl Nachbarsprachen im Umkreis von 100 km. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

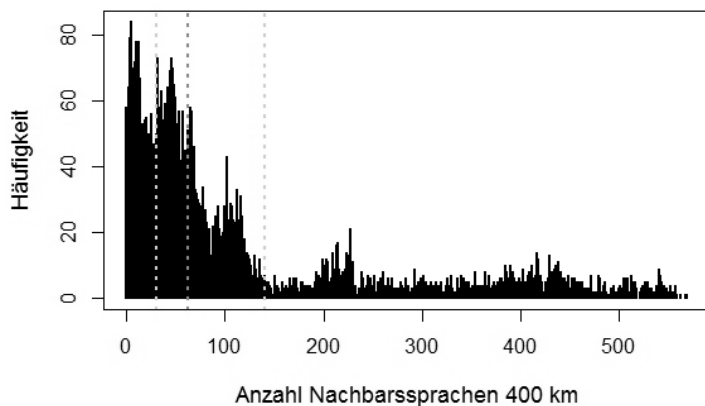


Abbildung 15: Häufigkeitsverteilung der Anzahl Nachbarsprachen im Umkreis von 400 km. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

Je grösser der Radius gewählt wird, desto mehr kristallisiert sich eine Zweiteilung heraus. Das bedeutet, dass eine Gruppe von Sprachen mit sehr hohen lokalen Dichten und ein Grossteil an Sprachen mit tieferen Dichten sich immer mehr aufteilen. Dies ist hauptsächlich auf die unregelmässige Verteilung der Sprachen zurückzuführen. In Gebieten, wo viele Sprachen vorzufinden sind, steigt die Dichte deutlich an je grösser der gewählte Radius ist. Hingegen wird dort, wo es wenige Sprachen hat, die Dichte mit zunehmendem Radius nur sehr gering grösser. In *Unterkapitel 15.1*, wo die Verteilung der lokalen Sprachendichten auf die Breitengrade dargestellt ist, wird darauf nochmals eingegangen.

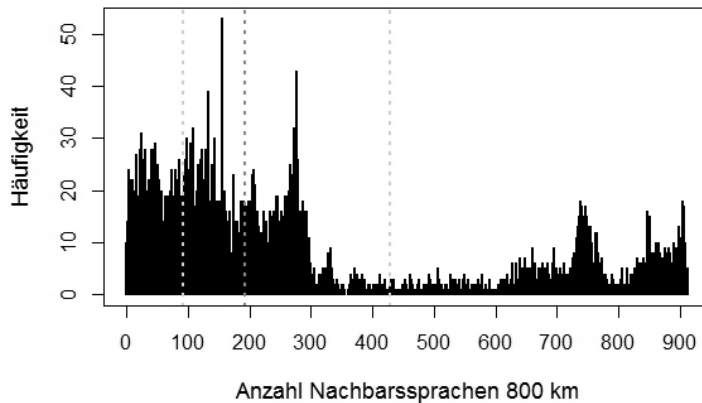


Abbildung 16: Häufigkeitsverteilung der Anzahl Nachbarsprachen im Umkreis von 800 km. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

9.2 GEOGRAPHISCH NÄCHSTER NACHBAR (GNN)

9.2.1 Berechnung

Die Distanz zum Geographisch Nächsten Nachbarn (GNN) beschreibt die kürzeste Distanz von einer Sprache zu der am nächsten gelegenen Sprache unabhängig davon, ob diese verwandt ist oder nicht. Mit dieser Distanz soll eine Aussage darüber getroffen werden, unter welchen Umweltbedingungen Sprachen näher beieinander liegen und dennoch eigenständige Sprachen bleiben oder sich auf kurze Distanzen separieren. Die GNN-Distanz wurde schon von Mace & Jordan (2011) verwendet um die Übertragung von kulturellen Merkmalen zwischen austronesischen Gesellschaften zu untersuchen.

In Abbildung 17 ist ein Beispiel zu den GNN-Distanzen abgebildet. Bei den Distanzen handelt es sich, wie bei allen Distanzen in dieser Arbeit, um sphärische Distanzen (Grosskreisdistanzen). Der entsprechende SQL-Code ist im Anhang vorzufinden. Bei 123 Sprachen beträgt die GNN-Distanz null Meter, das heisst sie haben die exakt gleiche Koordinate wie eine andere Sprache. Dieses Phänomen wurde bereits im Forschungsansatz beschrieben. Die Sprachen mit einer GNN-Distanz von 0 m werden weggelassen, da, wie schon begründet (*Abschnitt 6.1.2.1*), nicht sicher ist, dass diese tatsächlich exakt am gleichen Ort liegen.

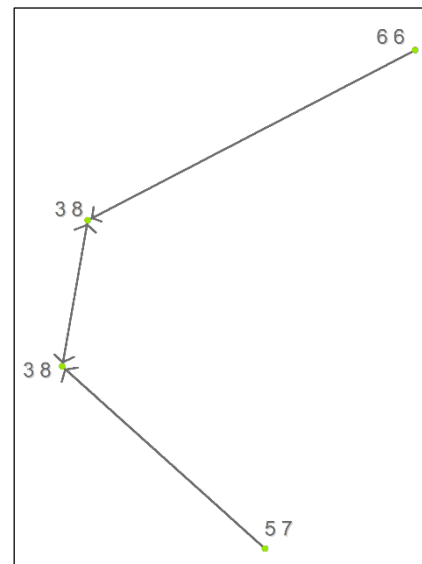


Abbildung 17: Vier Sprachen mit der GNN-Distanz [km]. Die Pfeile geben die Richtung zum nächsten Nachbarn an.

9.2.2 Häufigkeitsverteilung GNN-Distanz

Abbildung 18 zeigt auf, dass ein Viertel aller Sprachen innerhalb von 14 km eine Nachbarsprache und die Hälfte noch innerhalb von 30 km hat. Die maximale GNN-Distanz beträgt 3'174 km und betrifft die hawaiianische Sprache.

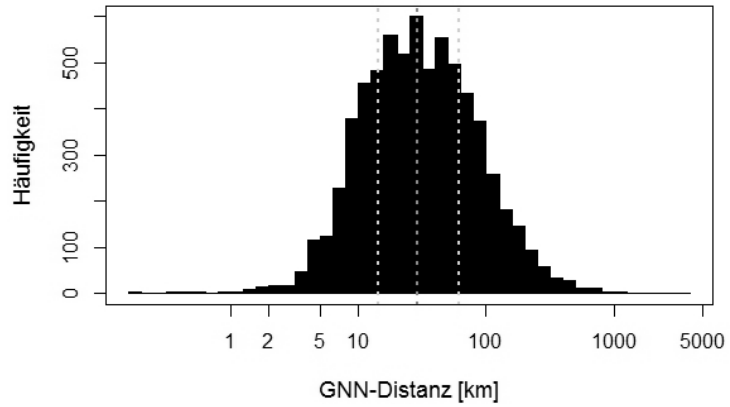


Abbildung 18: Häufigkeitsverteilung der Distanz zum geographisch nächsten Nachbarn (GNN) in logarithmierter Darstellung. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

10 SPRACHFAMILIENDIVERSITÄT

10.1 LOKALE SPRACHFAMILIENDICHTE

10.1.1 Berechnung

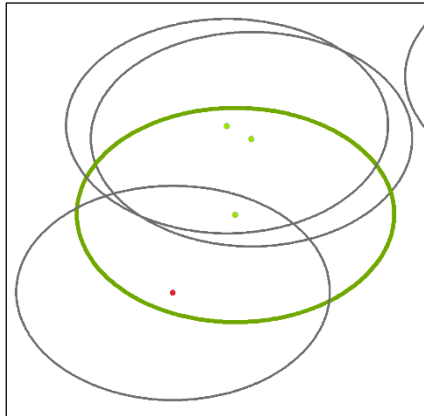


Abbildung 19: Die Sprachen als Punkte eingefärbt nach Sprachfamilie mit einem Umkreis mit Radius 400 km. Die Sprache mit dem grün eingefärbten Umkreis zählt eine Nachbarsprachfamilie.

Nachdem bereits die lokale Sprachendichte respektive die Anzahl an Nachbarsprachen in einem Umkreis von 100 km und die GNN-Distanz berechnet wurden, liegt nun der Fokus auf den Sprachfamilien, genauer gesagt auf der *Sprachfamiliendiversität*. Nichols (1990), die sich mit diesem Aspekt beschäftigte, berechnete dazu eine Dichte, die sich aus der Anzahl an Sprachfamilien pro Mio. Quadratmeilen pro Kontinent oder Subkontinent ergab. In der vorliegenden Arbeit soll ebenfalls eine Dichte berechnet werden, jedoch nicht wie bei Nichols eine einzelne Dichte über eine räumliche Einteilung, sondern wie schon bei der lokalen Sprachendichte, eine Dichte über die Nachbarschaft und für jede einzelne Sprache.

Berechnet wird wiederum die Anzahl an Nachbarn in einem bestimmten Umkreis um jede Sprache. Jedoch werden nun die verschiedenen Sprachfamilien gezählt, zu der die Nachbarsprachen gehören. Die Sprachfamilie zu der diejenige Sprache gehört, für welche die Anzahl Nachbarsprachfamilien berechnet wird, wird nicht gezählt. Das heißt, eine Sprache kann keine Nachbarsprachfamilie haben, obwohl sie Nachbarsprachen im Umkreis zählt (aber eben nur von ihrer eigenen Sprachfamilie). In Abbildung 19 ist zu erkennen, dass die Sprache mit dem grünen Umkreis nur eine Nachbarsprachfamilie (rot) hat. Die beiden Sprachen nördlich des mittleren Punktes hingegen haben keine Nachbarsprachfamilie. Der SQL-Code zu diesen Berechnungen ist ebenfalls im Anhang zu finden. Die Anzahl an Nachbarsprachfamilien wird in dieser Arbeit *lokale Sprachfamiliendichte* genannt, da sie im Grunde genommen auch eine Dichte beschreibt:

$$\text{Sprachfamiliendichte} = \frac{\text{Anzahl Nachbarsprachfamilien}}{\pi 400^2}$$

10.1.2 Häufigkeitsverteilung lokale Sprachfamiliendichte

Die Wahl des Radius gestaltet sich schwieriger als bei der lokalen Sprachendichte. Wie in Abbildung 20 zu sehen ist, zählen bei einem Radius von 100 km über 3'500 Sprachen keine andere Sprachfamilie. Bemerkenswert ist dabei, dass es eine Sprache gibt mit 23 Nachbarsprachfamilien. Bei einem Radius von 400 km (Abbildung 21) haben immer noch 1'304 Sprachen eine lokale Sprachfamiliendichte von null, über 900 eine von eins und über 1'000 eine von zwei.

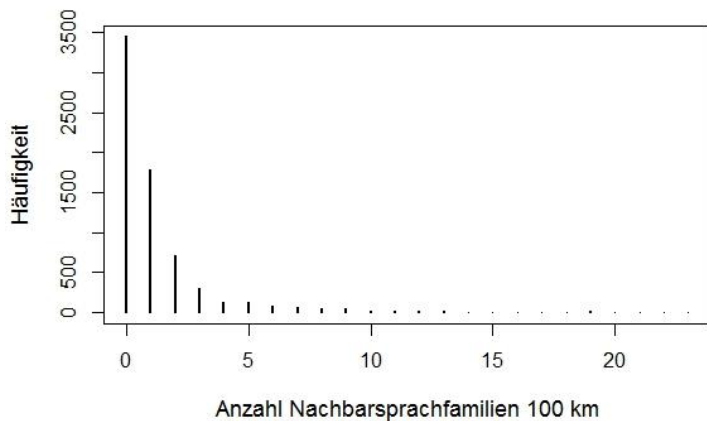


Abbildung 20: Häufigkeitsverteilung der Anzahl Nachbarsprachfamilien im Umkreis von 100 km.

Abbildung 21 verdeutlicht, dass die Häufigkeit ungefähr bis 20 Nachbarsprachfamilien stetig abnimmt. Danach variiert sie jedoch auf einem tiefen Niveau auf und ab. Bei Verwendung eines Radius von 800 km (Abbildung 22) zeigt sich bei den hohen Dichten nochmals ein stärkerer Effekt, denn dort nimmt die Häufigkeit wieder zu. So zählen beispielsweise 113 Sprachen 95 Nachbarsprachfamilien, in etwa gleich viel wie bei 27 und 10 Nachbarsprachfamilien. Dies lässt sich, wie schon bei den Nachbarsprachen, mit der Clusterbildung von Sprachfamilien erklären. Eine Ausführung dazu ist in *Unterkapitel 16.1* vorzufinden.

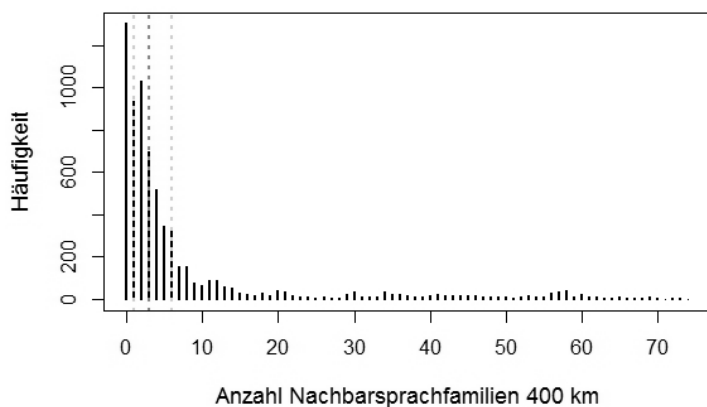


Abbildung 21: Häufigkeitsverteilung der Anzahl Nachbarsprachfamilien im Umkreis von 400 km. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

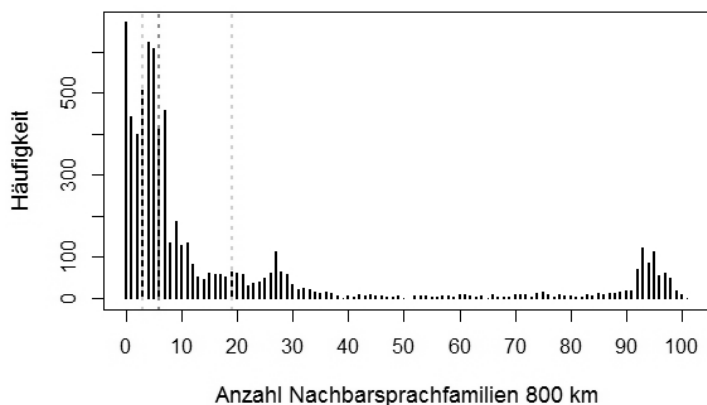


Abbildung 22: Häufigkeitsverteilung der Anzahl Nachbarsprachfamilien im Umkreis von 800 km. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

Der Radius des Umkreises, mit dem die Auswertungen gemacht werden sollen und welcher schlussendlich die Sprachfamiliendiversität bestimmt, wurde noch nicht festgelegt. Bei den Sprachfamilien interessiert

ein anderer Aspekt als bei den einzelnen Sprachen, weshalb der 100 km Radius nicht einfach übernommen werden konnte. Bei der Sprachendiversität geht es einerseits darum, unter welchen Umweltbedingungen die Sprachendiversifikation hoch ist, das heisst, ob viele Sprachen einer Sprachfamilie vorhanden sind. Andererseits interessiert, ob allgemein viele Sprachen vorhanden sind, sowohl verwandt als auch nicht verwandt, da die Umweltbedingungen offenbar dazu führen, dass sich keine Sprache durchsetzt. Bei den Sprachfamilien hingegen interessiert vor allem, unter welchen Umweltbedingungen mehrere Sprachfamilien nahe beieinander existieren respektive welche Umweltbedingungen dazu führen, dass sich Sprachen, obwohl geographisch relativ nahe beieinander, sich über die Zeit so stark voneinander entfernten, dass eine genealogische Verwandtschaft nicht mehr nachweisbar ist. Aus diesem Grund wurde entschieden, dass zur Berechnung der lokalen Sprachfamiliendichte ein grösserer Radius gewählt werden muss, da Sprachfamilien einen längeren Zeitraum der Ausprägung repräsentieren (siehe *Kapitel 3*). Ein Radius von 400 km wird dem gerecht und ist auch nicht zu gross, dass sich darin schon zu feste Cluster widerspiegeln.

Es gibt Sprachfamilien die sich über weite Gebiete ausgebreitet und dabei andere Sprachen verdrängt haben. Dies passierte oftmals aufgrund von kulturellen und historischen Prozessen. So haben sich beispielsweise viele dieser Sprachen mit der Erfindung der Landwirtschaft ausgebreitet (Diamond 1997, Diamond & Bellwood 2003). Die Verteilung der Sprachfamilien in solchen Gebieten kann somit kaum durch geographische Faktoren erklärt werden. Für die Analysen wird deshalb separat noch eine Analyse ohne die Sprachen der grossen Sprachfamilien gemacht. Eine grosse Sprachfamilie wird definiert als eine Familie mit mehr als 200 Sprachen. Insgesamt betrifft dies sieben Sprachfamilien mit 4'377 Sprachen.

10.2 GEOGRAPHISCH NÄCHSTER NACHBAR, NICHT-VERWANDT (GNNnv)

10.2.1 Berechnung

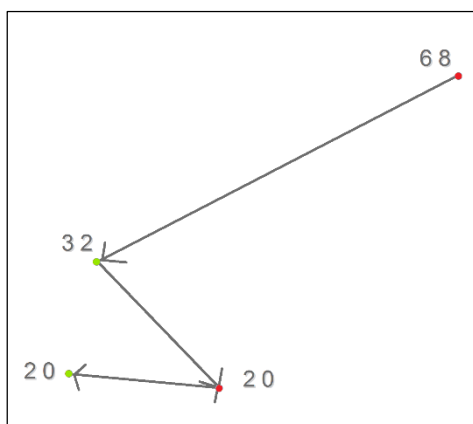


Abbildung 23: Vier Sprachen von zwei verschiedenen Sprachfamilien (grün, rot) und ihre GNNnv-Distanz.

Die GNNnv-Distanz beschreibt die kürzeste Distanz von einer Sprache zur nächsten Sprache, die nicht genealogisch verwandt ist, das heisst, sich nicht in derselben Sprachfamilie befindet. Abbildung 23 zeigt diesbezüglich ein einfaches Beispiel.

Bei der Berechnung resultiert für eine Sprache eine GNNnv-Distanz von null Metern. Diese wird für die Analyse, aus denselben Gründen wie schon bei der GNN-Distanz (siehe *Unterkapitel 9.2*) weggelassen.

10.2.2 Häufigkeitsverteilung der GNNnv-Distanz

Abbildung 24 verdeutlicht, dass die Hälfte der Sprachen innerhalb von 100 km eine nicht-verwandte Nachbarsprache hat. Bei einem Viertel der Sprachen liegt dieser Wert über 300 km. Die längste GNNnv-Distanz beträgt 6'021 km und betrifft die Sprache Rapa auf den Austral-Inseln im südlichen Pazifischen Ozean. Wie schon bei der lokalen Sprachfamiliendichte werden mit den GNNnv-Distanzen zusätzlich Analysen und Auswertungen ohne die Sprachen der grossen Sprachfamilien (mehr als 200 Mitgliedssprachen) gemacht. Die Gründe dafür sind dieselben wie in Abschnitt 10.1.2.

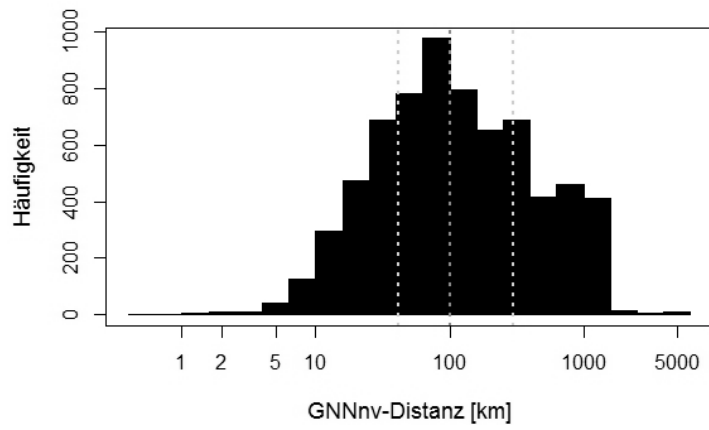


Abbildung 24: Häufigkeitsverteilung der Distanz zum geographisch nächsten, nicht-verwandten Nachbar (PNN) in logarithmierter Darstellung. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

11 GENEALOGISCHE DIVERSITÄT

In diesem Kapitel geht es darum, ein Mass für die genealogische Diversität zu berechnen und eine Antwort auf die Frage zu finden, ob die Sprachendiversifikation bei bestimmten Umweltbedingungen höher oder tiefer ist. Dazu muss auf die Taxonomie der Sprachdaten zurückgegriffen werden. Mit ihr lässt sich sagen, ob Sprachen miteinander verwandt sind, ob es Sprachen gibt die noch näher verwandt sind oder, ob dies die am nächsten verwandten Sprachen sind.

In der Literatur wurden verschiedene Ansätze auf unterschiedlichen Verwandtschaftsebenen gewählt, um ein Mass für die genealogische Diversität zu erhalten. Ein Ansatz besteht darin, die geographische Distanz und die strukturellen Unterschiede zwischen allen Sprachen zu berechnen (Holman, et al., 2007) und anschliessend die Resultate nach verwandten und nicht-verwandten Sprachen separat darzustellen. Dabei zeigte sich, dass der strukturelle Unterschied bei beiden gleichermassen mit der Distanz zunimmt, bei den nicht-verwandten einfach auf einem höheren Niveau. Weiter wurde das Ganze auch separat hinsichtlich der Aspekte Inland/Küste und tropisch/nicht-tropisch berechnet. Dieser Ansatz kann hier nicht verwendet werden. Zum einen, weil hier keine strukturellen Merkmale untersucht werden und zum anderen, da er einzig zwischen verwandt und nicht verwandt unterscheidet. In dieser Arbeit wird jedoch der Grad der Verwandtschaft miteinbezogen.

Es wurden mehrere Methoden gewählt um ein Mass für die genealogische Diversität und die Sprachendiversifikation zu ermitteln. Diese untersuchen jeweils unterschiedliche Aspekte. Zudem wurden die Analysen mit den Hammarström-Daten und den AUTOTYP-Daten gemacht, um zu prüfen, ob auch mit einer geringfügig sich unterscheidenden Taxonomie die gleichen Resultate erreicht werden.

11.1 MITTLERE VERWANDTSCHAFTSDISTANZ

11.1.1 Berechnung

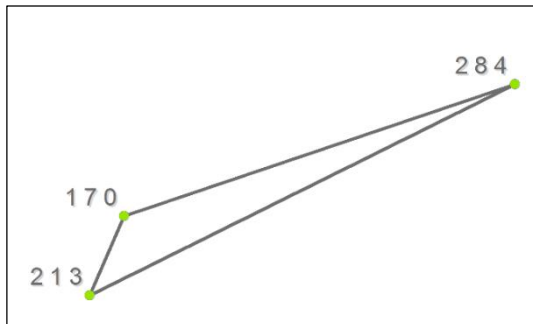


Abbildung 25: Drei verwandte Sprachen (grün) mit Angaben zur mittleren Verwandtschaftsdistanz [km].

Die mittlere Verwandtschaftsdistanz berechnet sich für jede Sprache, indem die durchschnittliche Distanz zu den nächsten Verwandten berechnet wird. Dies sind jene Sprachen, die sich in derselben untersten Gruppe einer Sprachfamilie befinden. Abbildung 25 veranschaulicht ein Beispiel einer solchen Gruppe.

Sprachen, deren nächste Verwandte keine Koordinaten haben, das heisst, die keinen Eintrag in den Hammarström-Daten besitzen, dafür in der dazugehörigen Taxonomie, fallen somit aus der Berechnung heraus, weil keine Distanz berechnet werden kann.

Insgesamt bleiben sodann noch 6'006 Sprachen übrig. Bei einigen Sprachen (29) beträgt diese mittlere Verwandtschaftsdistanz null Meter. Dies ist der Fall, wenn alle nächsten Verwandten die gleichen Koordinaten haben. Diese 29 Sprachen werden für die Analyse nicht verwendet, da es fraglich ist, ob diese Sprachen und all ihre nächsten Verwandten wirklich am exakt gleichen Ort liegen (siehe Abschnitt 6.1.2.1).

Auf der beigelegten CD ist eine Karte mit dem kompletten Linien-Netz der untersten Verwandtschaftsstufe zu finden. Ebenfalls dort und im Anhang ist der SQL-Code zur Berechnung der Distanzen enthalten. Dieser ist so aufgebaut, dass von jeder Sprache aus zu jeder anderen Sprache, die sich in derselben untersten Gruppe einer Sprachfamilie befindet, eine Linie erstellt wird. Danach wird pro Sprache die durchschnittliche Länge, der von ihr abgehenden Linien, berechnet.

Leider konnte für die AUTOTYP-Datenbank die mittlere Verwandtschaftsdistanz nicht berechnet werden. Zu über der Hälfte der Sprachen gibt es entweder nur Angaben zur obersten Einteilung (753), dem *stock*, oder nur eine Einteilung auf der zweitobersten Stufe (747), dem *mbranch*. Würde da die Verwandtschaftsdistanz berechnet werden, gäbe es extrem grosse Werte, da sich das Verwandtschaftsnetz über sehr weite Strecken zieht (siehe beigelegte CD). Diese mittlere Verwandtschaftsdistanz würde dann nicht mehr die möglichen Einflüsse der geographischen Faktoren auf die Sprachendiversifikation widerspiegeln, da die Sprachen in Gebieten mit sehr unterschiedlichen Umweltbedingungen lägen.

11.1.2 Häufigkeitsverteilung der mittleren Verwandtschaftsdistanz

Abbildung 26 zeigt die Häufigkeitsverteilung der mittleren Verwandtschaftsdistanz der Hammarström-Daten. Bei der Hälfte der Sprachen beträgt die mittlere Verwandtschaftsdistanz mehr als 85 km. Ein Viertel der Sprachen liegt nicht weiter als 40 km von seinen nächsten Verwandten entfernt. Die Sprache Salar, welche in China und ihre nächsten Verwandten im Iran gesprochen werden, weist mit 4'791 km die grösste Distanz auf.

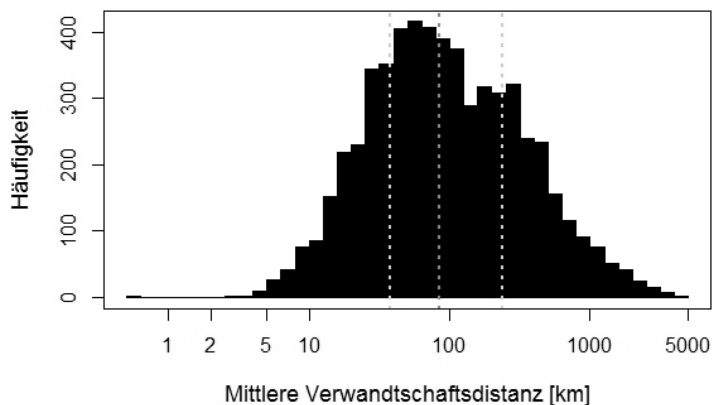


Abbildung 26: Häufigkeitsverteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

11.2 PHYLOGENETISCH NÄCHSTER NACHBAR (PNN)

11.2.1 Berechnung

Die Distanz zum phylogenetisch nächsten Nachbarn (PNN) wird mit denselben Verbindungslinien wie bei der mittleren Verwandtschaftsdistanz berechnet, wobei diesmal nur die kürzeste Distanz wichtig ist. Auch Mace & Jordan (2011) interessierten sich nebst der GNN-Distanz für den PNN. Jedoch war bei ihnen damit nicht eine Distanz gemeint, sondern die phylogenetisch nächste Gesellschaft, unabhängig von der Distanz.

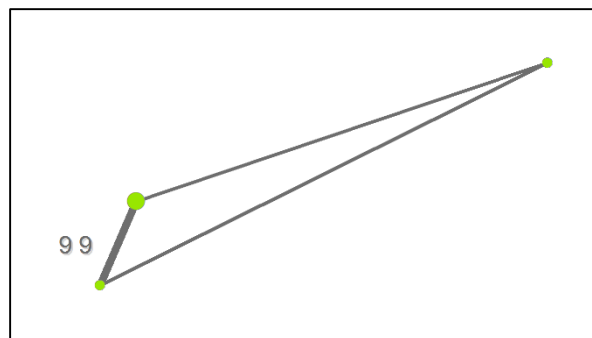


Abbildung 27: Drei verwandte Sprachen (grün). Hervorgehobene Sprache mit Angabe zur PNN Distanz [km].

Die PNN-Distanz macht eine gering andere Aussage als die mittlere Verwandtschaftsdistanz. Sie beschreibt keinen Durchschnitt, sondern die kürzeste Distanz, auf der sich zwei auf derselben untersten Taxonomiestufe befindliche Sprachen unterscheiden. Berücksichtigt man nur die PNN-Distanz, widerspiegelt dies vermutlich in differenzierterer Weise die Einflüsse von Umweltfaktoren auf die Sprachbildung. Beispielsweise beträgt in Abbildung 27 die PNN-Distanz für die hervorgehobene Sprache 99 km. Die mittlere Verwandtschaftsdistanz für dieselbe Sprache

ist 170 km. Möglicherweise verstärken die dortigen Umweltbedingungen die Separation von Sprachen, weshalb sich dort auf kürzerer Distanz zwei Sprachen entwickelt haben. Bei einigen Sprachen (89) beträgt die PNN-Distanz null Meter. Da auch hier, wie schon bei der mittleren Verwandtschaftsdistanz, nicht überprüft werden kann, ob das wirklich der Realität entspricht oder die Koordinaten zu grob gesetzt wurden, werden diese nicht in die Auswertungen miteinbezogen.

11.2.2 Häufigkeitsverteilung PNN-Distanz

In Abbildung 28 ist erkennbar, wie sich die PNN-Distanzen der einzelnen Sprachen verteilen. Bei über der Hälfte der Sprachen liegt die nächste verwandte Sprache näher als 56 km. Für die PNN-Distanzen der AUTOTYP-Datenbank ist die Häufigkeitsverteilung in Abbildung 29 zu finden. Die Distanzen sind dort deutlich höher. So liegt die nächste verwandte Sprache bei der Hälfte der Sprachen weiter als 91 km entfernt.

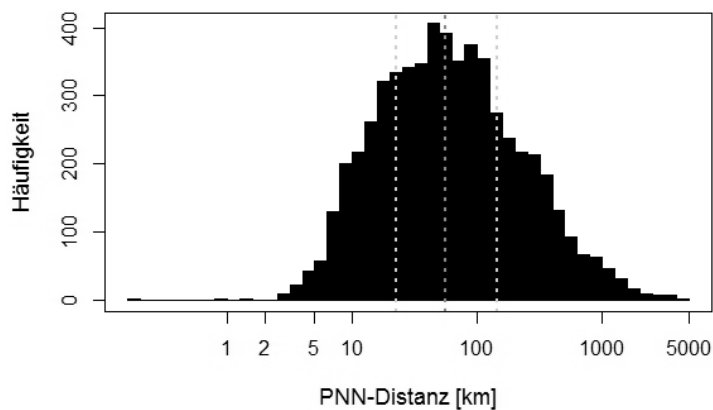


Abbildung 28: Häufigkeitsverteilung der Distanz zum phylogenetisch nächsten Nachbar (PNN) in logarithmierter Darstellung. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

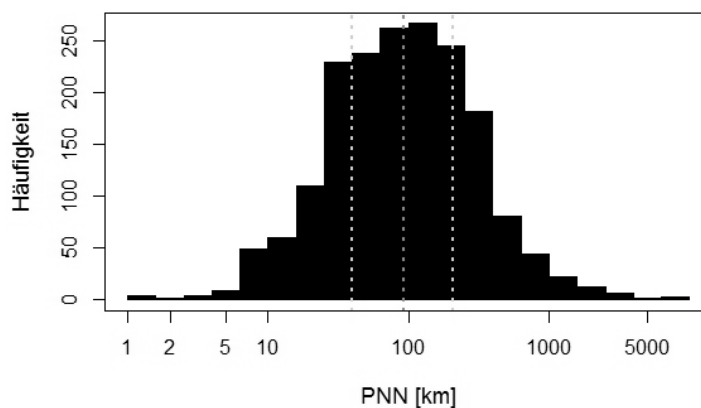


Abbildung 29: Häufigkeitsverteilung der Distanz zum phylogenetisch nächsten Nachbar (PNN) der AUTOTYP-Datenbank in logarithmierter Darstellung. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

Mithilfe der PNN-Distanz ist es möglich, die mittlere nächste-Nachbar Distanz (*mean nearest-neighbor distance*) (Clark & Evans 1954, O'Sullivan & Unwin 2010) für jede Sprachfamilie zu berechnen. In Abbildung 30 zeigt sich, dass diese für die Hälfte der Sprachfamilien mehr als 78 km beträgt. Das heisst, dass bei der Hälfte der Sprachfamilien der PNN durchschnittlich nicht weiter als dieser Wert entfernt ist.

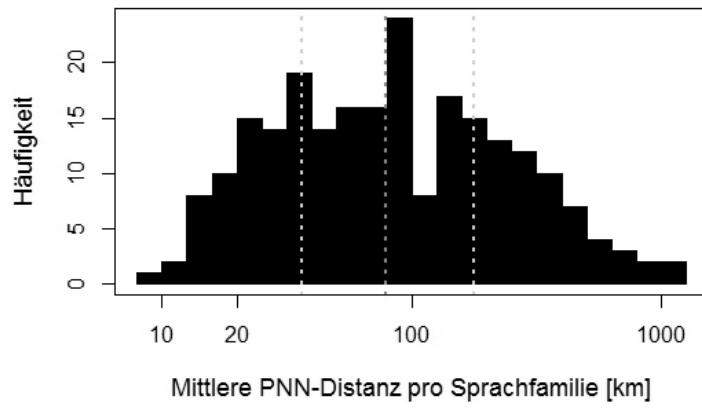


Abbildung 30: Verteilung der mittleren nächste-Nachbar Distanz der einzelnen Sprachfamilien in logarithmierter Darstellung. Senkrechte Linien: 25 %, 50 % und 75 % Quantil.

12 BERECHNUNG GEOGRAPHISCHE FAKTOREN

In den drei vorangegangenen Kapiteln wurde beschrieben wie die abhängige Variable, die Sprachendiversität, berechnet wird. In diesem Kapitel geht es nun darum zu erläutern wie die unabhängigen Variablen, die geographischen Faktoren, für jede einzelne Sprache bestimmt werden.

Wie aus der Beschreibung der Datengrundlagen (siehe *Kapitel 7*) ersichtlich ist, liegen die geographischen Faktoren entweder als Rasterdatei oder Shapefile vor. Diese werden mit den Sprachpunkten verschnitten, so dass jeder Sprache ein Wert für jeden geographischen Faktor zugewiesen werden kann. Ausnahmen bilden hier die Küstendistanz und die Höhendifferenz, die wie in Unterkapitel 7.4 und 7.6 beschrieben, für jede Sprache berechnet werden mussten.

Eine Sprache wird eigentlich nicht nur an einem Punkt — dem soziolinguistischen Zentrum, welches in der Datenbank vermerkt ist — gesprochen, sondern eine Sprache respektive ihre Sprecher sind über ein Gebiet verbreitet. Demzufolge müsste richtigerweise der Durchschnitt über dieses Gebiet ermittelt werden. Das Problem dabei ist jedoch, dass in den verwendeten Daten keine Information zum Ausbreitungsgebiet vorhanden ist und es auch häufig schwierig bis unmöglich ist, ein solches Sprachgebiet abzugrenzen. Entsprechend könnte auch einfach der Durchschnitt eines Umkreises mit einer bestimmten Größe gewählt werden. Das wiederum ist jedoch auch problematisch, da es Regionen gibt in denen in jedem Dorf eine andere Sprache gesprochen wird, wo somit ein Radius von 10 km angebracht wäre, und Regionen in denen sich die Sprecher einer Sprache über mehrere hundert Kilometer verteilen. Um zu prüfen, ob es einen signifikanten Unterschied gibt, ob ein Durchschnittswert oder der Wert am Ort des Sprachpunktes genommen wird, wurden für einen Umkreis mit dem Radius 50 km die Durchschnittswerte berechnet und mit den Werten der Punkt-Verschneidung verglichen. Abbildung 31 zeigt einen Vergleich zwischen dem jährlichen Niederschlag am Ort der Sprachkoordinate und dem Durchschnitt von 50 km um jede Sprache.

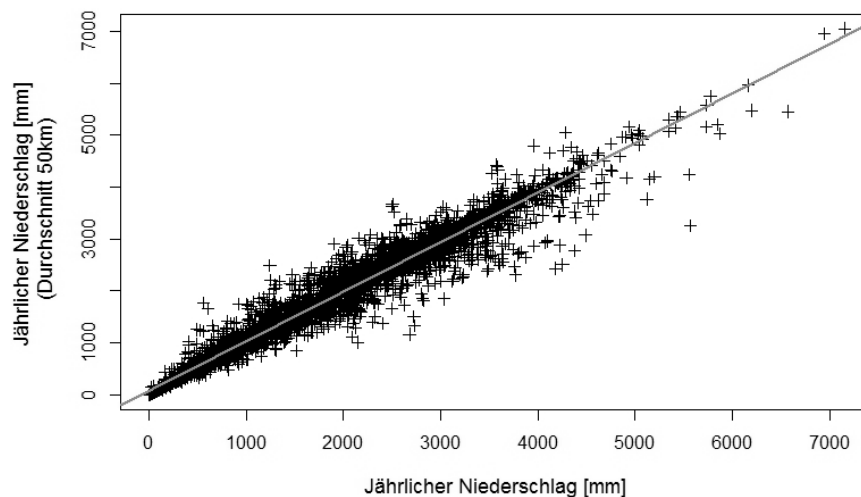


Abbildung 31: Die x-Achse zeigt den jährlichen Niederschlag am Ort der Sprachkoordinate und die y-Achse den durchschnittlichen Niederschlag 50 km rund um jede Sprache.

In der nachfolgenden Tabelle 4 sind die Korrelationskoeffizienten des Zusammenhangs zwischen den Werten der geographischen Faktoren am Ort der Sprache vs. den durchschnittlichen Werten der Faktoren in einem Umkreis von 50 km aufgeführt. Wie zu erkennen ist, sind die Korrelationskoeffizienten alle sehr nahe bei eins. Die zwei auf verschiedene Weise errechneten Werte der geographischen Faktoren unterscheiden sich also nur geringfügig. Deshalb wurde entschieden, anstatt irgendeinen Durchschnitt zu nehmen, der letztendlich auch nur wieder einen Kompromiss darstellen würde, einfach die Werte am Ort der Sprache zu verwenden. Auch wenn dabei für Sprachen, die über weite Gebiete verteilt sind, eher

zufällige Werte resultieren können. Da aber die meisten Sprachen eine kleine und nur sehr wenige eine grosse Gebietsfläche aufweisen (Currie & Mace 2009), sollte auf globalem Massstab (um den es in dieser Arbeit ja geht) dieser Fehler vernachlässigbar sein.

Tabelle 4: Korrelationskoeffizienten des Zusammenhangs zwischen den Werten der geographischen Faktoren am Ort der Sprachen vs. des Durchschnitts in 50 km Umkreis. Alle Werte sind hoch signifikant.

Geographischer Faktor	Korrelationskoeffizient
Mittlere jährliche Temperatur	0.97
Jährlicher Temperaturschwankungsbereich	0.998
Jährlicher Niederschlag	0.98
Saisonalität Niederschlag	0.99
Nettoprimärproduktion	0.96
Länge der Wachstumsphase	0.99
Höhenmodell	0.92

13 SOFTWARE

In dieser Arbeit wurde für die Auswertung, Analysen und Visualisierungen unterschiedliche Software verwendet. Die Sprachdaten und daraus resultierenden Ergebnisse wurden in einer Datenbank auf einem Microsoft SQL Server 2012 gehalten. Alle Berechnungen, wie beispielsweise die lokale Sprachendichte oder die PNN-Distanz, wurden, wenn nicht anders vermerkt, ebenfalls in dieser Datenbank mithilfe des *Microsoft SQL Server Management Studio* (Version 11.0.3128.0) ausgeführt. Die Verschneidungen der Sprachdaten mit den geographischen Faktoren wurden mithilfe des *Geospatial Modelling Environment* (Version 0.7.2.1) (Beyer 2009-2012) durchgeführt. Die Visualisierung der Daten wurde mit *Esri ArcMap 10.1* (Lizenztyp: Advanced) vorgenommen. Die statistischen Auswertungen und die Erstellung von Diagrammen wurden mithilfe des Statistikprogramms *R* (R Core Team 2013) durchgeführt.

14 ÜBERBLICK

14.1 AUFBAU

In diesem Teil der Arbeit werden die Resultate der verschiedenen Methoden präsentiert. Für alle drei Themenbereiche (Sprachendiversität, Sprachfamiliendiversität und genealogische Diversität) werden zuerst jeweils die Analysen der einzelnen geographischen Faktoren dargestellt. Danach wird für jedes Mass, wie zum Beispiel die lokale Sprachendichte, eine multiple Regressionsanalyse mit allen Faktoren durchgeführt und ein gemischtes Modell mit dem zufälligen Effekt der Sprachfamilie erstellt. Am Schluss jedes Themenbereiches folgt eine Übersicht über die Resultate und deren Interpretation. Zuerst folgt im anschliessenden Unterkapitel noch eine Zusammenfassung der verwendeten statistischen Analysen.

14.2 STATISTISCHE GRUNDLAGEN

Zwischen jedem Sprachenmass (abhängige Variable) und jedem geographischen Faktor (unabhängige Variable) wird eine Korrelation berechnet. Dabei wird jeweils die Methode von Spearman verwendet, aus welcher der Rangkorrelationskoeffizient ρ resultiert. Dieser bewegt sich wie der Korrelationskoeffizient r von Pearson zwischen -1 und 1, wobei bei einem ρ von 0 kein und je näher bei 1 respektive -1 ein starker Zusammenhang zwischen den Variablen besteht. Der Rangkorrelationskoeffizient hat gegenüber dem Korrelationskoeffizient von Pearson den Vorteil, dass er für den Signifikanztest keine Normalverteilung beider Variablen benötigt und auch keinen linearen Zusammenhang voraussetzt. Einige geographische Faktoren erfüllen das Kriterium der Normalverteilung nicht. Damit die Korrelationskoeffizienten miteinander verglichen werden können, wurde sowohl für die normal- wie auch für die nicht-normalverteilten Faktoren der Rangkorrelationskoeffizient von Spearman berechnet. Die beiden Korrelationskoeffizienten der normalverteilten Faktoren unterscheiden sich meist nur um zwei Punkte im Hundertstelbereich, wie bei einem Vergleich festgestellt werden konnte. Der Signifikanztest überprüft, ob sich ρ signifikant von Null unterscheidet (Dolić 2004). Bei den meisten in den folgenden Kapiteln berechneten Korrelationen ergibt sich ein niedriger p-Wert mit einem Wert deutlich unter 0.001. Es wird deshalb nur explizit auf die Signifikanz verwiesen, wenn das Signifikanzniveau von 0.001 überschritten wird und somit auf diesem Niveau keine Signifikanz besteht.

In einem weiteren Schritt wird für jedes Sprachenmass eine multiple Regressionsanalyse mit den verschiedenen geographischen Faktoren durchgeführt, inklusive der Berechnung des Bestimmtheitsmasses R^2 und der standardisierten Koeffizienten. Das R^2 misst wie gut die Regressionsfunktion an die empirischen Daten angepasst ist. Je grösser es ist, desto grösser ist der Anteil der durch das Regressionsmodell erklärten Streuung an der Gesamtstreuung. Für jedes R^2 und jeden Koeffizienten wird zudem der p-Wert berechnet, um zu prüfen ob das Signifikanzniveau von 0.001 überschritten wird. Um einen linearen Zusammenhang zwischen den Variablen und auch um eine Normalverteilung der Werte zu erreichen, wurde in einigen Fällen eine Transformation, entweder durch Logarithmieren oder Wurzelziehen, vorgenommen. In dieser Arbeit wird zum Teil auch das korrigierte Bestimmtheitsmass verwendet, damit Modelle mit einer unterschiedlichen Anzahl von Prädiktoren verglichen werden können, da damit die Anzahl der Regressoren nicht künstlich die Qualität der Anpassung erhöht (Backhaus et al. 2011). Auch für das R^2 wird jeweils die Signifikanz getestet. Für jedes Modell werden die Modellannahmen überprüft. Wenn es nicht anders vermerkt ist, erfüllen die Modelle diese Annahmen.

Von Currie & Mace (2012) wurde gezeigt, dass die Subsistenzstrategie einen Einfluss darauf haben kann, wie sich Umweltfaktoren auf die Sprachendiversität auswirken (siehe *Abschnitt 4.1.2*). Deshalb wird nebst dem erwähnten multiplen Regressionsmodell auch noch ein Modell berechnet, das eine Interaktion der Subsistenzstrategie mit den geographischen Faktoren berücksichtigt. Die beiden Modelle werden anschliessend mit einer Varianzanalyse (ANOVA) verglichen, um festzustellen welches die Daten besser beschreibt und ob die Subsistenzstrategie tatsächlich dazu führt, dass geographische Faktoren sich anders auf die Sprachenmasse auswirken.

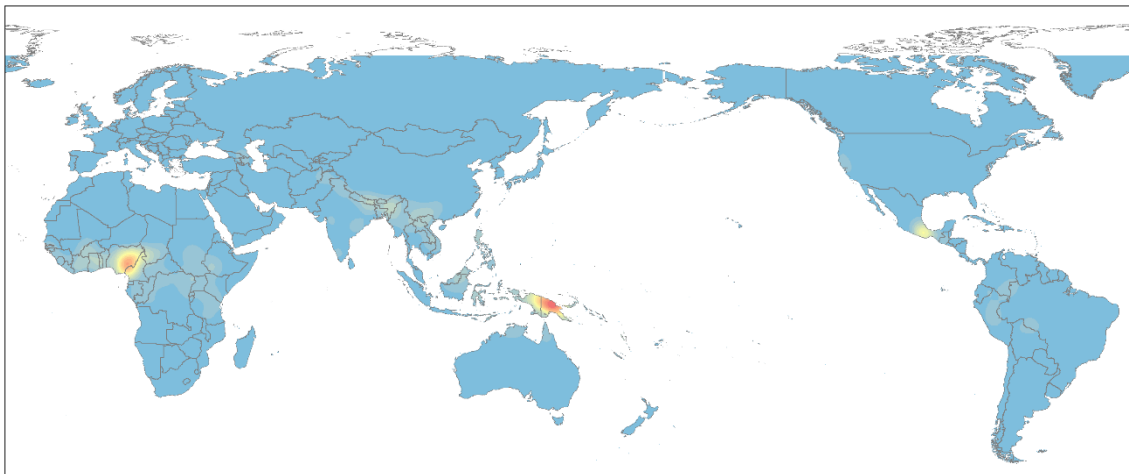
Sprachen derselben Sprachfamilie sind durch ihre gemeinsame Vergangenheit nicht unabhängig voneinander. Dies wird auch als *Galton's Problem* bezeichnet (Mace & Pagel 1994, Nettle 2009). So können kulturelle Faktoren dazu führen, dass verschiedene Familien unter bestimmten geographischen Faktoren eine andere Sprachendiversität aufweisen als Sprachen einer anderen Sprachfamilie. Um diesem Problem der zufälligen Effekte, die durch die Zugehörigkeit zu einer bestimmten Sprachfamilie entstehen, entgegen zu wirken, wird dies in einem statistischen Modell berücksichtigt. So genannte Mehrebenenanalysen oder gemischte Modelle (*linear mixed-effects models*) können mit hierarchischen Daten umgehen. Diese Modelle beinhalten sowohl *Fixed Effects*, wie auch *Random Effects*. Das *random intercept model*, welches zu diesen Modellen gehört, wird in dieser Arbeit verwendet. Als Gruppierungskriterium für mögliche *Random Effects* wird dabei die Zugehörigkeit zu einer Sprachfamilie verwendet. Sprachen einer gemeinsamen Sprachfamilie teilen demnach gemeinsame, unbeobachtete Variablen. Für jede Gruppe von Werten, in diesem Fall die Sprachfamilien, wird ein separater Schnitt der Regressionsgerade mit der y-Achse berechnet (Everitt & Hothorn 2011). Für alle gemischten Modelle wurde das R-Paket *nlme* verwendet (Pinheiro et al. 2013). Es wird zusätzlich noch ein Modell berechnet, welches wiederum eine Interaktion der Subsistenzstrategie mit den geographischen Faktoren berechnet. Die beiden gemischten Modelle werden anschliessend mit den beiden multiplen Regressionsmodellen, jeweils eines mit und eines ohne Interaktion mit der Subsistenzstrategie, verglichen.

Da sich beim Hinzufügen oder Weglassen die Koeffizienten einzelner Prädiktoren relativ stark verändern und da auch viele geographische Faktoren untereinander korrelieren, kann davon ausgegangen werden, dass eine starke Multikollinearität besteht. Dies wurde zusätzlich durch die Berechnung der *Variance Inflation Factors* (VIF) bestätigt (Backhaus et al. 2011). Dadurch ist es schwierig die p-Werte und die Koeffizienten zu interpretieren. Deshalb wurde jeweils noch ein Modell gerechnet, bei dem durch Reduktion der unabhängigen Variablen die Multikollinearität auf ein annehmbares Niveau sinkt. In diesem Modell wurden der Temperaturschwankungsbereich und die Länge der Wachstumsphase entfernt. Bei einer iterativen Überprüfung der VIF-Werte zeigte sich, dass dadurch die Multikollinearität gesenkt werden konnte. Zusätzlich wurde mit einer schrittweisen Überprüfung der AIC (Akaike information criterion)-Werte getestet, ob alle geographischen Faktoren zur Verbesserung des Modells beitragen. War dies nicht der Fall wurden die entsprechenden Faktoren aus dem Modell entfernt, was im Text jeweils vermerkt ist.

15 SPRACHENDIVERSITÄT

15.1 GLOBALE VERTEILUNG

Bevor analysiert wird, inwiefern die Sprachendiversität mit geographischen Faktoren zusammenhängt, wird zuerst die globale Verteilung betrachtet. Erstens werden die Werte auf einer Karte dargestellt, um zu sehen wie sich diese verteilen und zweitens wird noch genauer auf die geographische Breite eingegangen, da offenbar die Sprachendiversität mit dieser korreliert. Auf die geographische Länge wird nicht eingegangen. Diese wurde zwar auch überprüft, aber dort wurde festgestellt, dass sich keine Muster, ausser die der unregelmässig verteilten Landmassen, zeigen. Dies ist auch auf der Abbildung 32 zu sehen.



Eigene Darstellung. Quellen: UNEP (2013), Hammarström (2012)

Abbildung 32: Kerndichteschätzung der Sprachen. Rot bedeutet eine hohe Dichte, blau eine niedrige Dichte.

15.1.1 Lokale Sprachendichte

In Abbildung 33 ist die Verteilung der lokalen Sprachendichte über die Breitengrade zu sehen. Die Sprachfamilien mit den meisten Sprachen sind farblich hervorgehoben. In der Abbildung sind mehrere Peaks erkennbar. Die zwei grössten treten leicht nördlich respektive südlich des Äquators auf. In Abbildung 34 wurden die Werte der Nord- und Südhalbkugel übereinander gelegt. Dort sind diese beiden Peaks, die sich nun überlagern, nochmals deutlich sichtbar. Auch ist zu erkennen, dass ein zweiter (ca. bei 10 Grad) und ein dritter Peak (ca. bei 15 Grad) auf beiden Seiten des Äquators vorkommen. Der Spearman-Rangkorrelationskoeffizient beträgt -0.33.

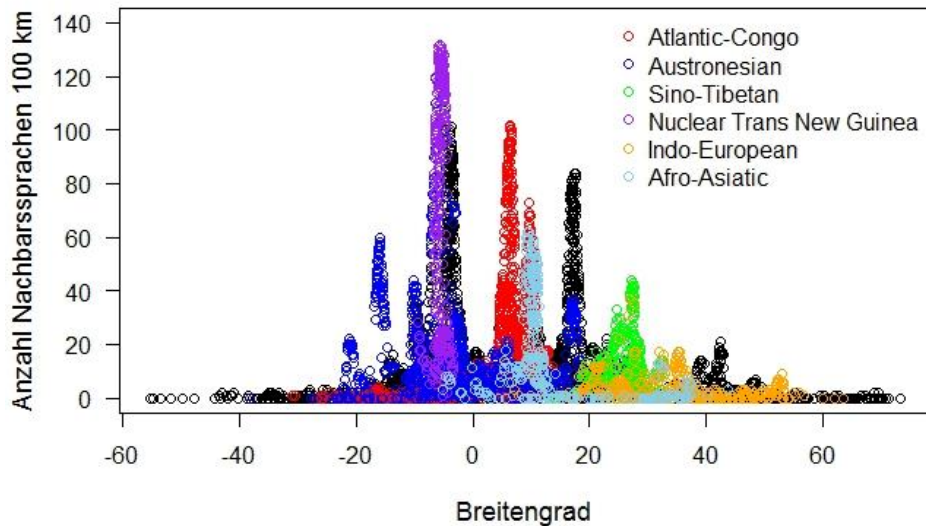


Abbildung 33: Verteilung der lokalen Sprachendichte über die Breitengrade. Farblich hervorgehoben sind die Sprachen der grössten Sprachfamilien.

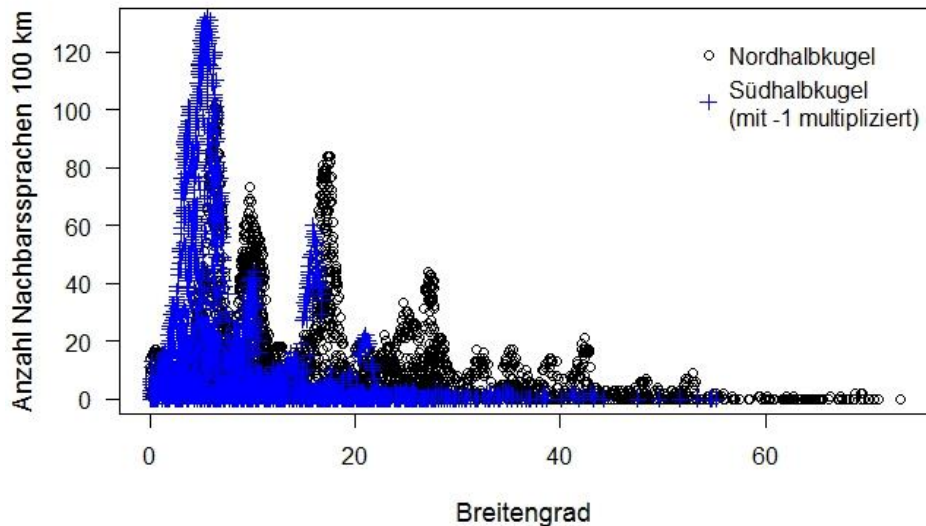


Abbildung 34: Verteilung der lokalen Sprachendichte über die Breitengrade. In blau die Werte der Südhalbkugel, in schwarze jene der Nordhalbkugel.

Der erste grosse Peak, welcher auf beiden Seiten des Äquators zu sehen ist, entsteht auf der Nordhalbkugel durch die hohen Dichten in Kamerun und Nigeria, und auf der Südhalbkugel durch die Sprachenvielfalt auf der Insel Neuguinea. Der zweite Peak ist nördlich des Äquators durch die Niger-Kongo-Sprachen und die Afroasiatischen Sprachen in Westafrika zu erklären. Südlich des Äquators entsteht der Peak durch die Vielzahl an überwiegend austronesischen Sprachen im südlichen Neuguinea. Der dritte Peak auf der Südhalbkugel ist auf die Sprachenvielfalt in der südpazifischen Republik Vanuatu zurückzuführen. Und auf der Nordhalbkugel erklärt sich der dritte Peak mit der hohen Sprachendichte der Otomangue-Sprachen in Mexiko.

Diesbezüglich stellt sich nun die Frage, ob die genannten drei Peaks sich zufällig überlagern oder ob ein Zusammenhang mit entsprechenden geographischen Faktoren besteht. Wie in Unterkapitel 7.7 gezeigt wurde, korrelieren einige geographische Faktoren mit der geographischen Breite. In den folgenden Abschnitten soll gezeigt werden, ob die Sprachendiversität mit einzelnen oder mehreren geographischen Faktoren erklärt werden kann. Möglicherweise können dadurch auch die Überlagerungen der hohen Dichten auf der Nord- und Südhalbkugel erklärt werden.

15.1.2 Geographisch Nächster Nachbar

In Abbildung 35 ist zu sehen wie sich die Distanzen zum Geographisch Nächsten Nachbar über die geographische Breite verteilen. Diese nehmen zum Äquator hin ab, wobei wiederum die extremsten Werte kurz davor auftreten. In Abbildung 36 ist die Zunahme der GNN-Distanz hin zu den Polen nochmals deutlicher sichtbar. Die Korrelation beträgt 0.27.

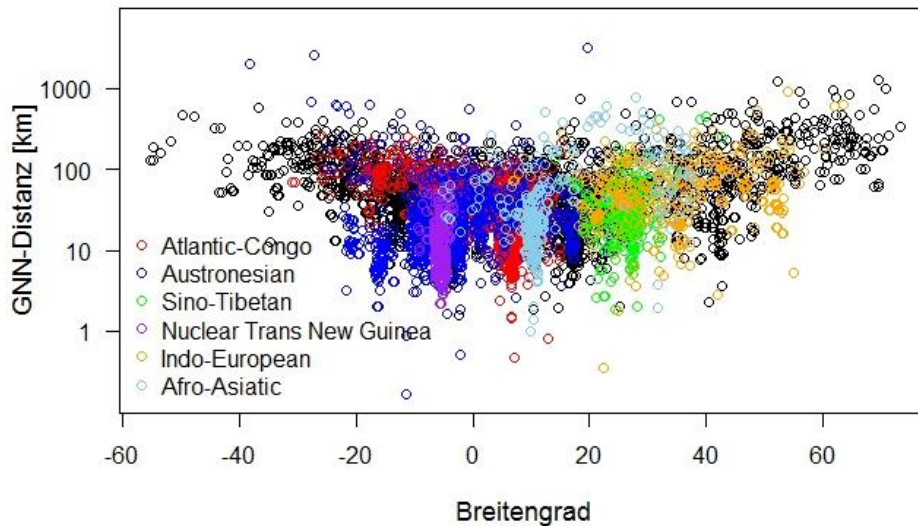


Abbildung 35: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Breitengrade. Farblich hervorgehoben sind die Sprachen der grössten Sprachfamilien.

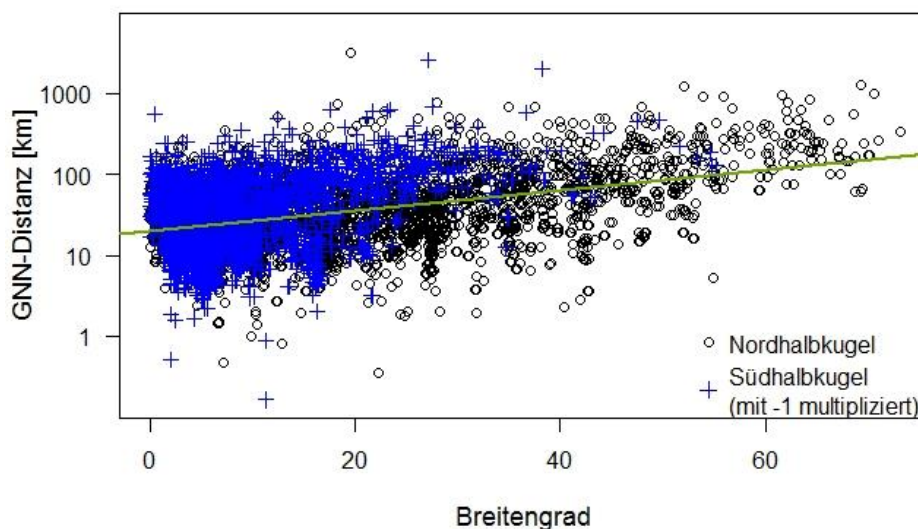


Abbildung 36: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Breitengrade. In blau die Werte der Südhalbkugel, in schwarz jene der Nordhalbkugel.

In Abbildung 35 zeigen die grossen Sprachfamilien zum Teil säulenartige Muster. Dies ist darauf zurückzuführen, dass diese alle ungefähr bei der gleichen geographischen Breite vorkommen, aber unterschiedlich dicht beieinander liegen. Die Sprachfamilie *Nuclear-Trans-Neu-Guinea* auf der Insel Neuguinea weist beispielsweise von West nach Ost immer mehr Sprachen auf, die somit eine immer kürzere GNN-Distanz haben. Trotz dieser säulenartigen Strukturen ist zu den Polen hin eine lineare Zunahme der logarithmierten GNN-Distanzen feststellbar. In den kommenden Kapiteln wird analysiert, welche geographischen Faktoren die Sprachendiversität erklären können.

15.2 MITTLERE JÄHRLICHE TEMPERATUR

15.2.1 Lokale Sprachendichte

In Abbildung 37 ist erkennbar, welche mittlere jährliche Temperatur bei jeder Sprache vorherrscht. Die höchsten lokalen Sprachendichten treten bei Werten grösser als 15 °C auf. Bei ca. über 28 °C sind dann wieder tiefere Dichten zu finden. Der Rangkorrelation nach Spearman beträgt 0.18.

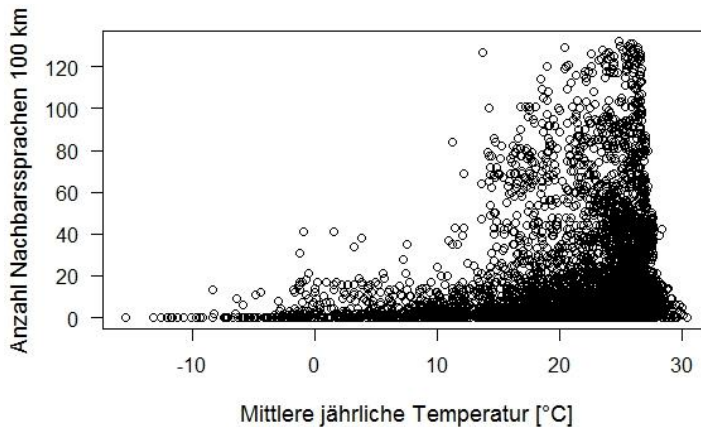


Abbildung 37: Verteilung der lokalen Sprachendichte über die mittlere jährliche Temperatur.

Fast über das ganze Temperaturspektrum, wo es Sprachen gibt, hat es Sprachen mit einer Sprachendichte von Null oder nur einigen wenigen Nachbarn. Sprachen mit einer hohen lokalen Dichte prägen jedoch das Bild. Die Rangkorrelation zeigt zwar einen positiven Zusammenhang zwischen der Anzahl Nachbarsprachen und der mittleren jährlichen Temperatur an, aber aufgrund der vielen tiefen Dichten, ist es dennoch schwierig dieses Resultat abschliessend zu interpretieren.

15.2.2 Geographisch Nächster Nachbar

Abbildung 38 verdeutlicht, dass die logarithmierte Distanz zum Geographisch Nächsten Nachbar mit zunehmender mittlerer jährlicher Temperatur abnimmt. Die Rangkorrelation beträgt -0.17.

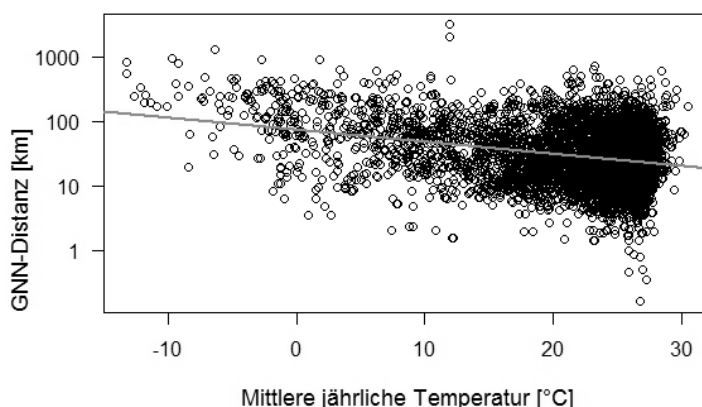


Abbildung 38: Verteilung der GNN-Distanz in logarithmierter Darstellung über die mittlere jährliche Temperatur.

Zwischen der GNN-Distanz und der mittleren jährlichen Temperatur besteht ein leichter negativer linearer Zusammenhang, d.h. der Geographisch Nächste Nachbar ist mit steigender Temperatur in zunehmend näherer Entfernung zu finden ist. Der Betrag des Korrelationskoeffizienten ist praktisch gleich stark ausgeprägt wie bei der lokalen Sprachendichte. Die abrupte Abnahme von Sprachen an Orten ab 28 °C ist

einerseits darauf zurückzuführen, dass es auf der Welt nicht viele Regionen gibt, die noch höhere Temperaturen aufweisen und andererseits sind solche Gebiete auch eher trockener als andere. Dies führt möglicherweise auch dazu, dass es dort weniger Sprachen hat. Die meisten dieser Regionen liegen in der Sahara. Ob sich dies bestätigt, wird sich in der Analyse mit dem jährlichen Niederschlag zeigen.

15.3 JÄHRLICHER TEMPERATURSCHWANKUNGSBEREICH

15.3.1 Lokale Sprachendichte

Die lokale Sprachendichte ist dort besonders hoch, wo es nur sehr wenig Temperaturschwankung während des Jahres gibt, wie in Abbildung 39 zu sehen ist. Ab einem Schwankungsbereich von ca. 25 °C hat kaum mehr eine Sprache mehr als 20 Nachbarsprachen innerhalb von 100 km. Der Rangkorrelationskoeffizient zeigt mit -0.4 einen deutlich höheren linearen Zusammenhang an, als noch bei der mittleren jährlichen Temperatur.

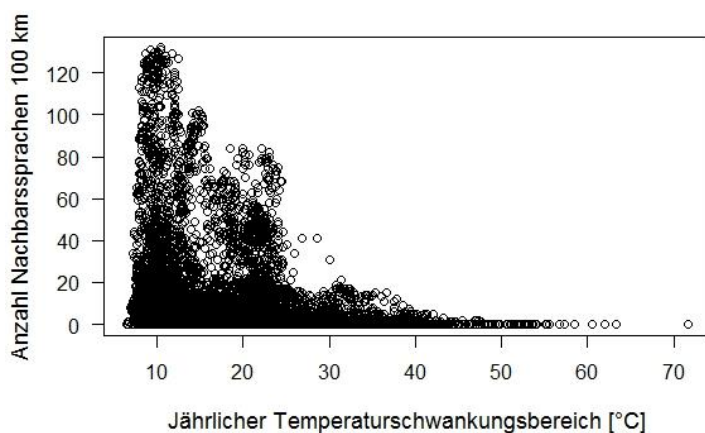


Abbildung 39: Verteilung der lokalen Sprachendichte über den jährlichen Temperaturschwankungsbereich.

Bei der Verteilung der lokalen Sprachendichten auf den Temperaturschwankungsbereich lassen sich drei Stufen ausmachen. Die erste Stufe, die von ca. 5 bis 15 °C reicht, wird vor allem von den Sprachen auf Neuguinea und den Austronesischen Sprachen gebildet. Gebiete mit einem solchen Temperaturschwankungsbereich sind nur in Äquatornähe zu finden. Die zweite Stufe von ca. 15 bis 25 °C beinhaltet vor allem Niger-Kongo, afroasiatische und sinotibetische Sprachen. Ab 25 °C sind fast nur noch sinotibetische und indogermanische Sprachen zu finden.

Ähnlich wie schon bei der Temperatur zeigt sich hier zwar klar in welche Richtung höhere Sprachendichten vorkommen, dennoch ist es schwierig eine generelle Aussage zu machen, da das Bild stark durch die niedrigen lokalen Dichten über den gesamten Bereich beeinträchtigt ist.

15.3.2 Geographisch nächster Nachbar

Die GNN-Distanz nimmt mit steigendem Temperaturschwankungsbereich deutlich zu, wie in Abbildung 40 klar zu erkennen ist. Der Rangkorrelationskoeffizient beträgt 0.37 und ist damit deutlich höher als bei der Temperatur mit -0.17.

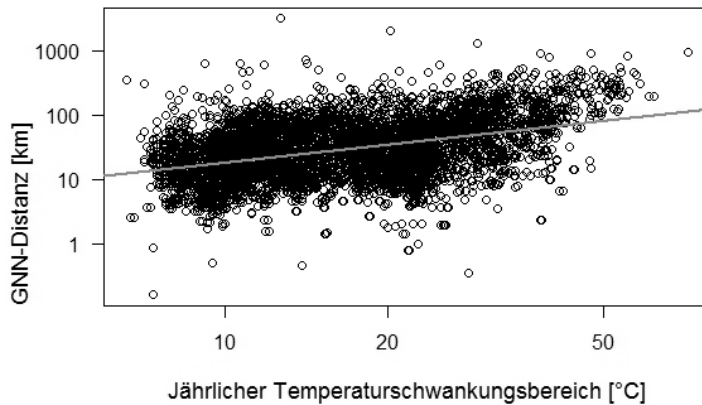


Abbildung 40: Verteilung der GNN-Distanz in logarithmierter Darstellung über den jährlichen Temperaturschwankungsbereich in logarithmierter Darstellung.

ρ ist ähnlich gross wie bei der lokalen Sprachdichte im vorangegangenen Abschnitt. In Abbildung 40 zeigt sich jeweils um 10 °C und um 20 °C eine Häufung, und zwischen beiden eine Lücke von Sprachen. Diese Lücke war schon im vorangegangenen Abschnitt bei den lokalen Dichten zu sehen. Die anschliessenden tieferen Distanzen um 20 °C sind vor allem auf sinotibetische und afroasiatische Sprachen zurückzuführen. Diese weisen offenbar eine leicht kleinere GNN-Distanz auf als andere Sprachen mit denselben Temperaturschwankungen.

15.4 JÄHRLICHER NIEDERSCHLAG

15.4.1 Lokale Sprachdichte

Die Verteilung der lokalen Sprachdichten auf den jährlichen Niederschlag ist stark gestreut. Wie in Abbildung 41 festzustellen ist, treten höhere Dichten mit zunehmendem Niederschlag auf. Doch selbst bei höherem Niederschlag, insbesondere ab 4'000 mm, gibt es weiterhin viele tiefe Dichten. Der Korrelationskoeffizient beträgt trotzdem 0.41.

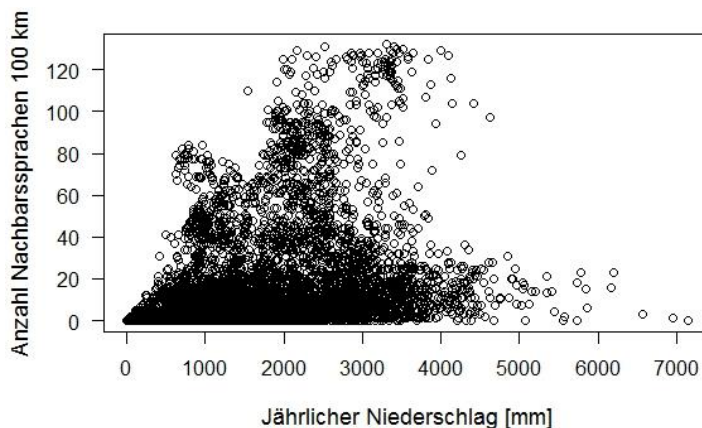


Abbildung 41: Verteilung der lokalen Sprachdichte über den jährlichen Niederschlag.

Mit diesem Wert bewegt sich der Zusammenhang zwischen der lokalen Dichte und dem jährlichen Niederschlag in der Grössenordnung des jährlichen Temperaturschwankungsbereichs. Der Fakt, dass ab 4'000 mm nur noch wenige Sprachen und fast nur solche mit tiefen lokalen Dichten vorkommen, hat zwei Gründe. Erstens gibt es nur sehr wenige Gebiete mit einem jährlichen Niederschlag über 4'000 mm und zweitens liegen fast alle Sprachen im Landesinnern von Neuguinea und Borneo, weshalb sie jeweils ähnliche Dichten haben.

15.4.2 Geographisch Nächster Nachbar

Abbildung 42 zeigt eindeutig, dass die logarithmierte Distanz zum Geographisch Nächsten Nachbar mit zunehmendem Niederschlag abnimmt. Der Rangkorrelationskoeffizient beträgt -0.37 . Somit ist dieser Wert vergleichbar mit jenem der lokalen Sprachendichte (0.41).

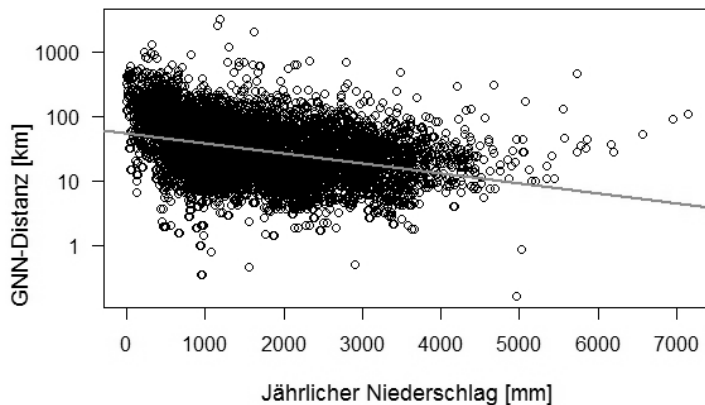


Abbildung 42: Verteilung der GNN-Distanz in logarithmierter Darstellung über den jährlichen Niederschlag.

Die Grössenordnung des ρ ist vergleichbar mit dem jährlichen Temperaturschwankungsbereich. Beide geographischen Faktoren vermögen einen ansehnlichen Teil der Varianz der GNN-Distanz zu erklären. Da diese beiden stark korrelieren, mit einem r von -0.79 , erklären sie möglicherweise ähnliche Teile der Varianz. Dies wird bei der multiplen Regressionsanalyse in *Unterkapitel 15.12* verdeutlicht werden.

15.5 SAISONALITÄT DES NIEDERSCHLAGS

15.5.1 Lokale Sprachendichte

Wie in Abbildung 43 erkennbar ist, treten, je unregelmässiger verteilt der Niederschlag über das Jahr fällt, kleinere lokale Dichten auf. Diese Abnahme von hohen Dichten ist allerdings nicht stetig, sondern weist zwischen 50 und 75 eine Lücke auf. Mit einem ρ von -0.054 besteht daher nur eine sehr geringe Korrelation zwischen der lokalen Dichte und der Saisonalität des Niederschlags.

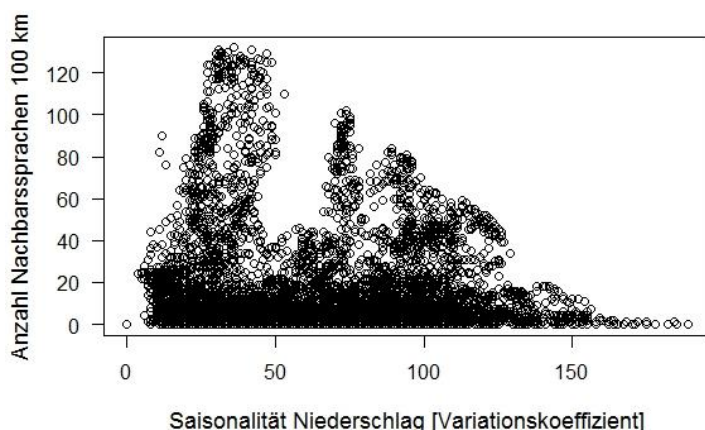


Abbildung 43: Verteilung der lokalen Sprachendichte über die Saisonalität des Niederschlags.

Im Gegensatz zum jährlichen Niederschlag scheint die Verteilung des Niederschlags über das Jahr kaum Auswirkungen auf die lokale Sprachendichte zu haben. Die Lücke zwischen den Werten 50 und 75 lässt sich am ehesten so erklären, dass es zwar viele Regionen mit diesen Werten gibt, jedoch keine Region mit

hohen Sprachendichten, die in ein solches Gebiet fällt. Regionen mit gleicher Saisonalität des Niederschlags verlaufen oft entlang der Breitengrade und da die Sprachen der beiden Sprachfamilien mit den höchsten Dichten auch eher west-ostwärts streuen, ist die Wahrscheinlichkeit deutlich höher, dass sie die gleiche Saisonalität des Niederschlags aufweisen. Da ausserdem die lokale Sprachendichte offenbar nicht stark von der jährlichen Verteilung des Niederschlags abhängig ist, ist auch nicht zu erwarten, dass die auftretenden Muster erklärt werden können.

15.5.2 Geographisch Nächster Nachbar

Mit zunehmender Saisonalität des Niederschlags nimmt die logarithmierte Distanz zum Geographisch Nächsten Nachbarn nur leicht zu (Abbildung 44). Der Rangkorrelationskoeffizient beträgt rund 0.05. Damit ist dieser Wert im Betrag sehr ähnlich wie jener bei der lokalen Sprachendichte (-0.054).

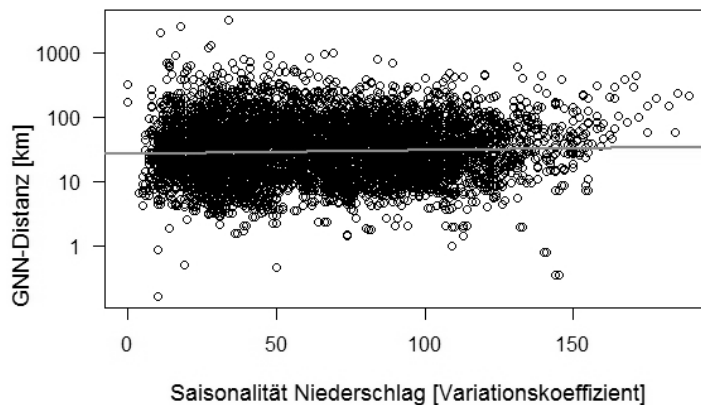


Abbildung 44: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Saisonalität des Niederschlags.

Die Lücke zwischen 50 und 75, die sich bei der lokalen Dichte zeigte, ist hier nur als kleine konkave Wölbung unterhalb von 10 km Distanz zu erkennen. Obwohl die Saisonalität des Niederschlags sowohl mit dem jährlichen Niederschlag als auch mit dem Temperaturschwankungsbereich korreliert, die beide eine deutlich stärkere Korrelation mit der GNN-Distanz aufweisen, besteht kaum ein Zusammenhang zwischen der Distanz zum Geographisch Nächsten Nachbar und der Verteilung des Niederschlags übers Jahr.

15.6 NETTOPRIMÄRPRODUKTION

15.6.1 Lokale Sprachendichte

Wie in Abbildung 45 ersichtlich ist, kommen die hohen lokalen Sprachendichten mit höherer mittlerer Nettoprimärproduktion vor. ρ beträgt 0.34. Ab ca. $5 \text{ gC}/\text{m}^2/\text{d}$ hat es kaum mehr Sprachen mit mehr als 20 Nachbarsprachen.

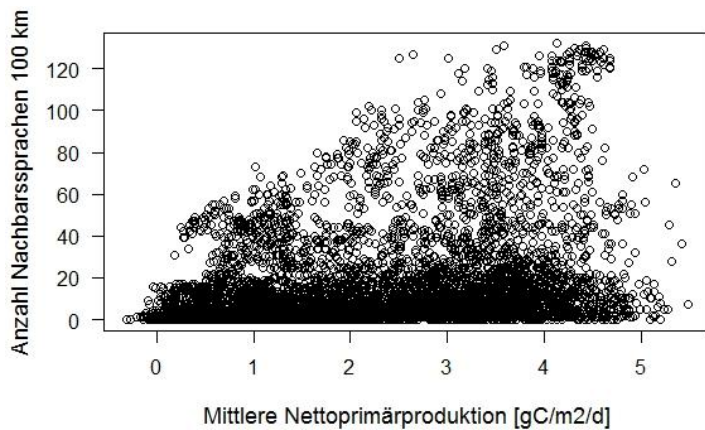


Abbildung 45: Verteilung der lokalen Sprachendichte über die mittlere Nettoprimärproduktion.

Mit einem ρ von 0.34 besteht ein beachtlicher Zusammenhang zwischen der NPP und der lokalen Sprachendichte. Das geringe Vorkommen von Sprachen über $5 \text{ gC/m}^2/\text{d}$ erklärt sich damit, dass es fast keine Gebiete auf der Welt gibt, die eine so hohe NPP aufweisen.

15.6.2 Geographisch Nächster Nachbar

Die logarithmierte Distanz zum Geographisch Nächsten Nachbarn nimmt mit zunehmender mittlerer Nettoprimärproduktion ab (Abbildung 46). Der Rangkorrelationskoeffizient beträgt -0.32. Dieser Wert ist damit jenem bei der lokalen Dichte mit 0.34 im vorangegangenen Abschnitt sehr ähnlich, wenn auch mit umgekehrtem Vorzeichen.

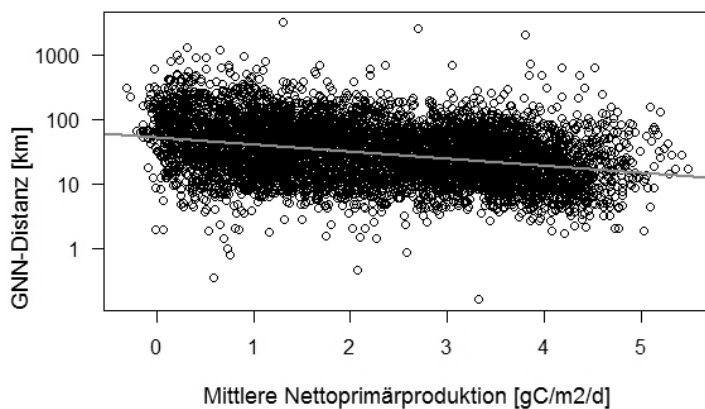


Abbildung 46: Verteilung der GNN-Distanz in logarithmierter Darstellung über die mittlere Nettoprimärproduktion.

Die NPP korreliert deutlich mit dem Temperaturschwankungsbereich und dem Niederschlag. Diese beiden Faktoren wiederum zeigen eine stärkere Korrelation mit der GNN-Distanz als die NPP. Bei der Auswertung des multiplen Regressionsmodells wird sich zeigen, ob die NPP zusätzliche Anteile erklären kann gegenüber den beiden anderen Faktoren.

15.7 LÄNGE DER WACHSTUMSPHASE

15.7.1 Lokale Sprachendichte

In Abbildung 47 ist zu sehen, dass ungefähr bis zu einer 280 Tage langen Wachstumsphase immer höhere lokale Sprachendichten vorkommen. Danach werden die höchsten vorkommenden lokalen Dichten immer tiefer. Bei einer Länge von 365 und 366 Tagen tritt von wenigen Nachbarsprachen bis über 130 das

ganze Spektrum auf. Der Rangkorrelationskoeffizient beträgt 0.36. Dies ist ein ähnlicher Wert wie beim jährlichen Niederschlag.

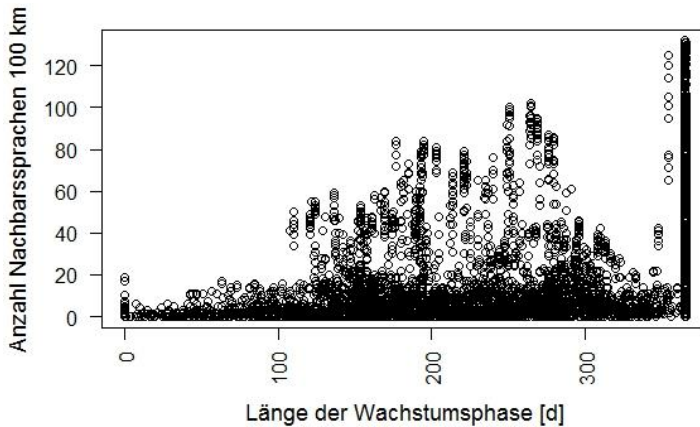


Abbildung 47: Verteilung der lokalen Sprachendichte über die Länge der Wachstumsphase.

Die gleichmäßige Verteilung von Sprachendichten bei einer Länge von 365 und 366 Tagen ist auf die unterschiedlichen Gebiete, wo diese auftreten, zurückzuführen. Hauptsächlich treten solch lange Wachstumsphasen auf Neuguinea, auf der Insel Borneo und im westlichen Amazonas Regenwald auf. Die Sprachendichten in Neuguinea gehören dabei zu den höchsten, wobei es auch dort Sprachen mit nur wenigen Nachbarsprachen gibt. Jene auf der Insel Borneo und im Regenwald erreichen lediglich Werte bis ca. 20 Nachbarsprachen. Aus diesen Gründen hat es bei 365 und 366 Tagen so viele verschiedene Dichten.

Da die Länge der Wachstumsphase stark mit dem jährlichen Niederschlag, dem Temperaturschwankungsbereich wie auch mit der NPP korreliert, wird sich erst bei der statistischen Auswertung in einem multiplen Regressionsmodell (*Unterkapitel 15.12*) zeigen, ob dieser geographische Faktor andere Teile der Varianz erklärt als die anderen mit ihm korrelierenden Faktoren.

15.7.2 Geographisch Nächster Nachbar

Die logarithmierte Distanz zum geographisch nächsten Nachbar nimmt mit der Länge der Wachstumsphase zu. Mit einem ρ von -0.33 bewegt sich die Stärke dieses Zusammenhangs in einem Bereich wie bei der lokalen Dichte. Auffällig ist wiederum das Vorkommen von diversen Dichten bei einer Länge von 365 und 366 Tagen. Im Gegensatz zur lokalen Sprachendichte tritt dieses Phänomen hier auch bei einer Länge von null Tagen deutlich auf.

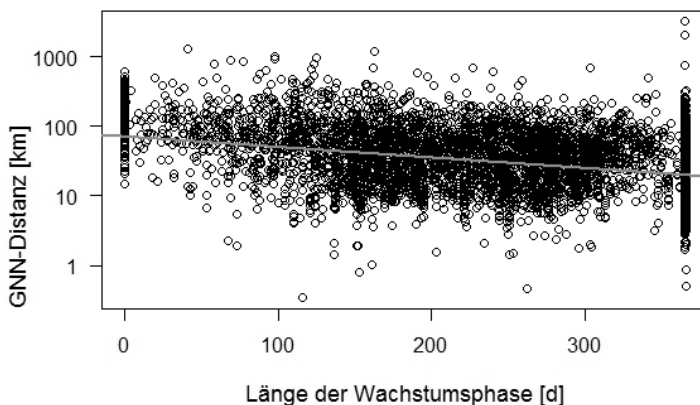


Abbildung 48: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Länge der Wachstumsphase.

Der Betrag von ρ ist leicht tiefer als bei der lokalen Sprachendichte. Eine Wachstumsphase von null Tagen, welche hier eine Anhäufung von diversen Dichten hat, kommt vor allem in Wüstenregionen vor. Da dies relativ grosse Flächen betrifft, fallen dementsprechend mehr Sprachen dort hinein. Hingegen hat es nur wenige Regionen mit einer Länge von nur wenigen Tagen, was die nicht so dicht vorkommenden Punkte zwischen 0 und 100 Tagen in Abbildung 48 zeigen.

15.8 KÜSTENDISTANZ

15.8.1 Lokale Sprachendichte

Abbildung 49 stellt die Verteilung der lokalen Sprachendichten auf die Distanz zur Küste dar. Auffällig daran sind die vier deutlichen Peaks. Obwohl dadurch keine kontinuierliche Abnahme der hohen Dichten von den Küsten weg beobachtet werden kann, ist dennoch deutlich erkennbar, dass näher an der Küste höhere Sprachendichten erwartet werden können. Der Rangkorrelationskoeffizient beträgt -0.22 .

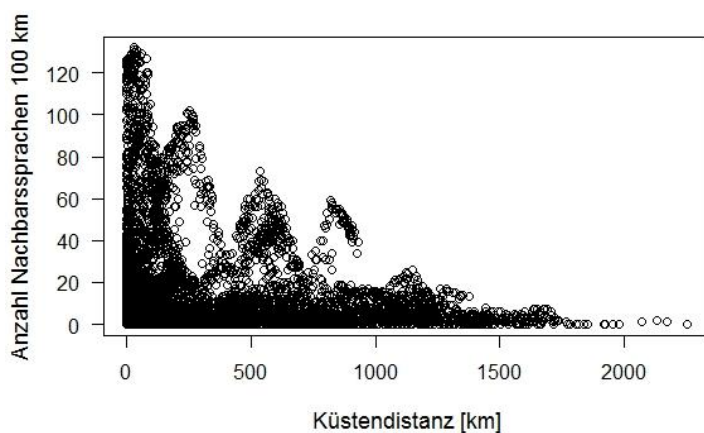


Abbildung 49: Verteilung der lokalen Sprachendichte über die Küstendistanz.

Die hohen Sprachendichten des ersten Peaks setzen sich hauptsächlich aus Sprachen der Insel Neuguinea und austronesischen Sprachen zusammen. Da sich diese fast alle auf Inseln befinden, können die Distanzen zur Küste nirgends gross sein. Der zweite Peak setzt sich ausschliesslich aus Niger-Kongo-Sprachen zusammen. Dieser bildet sich, da ein Cluster dieser Sprachen ca. 250 km von der Küste entfernt liegt. Der dritte Peak wird von sinotibetischen Sprachen, Niger-Kongo-Sprachen und afroasiatischen Sprachen gebildet. Letztere bilden zusammen einen Cluster ca. 500 km von der Küste entfernt. Auch die sinotibetischen Sprachen bilden im äussersten Osten von Indien in einer Entfernung von 500 km zum Meer einen Cluster.

15.8.2 Geographisch Nächster Nachbar

Je weiter eine Sprache von der Küste entfernt ist, desto weiter entfernt ist auch ihr nächster Nachbar (Abbildung 50). Obwohl dieser Trend klar zu erkennen ist, kommen in den ersten 100 km von der Küste entfernt diverse GNN-Distanzen vor. Die statistische Analyse ergibt ein ρ von 0.25 .

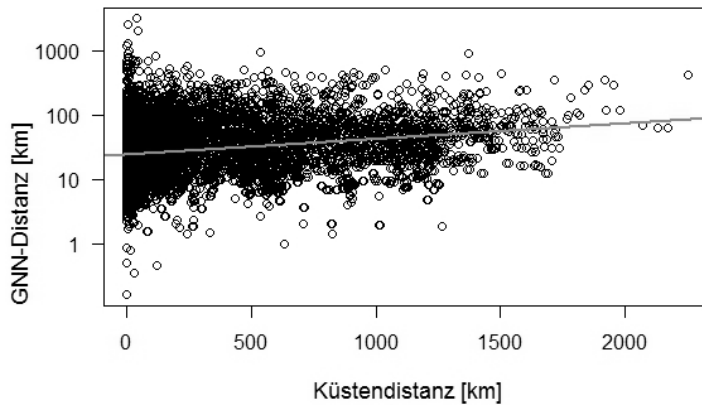


Abbildung 50: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Küstendistanz.

Alle bisher überprüften geographischen Faktoren hatten die Tendenz, dass Sprachen auf ungefähr gleicher geographischer Breite ähnliche Werte aufweisen. Sie korrelierten als klimatisch bedingte Variablen fast alle mit den Breitengraden. Dies ist bei der Küstendistanz nicht mehr der Fall (siehe *Unterkapitel 7.7*). So kommen hier Sprachen hoch im Norden vor, die eine ähnliche Küstendistanz aufweisen wie Sprachen am Äquator und umgekehrt. Die Distanz ist natürlich auch stark von der Form und Grösse der Landmassen abhängig. So sind Sprachen in Europa selten mehr als 1'000 km vom Meer entfernt.

15.9 HÖHENMODELL

15.9.1 Lokale Sprachendichte

In Abbildung 51 ist klar ersichtlich, dass hohe Sprachendichten vor allem in tieferen Regionen vorkommen. Weiter ist erkennbar, dass die hohen Dichten gegen 3'000 m ü.M. immer weniger werden. Der Rangkorrelationskoeffizient beträgt 0.065.

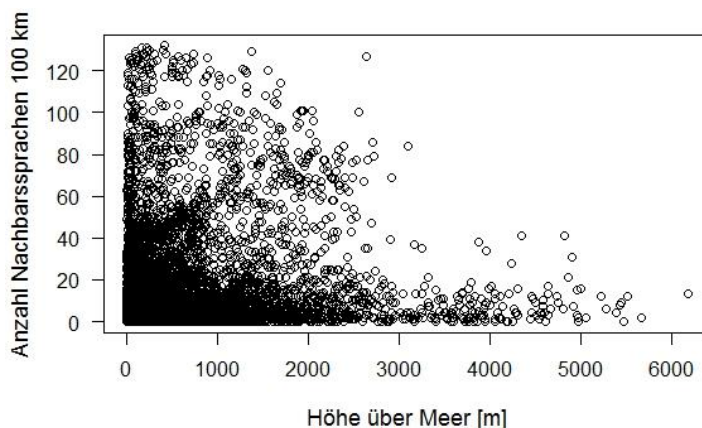


Abbildung 51: Verteilung der lokalen Sprachendichte über die Höhe über Meer.

Wie immer bei der lokalen Sprachendichte, gibt es auch hier tiefe Werte bei allen vorkommenden Höhen. Obwohl es scheint, dass bis ca. 3'000 m die hohen Dichten immer weniger werden, zeigt die Korrelation, dass praktisch kein Zusammenhang zwischen diesen beiden Variablen besteht. Die Sprachen ab 3'000 m sind fast nur noch sinotibetische und indogermanische Sprachen.

15.9.2 Geographisch Nächster Nachbar

Wie sich in Abbildung 52 zeigt, besteht kein Zusammenhang zwischen der logarithmierten GNN-Distanz und der Höhe über dem Meeresspiegel. Dies zeigt sich auch im ρ , welches lediglich -0.018 beträgt. Zudem ist dieser Wert nicht signifikant unterschiedlich von Null, weshalb nicht einmal der marginale Zusammenhang der dadurch angezeigt wird, interpretiert werden darf.

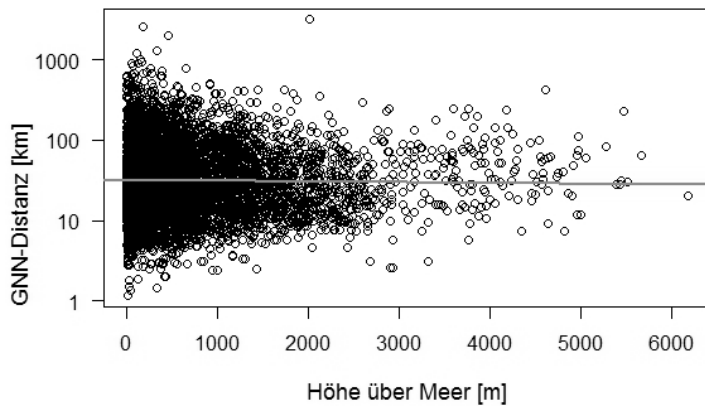


Abbildung 52: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Höhe über Meer.

Die Distanz zum nächsten Nachbarn scheint dementsprechend unabhängig von der Höhe über dem Meeresspiegel zu sein. Da jedoch nur sphärische Distanzen berechnet werden und keine Wegdistanzen, wird nicht berücksichtigt, dass Sprachen aufgrund der Topographie zum Teil viel weiter voneinander entfernt liegen, als es die Luftliniendistanz suggeriert. Ob sich die Topographie möglicherweise dennoch auf die Sprachendiversität auswirkt, wird sich im nächsten Unterkapitel zeigen.

15.10 HÖHENDIFFERENZ

15.10.1 Lokale Sprachendichte

In Abbildung 53 ist der Zusammenhang zwischen der Anzahl an Nachbarsprachen und der maximalen Höhendifferenz 50 km um jede Sprache dargestellt. In der Verteilung ist keine klare Tendenz zu erkennen. Zudem zeigen sich viele verschiedene Muster. Dennoch beträgt der Rangkorrelationskoeffizient relativ hohe 0.37 .

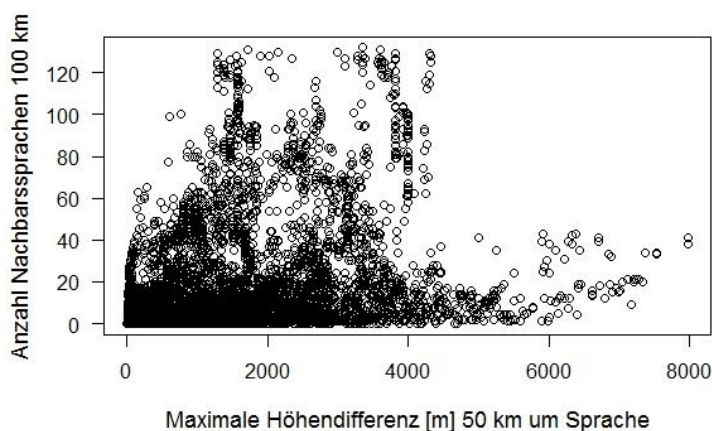


Abbildung 53: Verteilung der lokalen Sprachendichte über die maximale Höhendifferenz 50 km um die jeweilige Sprache.

Die zahlreichen Muster in der Verteilung ergeben sich, da Sprachen in räumlichen Clustern jeweils annähernd ähnliche Höhendifferenzen haben wie andere Sprachen in demselben Cluster. Da jedoch einige Sprachen am Rand und andere in der Mitte eines Clusters liegen, variiert die Anzahl an Nachbarsprachen. Deshalb sind in Abbildung 53 senkrechte Muster zu erkennen.

15.10.2 Geographisch nächster Nachbar

Wie in Abbildung 54 deutlich zu sehen ist, nimmt die logarithmierte Distanz zum Geographisch Nächsten Nachbarn mit der maximalen Höhendifferenz 50 km um jede Sprache ab. Das ρ beträgt -0.3, welches somit tiefer ist als bei der lokalen Sprachendichte (0.37).

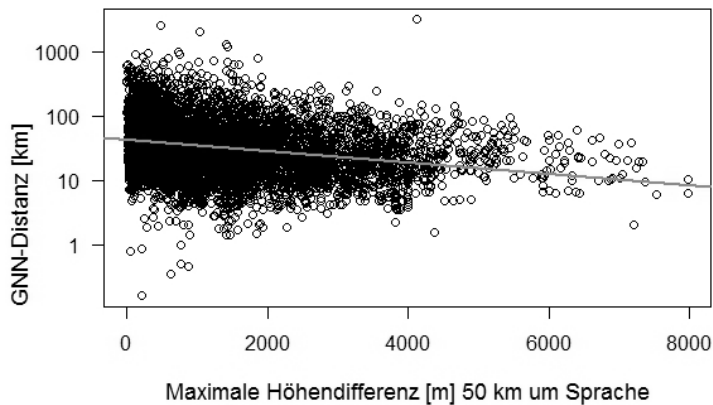


Abbildung 54: Verteilung der GNN-Distanz in logarithmierter Darstellung über die maximale Höhendifferenz 50 km um die jeweilige Sprache.

In topographisch extremeren Gebieten liegen die Sprachen offensichtlich näher beieinander als in eher flachen Gebieten. Da keine Korrelation in Zusammenhang mit der Höhe über dem Meeresspiegel an sich nachgewiesen werden konnte, scheint es, dass die topographische Heterogenität nur in einem gewissen lokalen Umkreis der Sprachen ausschlaggebend ist.

15.11 ZUSAMMENFASSUNG EINZELNE PRÄDIKTOREN

In der folgenden Tabelle 5 sind nochmals alle Rangkorrelationskoeffizienten zwischen der lokalen Sprachendichte respektive der GNN-Distanz und den einzelnen geographischen Faktoren aufgeführt.

Tabelle 5: Übersicht über die Rangkorrelationskoeffizienten ρ der lokalen Sprachendichte und der GNN-Distanz für jeden einzelnen geographischen Faktor. Die grössten Werte sind fett hervorgehoben. Kursive Werte sind nicht signifikant. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1.

Geo. Faktor	Lokale Sprachendichte		GNN-Distanz	
	ρ	Signifi- kanz	ρ	Signifi- kanz
Breitengrad	-0.325	***	0.267	***
Temperatur	0.180	***	-0.173	***
Temperaturschwankungsbereich	-0.395	***	0.373	***
Niederschlag	0.410	***	-0.370	***
Saisonalität Niederschlag	-0.054	***	0.048	***
NPP	0.336	***	-0.319	***
Wachstumsphase	0.361	***	-0.326	***
Küstendistanz	-0.221	***	0.254	***
Höhenmodell	0.065	***	<i>-0.018</i>	**
Höhendifferenz	0.371	***	-0.303	***

Der Korrelationskoeffizient bei den GNN-Distanzen hat jeweils umgekehrte Vorzeichen, wobei gilt, je grösser dieser ist, desto weiter liegen die Sprachen auseinander respektive desto kleiner ist die Sprachendiversität. Dies im Gegensatz zur lokalen Sprachendichte, bei der gilt, je grösser diese ist, desto grösser die Sprachendiversität. Die Beträge der Koeffizienten liegen pro geographischer Faktor recht nahe beieinander. Die grösste Differenz beträgt lediglich 0.073 und betrifft die Höhendifferenz. Die lokale Sprachendichte weist mit dem Niederschlag den grössten Zusammenhang auf. Danach folgen der Temperaturschwankungsbereich und die Höhendifferenz. Die GNN-Distanz hingegen weist mit dem Temperaturschwankungsbereich die stärkste Korrelation auf. Am zweitstärksten ist der Zusammenhang zum Niederschlag und am drittstärksten zur Länge der Wachstumsphase.

15.12 MULTIPLE REGRESSIONSANALYSE

Bis jetzt wurde der Zusammenhang zwischen der Sprachendiversität und den verschiedenen geographischen Faktoren jeweils einzeln betrachtet. Nun wird eine multiple Regressionsanalyse durchgeführt, um zu analysieren, ob und wie stark die geographischen Faktoren insgesamt die Sprachendiversität erklären können.

15.12.1 Lokale Sprachendichte

In Tabelle 6 sind die Resultate der multiplen Regressionsanalyse aufgeführt. Darin ist auch vermerkt, welche Faktoren welcher Transformation unterzogen wurden. Ebenfalls log₁₀-transformiert wurde die abhängige Variable, die lokale Sprachendichte. Die Koeffizienten geben jeweils den Effekt an, den ein Anstieg der unabhängigen Variablen um eine Einheit auf die abhängige Variable hat. Damit diese Werte, die sich auf sehr unterschiedlich grosse Zahlen und Bereiche beziehen, untereinander verglichen werden können, sind daneben noch die standardisierten Koeffizienten aufgeführt. Der Achsenabschnitt gibt den Schnittpunkt der Regressionsgerade mit der Y-Achse an (Backhaus et al. 2011). Eine schrittweise Analyse der AIC-Werte hat ergeben, dass das beste Modell erreicht wird, wenn alle geographischen Faktoren in die multiple Regressionsanalyse einbezogen werden.

Tabelle 6: Multiple Regressionsanalyse mit der logarithmierten lokalen Sprachendichte als abhängige Variable und den geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "***"<0.01 "**"<0.05 " "<0.1 " "<1.

Geo. Faktor	Transformation	Koeffizienten	Standardisierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		0.051	0.926	0.65739	
Temperatur		0.019	0.126	< 2e-16	***
Temperaturschwankungsbereich	log ₁₀	-0.628	-0.122	< 2e-16	***
Niederschlag	√	0.006	0.072	2.24E-10	***
Saisonalität Niederschlag	√	0.049	0.111	< 2e-16	***
NPP		-0.043	-0.056	2.05E-08	***
Wachstumsphase		0.001	0.085	8.10E-11	***
Küstendistanz	√	-0.003	-0.035	9.94E-06	***
Höhenmodell+1	log ₁₀	0.038	0.025	0.00365	**
Höhendifferenz	√	0.014	0.228	< 2e-16	***

p-Wert	<2E-16	***
Bestimmtheitsmass R ²	0.373	

Insgesamt kann mit diesem linearen Modell 37.3 % der gesamten Streuung der lokalen Sprachendichte erklärt werden. Das heisst auch, dass rund 63 % nicht durch dieses Modell erklärt werden können. Da mehrere einzelne geographische Faktoren einen relativ starken Zusammenhang mit der lokalen Sprachendichte aufweisen, hätte möglicherweise ein höherer Wert erwartet werden können. Jedoch gilt es dabei zu berücksichtigen, dass viele der Faktoren untereinander korrelieren. Somit überlagern sich ihre Erklärungen der Streuung der lokalen Sprachendichte teilweise. Dies ist in Tabelle 7 weniger der Fall, da dort die Multikollinearität weniger stark ist. Wenn noch die geographische Breite im ersten multiplen Regressionsmodell ergänzt wird, ergibt sich ein korrigiertes R², welches nur 0.002 Punkte grösser ist, im Vergleich zum korrigierten R² des Modells in Tabelle 6. Das heisst, dass die Breitengrade kaum zusätzliche Erklärungen zur Grösse der lokalen Sprachendichte liefern. Da einige geographische Faktoren mit den Breitengraden korrelieren, hängt auch die Sprachendichte mit diesen zusammen. Die geographischen Faktoren können somit auch ohne die geographische Breite die Zunahme der Sprachendiversität in

Richtung Äquator gleich gut erklären. Wird das multiple Regressionsmodell in Tabelle 6 um den Faktor Subsistenzstrategie ergänzt, lässt sich damit 40.78 % der Varianz der Sprachendichte erklären. Dies sind gut 3 Prozentpunkte mehr als im ursprünglichen Modell. Ein Modellvergleich folgt in Unterkapitel 15.14.

Tabelle 7: Multiple Regressionsanalyse mit der logarithmierten lokalen Sprachendichte als abhängige Variable und sechs geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 "****"<0.1 "****"<1.

Geo. Faktor	Transformation	Koeffizienten	Standardisierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		-0.859	0.920	< 2e-16	***
Temperatur		0.027	0.182	< 2e-16	***
Niederschlag	√	0.012	0.144	< 2e-16	***
Saisonalität Niederschlag	√	0.027	0.061	< 2e-16	***
Küstendistanz	√	-0.005	-0.055	7.53E-14	***
Höhenmodell+1	log10	0.059	0.038	5.82E-06	***
Höhendifferenz	√	0.013	0.224	< 2e-16	***

p-Wert	<2E-16	***
Bestimmtheitsmass R ²	0.355	

In Tabelle 7 sind die Resultate der multiplen Regressionsanalyse ohne den Temperaturschwankungsbereich, die Wachstumsphase und die NPP dargestellt. Bei einer schrittweisen Analyse der AIC-Werte ergab sich ohne NPP das beste Modell, weshalb dieser Faktor, nebst den beiden Faktoren mit hohen VIF-Faktoren, ebenfalls gestrichen wurde. Interessant ist, dass die NPP ohne die beiden ersteren Faktoren offenbar keinen signifikanten Einfluss mehr auf die Sprachendichte hat. Dies kann wahrscheinlich mit einer *net suppression* erklärt werden (Wuensch 2012). Das Modell erklärt zwei Prozentpunkte weniger der Varianz der lokalen Dichte, was nicht viel ist, da immerhin drei Faktoren gestrichen wurden. Die Höhendifferenz, die Temperatur und der Niederschlag haben den grössten Einfluss auf die Sprachendichte.

15.12.2 Geographisch Nächster Nachbar

Die Resultate der multiplen Regressionsanalyse mit der GNN-Distanz als abhängiger Variable und den geographischen Faktoren als unabhängigen Variablen sind in Tabelle 8 dargestellt. Damit die GNN-Distanz normalverteilt ist, wird sie log₁₀-transformiert.

Tabelle 8: Multiple Regressionsanalyse mit der GNN-Distanz als abhängige Variable und den geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1.

Geo. Faktor	Transformation	Koeffizienten	Standardisierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		2.052	1.476	< 2e-16	***
Temperatur		-0.014	-0.098	< 2e-16	***
Temperaturschwankungsbereich	log ₁₀	0.468	0.091	< 2e-16	***
Niederschlag	√	-0.003	-0.032	7.20E-04	***
Saisonalität Niederschlag	√	-0.043	-0.096	< 2e-16	***
NPP		0.018	0.024	4.08E-03	**
Wachstumsphase		-0.001	-0.065	3.45E-09	***
Küstendistanz	√	0.003	0.031	2.21E-06	***
Höhenmodell+1	log ₁₀	-0.010	-0.006	3.72E-01	
Höhendifferenz	√	-0.009	-0.151	< 2e-16	***

p-Wert	<2E-16	***
Bestimmtheitsmass R ²	0.309	

Der Anteil der erklärten Streuung an der Gesamtstreuung beträgt 30.9 % und ist somit rund 6 Prozentpunkte (Vergleich der korrigierten Bestimmtheitsmasse) tiefer als bei der lokalen Sprachendichte. Mit den geographischen Faktoren kann bei der Sprachendichte mehr erklärt werden als bei der GNN-Distanz. Wie in Tabelle 8 zu sehen ist, hat die Höhe keinen signifikanten Einfluss auf die GNN-Distanz. Der Koeffizient der NPP erreicht das Signifikanzniveau von 0.001 nicht. Alle anderen Koeffizienten haben ein Signifikanzniveau kleiner als 0.001. Wenn wiederum das gleiche Modell noch mit Einbezug der geographischen Breite gerechnet wird, wie schon bei der lokalen Dichte, ergibt sich das exakt gleiche Bestimmtheitsmass. Das heisst, auch die Korrelation der GNN-Distanz mit der geographischen Breite kann vollständig durch die geographischen Faktoren erklärt werden. Wird im Modell noch die Subsistenzstrategie ergänzt, steigt das korrigierte Bestimmtheitsmass lediglich um 1.5 Prozentpunkte.

In Tabelle 9 sind die Resultate einer multiplen Regressionsanalyse ohne die Faktoren Temperaturschwankungsbereich und Länge der Wachstumsphase dargestellt. Wie bereits erwähnt, korrelieren diese stark mit anderen Faktoren und weisen hohe VIF-Werte auf. Eine schrittweise Überprüfung der AIC-Werte hat ergeben, dass das beste Modell erreicht wird, wenn alle anderen Faktoren in das Modell einbezogen werden. Aus den Resultaten geht hervor, dass wiederum die Höhendifferenz, die Temperatur und der Niederschlag den grössten Einfluss auf die GNN-Distanz haben. Die NPP und das Höhenmodell erreichen das Signifikanzniveau von 0.001 nicht.

Tabelle 9: Multiple Regressionsanalyse mit der logarithmierten GNN-Distanz als abhängige Variable und sieben geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 ". "<0.1 " "<1.

Geo. Faktor	Trans-formation	Koeffizienten	Standardisierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		2.723	1.476	< 2e-16	***
Temperatur		-0.021	-0.142	< 2e-16	***
Niederschlag	√	-0.024	-0.089	< 2e-16	***
Saisonalität Niederschlag	√	-0.007	-0.054	< 2e-16	***
NPP		-0.010	-0.013	0.0837	.
Küstendistanz	√	0.004	0.045	2.68E-13	***
Höhenmodell+1	log10	-0.023	-0.014	0.0383	*
Höhendifferenz	√	-0.009	-0.151	< 2e-16	***

p-Wert	<2E-16	***
Bestimmtheitsmass R ²	0.294	

15.13 MEHREBENENANALYSE

In diesem Kapitel soll mithilfe eines gemischten Modells herausgefunden werden, ob die Sprachfamilie einen Einfluss darauf hat, auf welchem Niveau geographische Faktoren Einfluss auf die Sprachendiversität haben. Dazu werden wie schon im Unterkapitel 15.12 die geographischen Faktoren Temperaturschwankungsbereich und Wachstumsphase nicht mit einbezogen, da das Modell sonst eine zu starke Multikollinearität aufweist. Wie bereits am Anfang dieses Teils der Arbeit erwähnt wurde, wird die Sprachfamilie als Gruppe dieses Modells verwendet. Um zu überprüfen, inwiefern die von Currie & Mace (2012) erzielten Resultate, die zeigten, dass die Subsistenzstrategie entscheidend ist, wenn es um die Einflüsse der einzelnen geographischen Faktoren geht, stimmen, sind die gemischten Modelle am Ende der beiden folgenden Abschnitte jeweils noch pro Subsistenztyp aufgeführt. Dargestellt sind jeweils nur die Typen Landwirtschaft und Jäger und Sammler, die Typen Hirten und sesshafte Jäger und Sammler sind im Anhang zu finden. Da diese nur sehr wenige Sprachen betreffen und nur in sehr wenigen Gebieten vorkommen, werden deren Zusammenhänge mit den Umweltfaktoren nicht genauer betrachtet, weil sie zu stark durch die lokalen Gegebenheiten beeinflusst sind.

15.13.1 Lokale Sprachendichte

Entsprechend der Sprachfamilien-Gruppierung gibt es im gemischten Modell 421 Gruppen. Die Resultate der Analyse sind in Tabelle 10 zu finden. Eine schrittweise Überprüfung der AIC-Werte hat ergeben, dass das beste Modell ohne die NPP und ohne das Höhenmodell erreicht wird. Diese beiden Faktoren beeinflussen die lokale Sprachendichte nicht signifikant. Die Höhendifferenz gefolgt vom Niederschlag hat den grössten Einfluss. Die Temperatur hat weniger Einfluss, wenn die Sprachfamilien-Gruppen berücksichtigt werden.

Tabelle 10: Mehrebenenanalyse der lokalen Sprachendichte mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1.

Geo. Faktor	Trans-formation	Koeffizienten	Standardisierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		-0.656	0.817	0	***
Temperatur		0.015	0.103	0	***
Niederschlag	√	0.013	0.157	0	***
Saisonalität Niederschlag	√	0.037	0.082	0	***
Küstendistanz	√	-0.003	-0.038	0	***
Höhendifferenz	√	0.012	0.200	0	***

Anzahl Sprachen	6684
Anzahl Gruppen	421

In Tabelle 11 sind die Resultate der Mehrebenenanalyse jeweils noch pro Subsistenztyp aufgeführt. Werden die Koeffizienten zwischen dem Modell der Landwirtschaft und jenem der Jäger und Sammler verglichen, zeigen sich Unterschiede bezüglich der Einflussstärke einzelner Faktoren. So hat bei der Landwirtschaft die Höhendifferenz und bei den Jäger und Sammlern die Saisonalität des Niederschlags den stärksten Einfluss. Bei beiden Subsistenztypen erreicht der Zusammenhang zum Höhenmodell und zur NPP das Signifikanzniveau von 0.001 nicht.

Tabelle 11: Mehrebenenanalyse der lokalen Sprachendichte aufgeteilt nach Subsistenztyp Landwirtschaft und Jäger und Sammler. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "***"<0.05 " "<0.1 " "<1.

Geo. Faktor	Landwirtschaft			Jäger und Sammler		
	Koeffizienten	Standardisierte Koeffizienten	Signifikanz	Koeffizienten	Standardisierte Koeffizienten	Signifikanz
(Achsenabschnitt)	-0.736	0.779	***	-0.472	0.383	***
Temperatur	0.015	0.096	***	0.007	0.073	***
Niederschlag	0.013	0.148	***	0.011	0.127	***
Saisonalität Niederschlag	0.036	0.080	***	0.059	0.128	***
NPP	-0.014	-0.018	*	0.005	0.006	
Küstendistanz	-0.003	-0.032	***	-0.005	-0.055	***
Höhenmodell+1	0.014	0.008		-0.062	-0.038	*
Höhendifferenz	0.012	0.206	***	0.005	0.063	***

Anzahl Sprachen	5624			685		
Anzahl Gruppen	188			146		

15.13.2 Geographisch Nächster Nachbar

In Tabelle 12 sind die Resultate der Mehrebenenanalyse mit der GNN-Distanz abgebildet. Eine iterative Überprüfung der AIC-Werte hat ergeben, dass, wie schon bei der Sprachendichte, das beste Modell ohne die NPP und das Höhenmodell erreicht wird. Es gibt somit keinen signifikanten Einfluss der NPP und des Höhenmodells auf die Distanz zur nächsten Nachbarsprache. Wie schon bei der lokalen Sprachendichte ist der Einfluss der Höhendifferenz gefolgt vom Niederschlag und der Temperatur am grössten.

Tabelle 12: Mehrebenenanalyse der GNN-Distanz mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "***"<0.05 " "<0.1 " "<1.

Geo. Faktor	Transformation	Koeffizienten	Standardisierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		2.575	1.547	0	***
Temperatur		-0.015	-0.100	0	***
Niederschlag	√	-0.008	-0.100	0	***
Saisonalität Niederschlag	√	-0.024	-0.053	0	***
Küstendistanz	√	0.004	0.039	0	***
Höhendifferenz	√	-0.008	-0.129	0	***

Anzahl Sprachen	6566
Anzahl Gruppen	421

In Tabelle 13 sind die Resultate von Mehrebenenanalysen mit den Sprachfamilien-Gruppen aufgeteilt nach Subsistenztyp aufgeführt. Während bei der Landwirtschaft der Einfluss der Höhendifferenz deutlich am stärksten ist, hat bei den Jägern und Sammlern hingegen die Temperatur, relativ dicht gefolgt von der Saisonalität des Niederschlags, den stärksten Einfluss auf die GNN-Distanz. Die NPP und die Höhe über Meeresspiegel haben beide keinen signifikanten Einfluss auf die Sprachendiversität der Sprachen mit Subsistenztyp Landwirtschaft. Die Unterschiede zwischen den beiden Subsistenztypen sind grösser als bei der lokalen Sprachendichte.

Tabelle 13: Mehrebenenanalyse der lokalen Sprachendichte aufgeteilt nach Subsistenztyp. Die Zeichen für die Signifikanz sind: "****" <0.001 "***"<0.01 "**"<0.05 "."<0.1 ""<1.

Geo. Faktor	Landwirtschaft			Jäger und Sammler		
	Koeffizienten	Standardisierte Koeffizienten	Signifikanz	Koeffizienten	Standardisierte Koeffizienten	Signifikanz
(Achsenabschnitt)	2.509	1.536	***	2.639	1.843	***
Temperatur	-0.013	-0.080	***	-0.013	-0.128	***
Niederschlag	-0.007	-0.087	***	-0.004	-0.050	*
Saisonalität Niederschlag	-0.023	-0.051	***	-0.051	-0.112	***
NPP	-0.001	-0.002		-0.057	-0.066	**
Küstendistanz	0.003	0.030	***	0.003	0.031	.
Höhenmodell+1	0.008	0.005		0.078	0.048	**
Höhendifferenz	-0.008	-0.131	***	-0.007	-0.086	***

Anzahl Sprachen	5514			679		
Anzahl Gruppen	188			146		

15.14 MODELLVERGLEICH

Die verschiedenen Modelle wurden mithilfe eines ANOVA-Tests miteinander verglichen. Für diesen Vergleich wurden alle geographischen Faktoren in die Modelle einbezogen. Miteinander verglichen wurden das multiple Regressionsmodell und das gemischte Modell mit den Sprachfamilien als Gruppierungskriterium. Zudem wurden die Modelle noch unterschieden, je nachdem ob die Subsistenzstrategie auch in das Modell einbezogen wurde oder nicht.

15.14.1 Lokale Sprachendichte

Es zeigte sich, dass alle Modelle der lokalen Sprachendichte signifikant (p -Wert < 0.001) unterschiedlich voneinander sind. Wie in Tabelle 14 zu erkennen ist, hat von diesen Modellen das gemischte Modell mit den Sprachfamilien-Gruppen inklusive dem Subsistenzfaktor den tiefsten AIC-Wert. Dieses Modell beschreibt die Daten somit am besten. Das bedeutet, dass die Zugehörigkeit zu einer bestimmten Sprachfamilie *per se* das Niveau der Sprachendichte beeinflusst und die Subsistenzstrategie mit den geographischen Faktoren interagiert.

Tabelle 14: Vergleich verschiedener Modelle zur lokalen Sprachendichte: Zwei multiple Regressionsmodelle und zwei Mehrebenenanalysen jeweils mit und ohne Subsistenzfaktor als Interaktion.

AIC	Multiples Regressionsmodell		Mehrebenenanalyse	
	Multiples Regressionsmodell	Multiples Regressionsmodell mit Faktor Subsistenzstrategie	Sprachfamilie	Sprachfamilie, Faktor Subsistenzstrategie
	8210.820	7858.585	6118.895	5972.575

15.14.2 Geographisch Nächster Nachbar

Auch für die GNN-Distanz zeigte sich, dass alle Modelle signifikant unterschiedlich voneinander sind, was bedeutet, dass ein Unterschied resultiert, wenn zusätzlich zufällige Effekte durch die Sprachfamilie einbezogen werden. Wie in Tabelle 15 anhand der AIC-Werte gesehen werden kann und mittels einem ANOVA-Test ermittelt wurde, ist wiederum das Modell mit den Sprachfamilien-Gruppen inklusive dem Faktor Subsistenzstrategie jenes, das die Daten am besten beschreibt.

Tabelle 15: Vergleich verschiedener Modelle zur lokalen Sprachdichte: Zwei multiple Regressionsmodelle und zwei Mehrebenenanalysen jeweils mit und ohne Subsistenzfaktor als Interaktion.

AIC	Multiples Regressionsmodell		Mehrebenenanalyse	
	Multiples Regressionsmodell	Multiples Regressionsmodell mit Faktor Subsistenzstrategie	Sprachfamilie	Sprachfamilie, Faktor Subsistenzstrategie
	5529.352	5440.358	4879.606	4867.157

15.15 VERGLEICH DER SPRACHENDIVERSITÄTS-MASSE

Sowohl die lokale Sprachendichte, wie auch die Distanz zum Geographisch Nächsten Nachbarn, sollen die Sprachendiversität beschreiben. Deshalb werden hier die beiden Masse miteinander verglichen, um zu sehen wie stark sie voneinander abweichen. In Abbildung 55 ist ein Vergleich zwischen der GNN-Distanz und der Anzahl an Nachbarsprachen dargestellt.

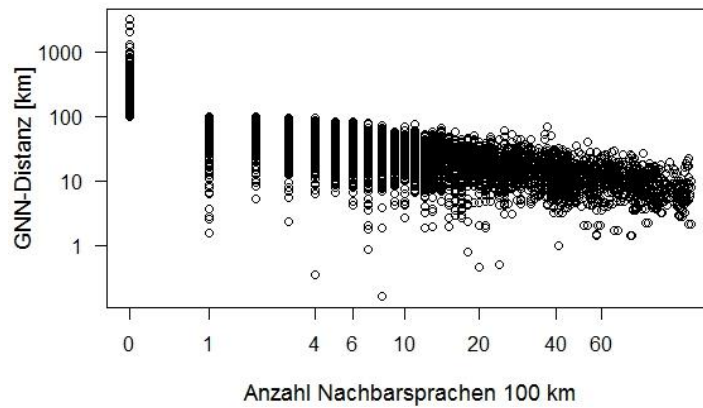


Abbildung 55: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Anzahl an Nachbarsprachen in logarithmierter Darstellung.

Was in Abbildung 55 sofort auffällt, ist, dass bei einem Wert von null Nachbarsprachen die GNN-Distanz nur Werte grösser 100 km liefert. Dies ist darauf zurückzuführen, dass mit der Methode der lokalen Sprachendichte nur Nachbarsprachen bis in maximal 100 km Entfernung erfasst werden. Die Methode mit der GNN-Distanz liefert hingegen auch Werte für Sprachen deren nächste Nachbarsprache weiter als 100 km entfernt ist. Ansonsten nimmt die GNN-Distanz mit zunehmender Anzahl an Nachbarsprachen ab. Der Rangkorrelationskoeffizient beträgt -0.83. Weiter ist zu sehen, dass es fast bei allen Sprachendichten relativ tiefe GNN-Distanzen gibt. Dies ist darauf zurückzuführen, dass, auch wenn die Dichte allgemein tief ist, der nächste Nachbar trotzdem sehr nahe sein kann.

In Tabelle 16 ist ein Vergleich der Mehrebenenanalyse der beiden Sprachendiversitätsmasse dargestellt. Da das Modell der GNN-Distanz weniger Beobachtungen hat, wurden die betreffenden Sprachen für den Vergleich im Modell der lokalen Sprachendichte ebenfalls weggelassen. Weiterhin ist zu sehen, dass die grössten Unterschiede zwischen dem Einfluss des Niederschlags und dem Einfluss der Höhendifferenz bestehen. Beide haben einen stärkeren Einfluss auf die lokale Sprachendichte als auf die GNN-Distanz. Dabei muss natürlich berücksichtigt werden, dass die geographischen Faktoren, wie in Unterkapitel 15.12 gezeigt wurde, insgesamt mehr der Streuung der Sprachendichte (37.3 %) erklären konnten als der GNN-Distanz (30.9 %).

Tabelle 16: Vergleich der Mehrebenenanalysen der lokalen Sprachendichte und GNN-Distanz. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 " " <0.1 " " <1.

Geo. Faktor	Mehrebenenanalyse – Sprachfamilien					
	Lokale Sprachendichte			GNN-Distanz		
	Koeffizienten	Standard-isierte Koeffizienten	Signifi-kanz	Koeffizienten	Standard-isierte Koeffizienten	Signifi-kanz
(Achsenabschnitt)	-0.649	0.813	***	2.552	1.547	***
Temperatur	0.015	0.104	***	-0.014	-0.094	***
Niederschlag	0.013	0.162	***	-0.008	-0.095	***
Saisonalität Niederschlag	0.036	0.080	***	-0.025	-0.057	***
NPP	-0.005	-0.006		-0.009	-0.012	
Küstendistanz	-0.004	-0.039	***	0.003	0.034	***
Höhenmodell+1	-0.003	-0.002		0.016	0.010	
Höhendifferenz	0.012	0.202	***	-0.008	-0.132	***
Anzahl Sprachen	6566			6566		
Anzahl Gruppen	421			421		

15.16 INTERPRETATION

15.16.1 Masse für Sprachendiversität

Bevor die Resultate interpretiert werden, soll zuerst noch allgemein auf die lokale Sprachendichte und die GNN-Distanz als Masse für die Sprachendiversität eingegangen werden. Das Mass der lokalen Sprachendichte bringt einige Nachteile mit sich, wie im vorherigen Unterkapitel gezeigt werden konnte. Zwei Hauptprobleme stellen sich bei diesem Mass der Diversität: erstens die Grösse des Radius und zweitens die Cluster-Effekte. Mit 100 km wurde ein Radius verwendet, der einer Distanz entspricht in der sich Sprecher oder auch Gruppen einer Sprache im Verlauf ihres Lebens bewegen und somit Kontakt mit anderen Sprachen haben können. Diese 100 km scheinen global gesehen ein guter Kompromiss zu sein. Das Problem ist jedoch, dass dieser Wert für gewisse Gesellschaften, wie zum Beispiel Nomaden, zu klein ist und für andere, die beispielsweise im Amazonas-Becken leben, zu gross. Das ist auch der Grund dafür, dass über 821 Sprachen keine einzige Nachbarsprache innerhalb eines Umkreises mit dem Radius 100 km haben. Diese Sprachen wurden auch in die Analysen miteinbezogen, aber im Grunde sind sie nicht sehr aussagekräftig, weil sowohl eine Sprache, die mit einem Radius von 200 km 10 Nachbarsprachen hätte, wie auch eine Sprache die dann immer noch keine hätte, bei 100 km null Nachbarsprachen aufweisen kann. Bei der Kerndichteschätzung (*Kernel Density Estimation*) ist die Schwierigkeit der Wahl eines geeigneten Radius ein bekanntes Problem (O'Sullivan & Unwin 2010). Das zweite Problem, dass sich bei der lokalen Sprachendichte stellt, ist der Cluster-Effekt. Wie in Abschnitt 9.1.2 bei der Häufigkeitsverteilung der lokalen Sprachendichte mit Radius 400 km gesehen werden konnte, gibt es Regionen mit besonders vielen Sprachen, die bei grösserem Nachbarschaftsradius entsprechend auch deutlich mehr Nachbarn haben, während in Regionen mit wenig Sprachen die Anzahl an Nachbarn mit grösser werdendem Radius nur sehr wenig steigt. So erhalten Sprachen in Clustern mehr Gewicht.

Die GNN-Distanz ist insofern eine bessere Methode als die lokale Sprachendichte, da sie für jede Sprache einen differenzierten Wert liefert, im Gegensatz zu den vielen Sprachen mit keinem Nachbarn innerhalb von 100 km. Dennoch hat die GNN-Distanz als Mass der Sprachendiversität auch einige Nachteile. So kann auch in einer Region, in der es sehr wenige Sprachen hat, eine tiefe GNN-Distanz resultieren, wenn zwei Sprachen nahe beieinander liegen. Eine solche Sprache hat dann möglicherweise einen ähnlichen Wert, wie eine andere, die in einer Region mit einer hohen Sprachendiversität vorkommt.

Die Beträge der Korrelationskoeffizienten der verschiedenen geographischen Faktoren mit den beiden Sprachendiversitätsmasse unterscheiden sich nicht stark. Dennoch kann deutlich mehr (6 Prozentpunkte) der Varianz der lokalen Sprachendichte mit geographischen Faktoren erklärt werden.

15.16.2 Einordnung der Resultate

Die Resultate der Analyse des Zusammenhangs zwischen der Sprachendiversität und den geographischen Faktoren als Prädiktoren haben gezeigt, dass diese die Diversität bis zu einem gewissen Anteil erklären können. Am Anfang dieser Arbeit in Kapitel 4 *Forschungsstand* wurden verschiedene bestehende Hypothesen vorgestellt, welche verschiedene Einflüsse auf die Sprachendiversität erläutern. Nun werden diese nochmals herangezogen um sie mit den erzielten Resultaten neu zu beurteilen. Dazu werden zuerst die Resultate der multiplen Regressionsanalysen und die Mehrebenenanalysen ohne den Effekt der Subsistenzstrategie herangezogen. Auf die Unterschiede der verschiedenen Subsistenzstrategien wird danach noch im Speziellen eingegangen.

Vor der Besprechung der verschiedenen Hypothesen wird noch ein weiteres Resultat aufgezeigt, welches in dieser Arbeit diskutiert wurde, da es von verschiedenen Forschern immer wieder thematisiert wurde (Cashdan 2001, Collard & Foley 2002, Mace & Pagel 1994, Nettle 1998, Currie & Mace 2009). Dabei handelt es sich um den *Breitengrad-Gradienten der Sprachendiversität*. Dieser konnte vollständig durch die klimatisch bedingten geographischen Faktoren (die ebenfalls von der Breitenlage abhängig sind)

erklärt werden. Zudem stimmen die bisherigen Beobachtungen nicht ganz, wonach die Sprachendiversität zum Äquator hin zunimmt. In mehreren Abbildungen (Abbildung 33-Abbildung 36) konnte gezeigt werden, dass diese nur bis kurz davor zunimmt. Um den Äquator sind die Sprachendichten tiefer.

Die *topographische Hypothese* besagt, dass in bergigen Regionen eine höhere Sprachendiversität erwartet wird, da sich durch die Topographie Gesellschaften eher separieren (Currie & Mace 2012). Einerseits zeigte sich in dieser Arbeit, dass kaum ein Zusammenhang zwischen der Höhe, auf der eine Sprache liegt, und der Sprachendiversität besteht. Andererseits war zu erkennen, dass je grösser die Höhendifferenz in einem Umkreis von 50 km um eine Sprache ist, desto grösser ist die Sprachendiversität. Sowohl in den multiplen Regressionsanalysen wie auch in der Mehrebenenanalyse war die Höhendifferenz stets der stärkste Einflussfaktor auf die Sprachendiversität. Diese Resultate bestätigen die topographische Hypothese, die nicht allein aufgrund der Höhe eine erhöhte Diversität erwarten würde, sondern aufgrund der topographischen Heterogenität, die in dieser Arbeit mit der Höhendifferenz beschrieben wurde. Interessant ist, dass Currie & Mace (2009) einen positiven Zusammenhang zwischen der Gebietsfläche einer Sprache und der Standardabweichung der Höhe über dasselbe Gebiet feststellen konnten. Da sie die Fläche, über die eine Sprache verbreitet ist, als indirektes Mass für die Sprachendiversität verwendeten, heisst das, dass die Sprachendiversität mit der topographischen Diversität sinkt. Sie kamen in ihrer Arbeit zu einem gegenteiligen Ergebnis und konnten die topographische Hypothese nicht bestätigen. Auch wenn es längst nicht der einzige Faktor ist, der einen Einfluss auf die Sprachendiversität hat, so ist die topographische Heterogenität doch ein wichtiger Einflussfaktor, wie es schon Cashdan (2001) für die ethnische Diversität und Stepp et al. (2005) mit visuellen Mitteln zeigten. Die Höhendifferenz alleine kann rund 13 % (lokale Sprachendichte) respektive 8 % (GNN-Distanz) erklären.

Die nächste Hypothese, die *Umweltrisikohypothese*, besagt, dass die Sprachendiversität vom Risiko bestimmt wird, nicht genügend Existenzgrundlagen zu haben (Currie & Mace 2012). In der bisherigen Forschung wurde als Mass für das Risiko oft die Länge der Anbausaison verwendet (Nettle 1998, Currie & Mace 2009). Auch in der vorliegenden Arbeit wurde deshalb der Zusammenhang zwischen der Sprachendiversität und der Länge der Wachstumsphase analysiert. Sowohl für die lokale Sprachendichte wie auch die GNN-Distanz konnte ein signifikanter Zusammenhang zur Länge der Wachstumsphase nachgewiesen werden. Bei beiden zeigte sich: je länger diese ist, desto höher ist die Sprachendiversität. Dieselben Resultate zeigten sich auch bei Nettle (1998) und Currie & Mace (2009), wobei beide die Länge der Anbausaison in Monaten und nicht wie hier in Tagen berechnet hatten. Das Problem mit diesem geographischen Faktor ist jedoch, dass er stark mit den Klima-Faktoren korreliert, da er auf ihrer Grundlage berechnet wird. Es kann deshalb nicht gesagt werden, ob die Länge der Anbauphasen direkt einen Einfluss auf die Sprachendiversität hat oder ob es doch eher einzelne Klima-Faktoren sind, die entscheidend darauf einwirken. Als weitere Indizien für das ökologische Risiko werden in dieser Arbeit die Faktoren Temperaturschwankungsbereich und die Saisonalität des Niederschlags herbeigezogen. Bei beiden zeigt sich, dass je grösser die Variation ist, desto kleiner ist die Sprachendiversität. Im Mehrebenenmodell hatte die Saisonalität des Niederschlags deutlich weniger Einfluss als andere Klima-Faktoren. Sowohl bei der lokalen Sprachendichte wie auch bei der GNN-Distanz erklärt der Temperaturschwankungsbereich allein mehr als 15 % von deren Streuung. Somit spricht auch dieser Faktor für die Umweltrisikohypothese. In einem ähnlichen Zusammenhang konnte Cashdan (2001) zeigen, dass in Regionen mit variablen klimatischen Bedingungen eine tiefe ethnische Diversität zu finden ist. Ein weiterer Faktor, der das ökologische Risiko beschreibt, ist die Küstendistanz. Erstens sind in Küstennähe extreme Klimata reduziert und zweitens steht dort während des gesamten Jahres proteinreiche Nahrung zu Verfügung (Nichols 1990). Beides sind wichtige Faktoren die das ökologische Risiko senken. In dieser Arbeit konnte gezeigt werden, dass isoliert betrachtet ein Zusammenhang zwischen der Sprachendiversität und der Küstendistanz besteht. Auch im Mehrebenenmodell besteht ein signifikanter Einfluss auf die Sprachendiversität, jedoch ist es der am wenigsten starke Einfluss, der noch signifikant ist. Deshalb kann durch diese Resultate zwar bestätigt werden, dass die Distanz zur Küste einen Einfluss auf die Sprachendiversität hat, jedoch ein nicht sehr grosser. Womöglich ist die Küstendistanz

kein geeigneter Faktor, um das ökologische Risiko zu beschreiben. Die Auswertung der geographischen Faktoren, die das ökologische Risiko beschreiben, unterstützen die Umweltrisikohypothese, wobei auch hier zu beachten ist, dass nicht dieses Risiko allein über die Sprachendiversität entscheidet. Die einzelnen Faktoren korrelieren zum Teil auch stark miteinander und erklären so in einem multiplen Regressionsmodell ähnliche Anteile der Streuung. Immerhin sind allein die vier Faktoren fähig darin 24 % (lokale Sprachendichte) respektive 22.2 % (GNN-Distanz) der Streuung zu erklären.

Abschliessend entsprechende Ausführungen zur *Umweltproduktivität-Hypothese*. Diese besagt, dass je grösser die Produktivität der Umwelt ist, desto höher ist die Sprachendiversität (Currie & Mace 2012). Diese Hypothese soll mithilfe der NPP, dem Niederschlag und der Temperatur überprüft werden. Bei allen drei Faktoren konnte dahingehend ein signifikanter Zusammenhang festgestellt werden, dass je höher deren Werte sind, desto grösser ist auch die Sprachendiversität. Diese drei Faktoren zusammen können rund 20 % der Streuung der Sprachendiversität erklären. Im gemischten Modell mit den Sprachfamilien-Gruppen hatten jedoch nur die Temperatur und der Niederschlag einen signifikanten Einfluss auf die beiden Masse. Der Einfluss der NPP erreichte, nachdem die Multikollinearität des Modells reduziert wurde, das Signifikanzniveau von 0.001 nicht. Interessant ist, dass Currie & Mace (2012) erst dann einen signifikanten Zusammenhang zwischen der NPP und der Sprachendiversität feststellen konnten als sie die Sprachen nach Subsistenzstrategien aufteilten. Auch in dieser Arbeit zeigte sich, dass einzig bei den Sprachen deren Sprecher Jäger und Sammler sind, die NPP einen signifikant (p -Wert < 0.01) positiven Einfluss auf die Sprachendiversität hatte. Das Thema Subsistenzstrategie wird später in diesem Abschnitt noch weiter ausgeführt. Dass der Niederschlag einen positiven Einfluss auf die Sprachendiversität respektive auch auf die ethnische Diversität hat, wurde schon von mehreren Autoren festgestellt (Birdsell 1953, Cashdan 2001, Collard & Foley 2002). Auch die Umweltproduktivität-Hypothese kann durch die erzielten Resultate unterstützt werden. Jedoch muss auch hier festgehalten werden, dass zwar ein Zusammenhang nachgewiesen werden konnte aber, dass auch 80 % der Streuung durch geographische Faktoren, welche die Produktivität beschreiben, nicht erklärt werden können.

Mit den Resultaten dieser Arbeit kann keine der drei Hypothesen widerlegt werden. Die Einflüsse, welche die geographischen Faktoren laut den Hypothesen auf die Sprachendiversität haben, konnten bestätigt werden. Was jedoch mit Sicherheit gesagt werden kann, ist, dass keine Hypothese allein die Sprachendiversität zu erklären vermag. Wie gezeigt werden konnte, kann mit allen Hypothesen zusammen respektive den entsprechenden geographischen Faktoren, insgesamt ein Drittel (37.3 % und 31 %) der Sprachendiversität erklärt werden. Dies deckt sich gut mit den Beobachtungen die Moore et al. (2002) für Afrika gemacht haben. Sie konnten mit verschiedenen Umweltfaktoren 36 % der Variation der Sprachendiversität erklären. Diese Grössenordnung lässt sich nun auch für die ganze Welt bestätigen. Weltweit kann circa ein Drittel der Streuung der Sprachendiversität mit Umweltfaktoren erklärt werden.

Wird das multiple Regressionsmodell noch um den Faktor Subsistenzstrategie ergänzt, können sogar rund 41 % (lokale Sprachendichte) respektive 32.4 % (GNN-Distanz) erklärt werden. Von Currie & Mace (2009, 2012) wurde mehrmals darauf hingewiesen, dass je nachdem welche Subsistenzstrategie Sprecher einer Sprache verfolgen, andere Umweltfaktoren entscheidend sein könnten. In dieser Arbeit konnte gezeigt werden, dass wenn die Subsistenzstrategie als Faktor mit Interaktion ins Mehrebenenmodell einbezogen wird, sich dieses verbessert. Erklärt wird das damit, dass Sprachen mit unterschiedlichen Subsistenztypen anders durch die geographischen Faktoren beeinflusst werden. Die grössten Unterschiede zwischen Landwirten und Jäger und Sammlern bestehen bezüglich der Einflüsse der Saisonalität des Niederschlags und der Höhendifferenz. Die Verteilung des Niederschlags über das Jahr hat einen höheren Einfluss auf die Sprachendiversität der Jäger-und-Sammler-Gesellschaften, während es sich bei der Höhendifferenz umgekehrt verhält. Ausser bei der NPP und dem Höhenmodell haben alle geographischen Faktoren einen signifikanten Einfluss auf die Sprachendiversität der Landwirtschaft. Bei den Jägern und Sammlern sieht es ähnlich aus. Das Modell, welches die Subsistenzstrategie berücksichtigt, erklärt die Daten besser, was die Hypothese von Currie & Mace (2012) unterstützt, wonach die Einflüsse der Umweltfaktoren je nach Subsistenztyp anders sind. Jedoch kann daraus nicht gefolgert werden, dass je nach Subsistenztyp andere

geographische Faktoren die Sprachendiversität beeinflussen. Die Subsistenztypen Jäger-und-Sammler, sesshafte Jäger-und-Sammler und Hirten konzentrieren sich auf einzelne Regionen, wodurch einzelne geographische Faktoren einen unterschiedlich starken Einfluss haben. Es ist gut möglich, dass je nach Subsistenzstrategie die von Sprechern einer Sprache verfolgt wird, unterschiedliche Faktoren verschieden starke Auswirkungen auf die Sprachenseparation oder auf das soziale Netzwerk haben, jedoch kann dies nicht aus den erzielten Resultaten geschlossen werden.

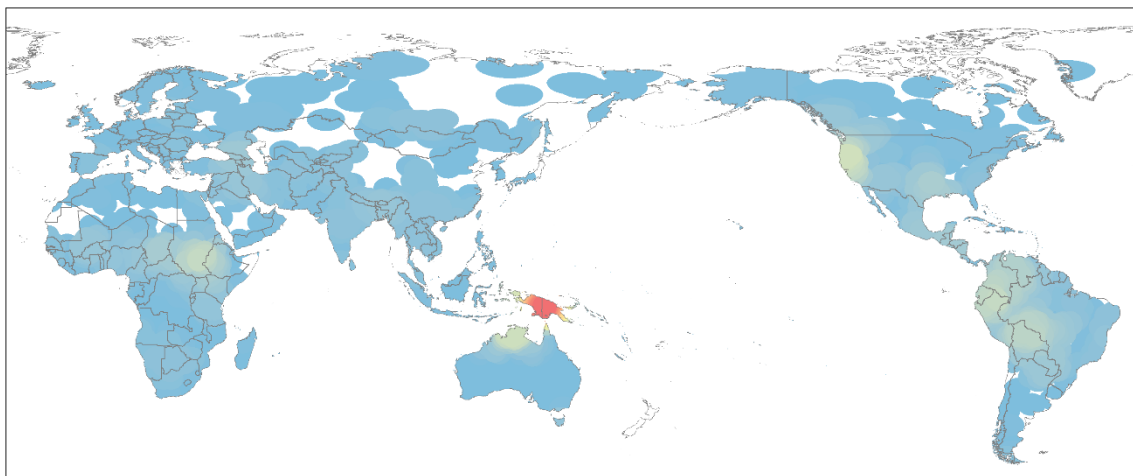
Abschliessend kann zur Sprachendiversität festgehalten werden, dass sie zwar zu einem Drittel durch die geographischen Faktoren erklärt werden kann, aber diese längst nicht die komplette Diversität erklären können. Die restlichen zwei Drittel lassen sich möglicherweise durch historische und kulturelle Faktoren erklären, die ebenfalls zu Sprachenseparation und Sprachendiversifizierung führen.

16 SPRACHFAMILIENDIVERSITÄT

In diesem Kapitel steht die Sprachfamiliendiversität im Zentrum. Zuerst wird der Zusammenhang zwischen der Sprachfamiliendichte respektive der Distanz zum nächsten nicht-verwandten Nachbarn und den geographischen Faktoren einzeln analysiert. Da die Sprachfamiliendiversitäten stark durch die grossen Sprachfamilien geprägt sind, da diese oftmals eine spezielle Ausbreitungsgeschichte besitzen (siehe *Abschnitt 10.1.2*), werden dieselben Auswertungen deshalb auch noch ohne die grossen Sprachfamilien durchgeführt. Oftmals zeigen sich in den Abbildungen Muster, die entweder schon bei der lokalen Sprachendichte oder der GNN-Distanz sichtbar waren. Gehen diese Muster auf dieselben Gründe zurück wie dort, werden diese hier nicht nochmals erläutert. Ist dies der Fall, hat das häufig mit der globalen Verteilung der geographischen Faktoren zu tun und nicht mit dem Sprachenmass. Danach folgen wie schon bei der Sprachendiversität ein multiples Regressionsmodell und eine Mehrebenenanalyse und zum Abschluss eine vertiefte Interpretation.

16.1 GLOBALE VERTEILUNG

Bevor die Zusammenhänge zwischen der Sprachfamiliendiversität und den geographischen Faktoren näher analysiert werden, steht zuerst die globale Verteilung im Fokus. Dies soll einen Vergleich mit der Sprachendiversität und deren Verteilungsmuster auf der Welt möglich machen. Die Verteilung nach Längengraden wurde ebenfalls überprüft, wobei dabei kein Zusammenhang festgestellt werden konnte. In Abbildung 56 sind die 400 km grossen Umkreise zur Berechnung der lokalen Sprachfamiliendichte zu sehen. Auf allen grossen Landmassen hat es Regionen, wo viele Sprachfamilien vorkommen, ausser in Eurasien, wo sich keine Region deutlich hervorhebt.



Eigene Darstellung. Quellen: UNEP (2013), Hammarström (2012)

Abbildung 56: Umkreise mit Radius 400 km um jede Sprache, eingefärbt nach Anzahl Nachbarsprachfamilien. Rot bedeutet eine hohe Dichte, blau eine niedrige Dichte.

16.1.1 Lokale Sprachfamiliendichte

In Abbildung 57 ist die Verteilung der Sprachfamiliendichten auf die geographische Breite zu sehen. Farblich hervorgehoben sind Sprachen, die zu den grössten Sprachfamilien gehören. Auffällig ist der grosse Peak südlich des Äquators, welcher sich hauptsächlich aus Sprachen auf Neuguinea zusammensetzt. Ein weiterer Peak bei 40° Nord ist auf eine Ansammlung vieler Sprachen unterschiedlicher Sprachfamilien an der Westküste Amerikas auszumachen. Bei den lokalen Sprachendichten traten diese Sprachen, im Gegensatz zu jenen auf der Insel Neuguinea, kaum in Erscheinung, da sie keine hohen Sprachendichten aufweisen. Der kleine Peak bei 10° ist auf Sprachen unterschiedlicher Familien im Sudan zurückzuführen. Weitere Hotspots, die in dieser Abbildung nur

schwer zu erkennen, jedoch in Abbildung 56 gut zu sehen sind, befinden sich im Norden Australiens, im Grenzgebiet von Bolivien und Brasilien und im Grenzgebiet von Peru, Ecuador und Kolumbien. Festgestellt werden kann auch, dass die Sprachen der grossen Sprachfamilien Sinotibetisch, Indogermanisch und Niger-Kongo-Sprachen kaum je mehr als fünf Nachbarsprachfamilien aufweisen. In Abbildung 58 wurden die Nord- und Südhalbkugel übereinander gelegt. Der Rangkorrelationskoeffizient der darin enthaltenen Daten beträgt -0.21 und jener ohne die grossen Sprachfamilien -0.51.

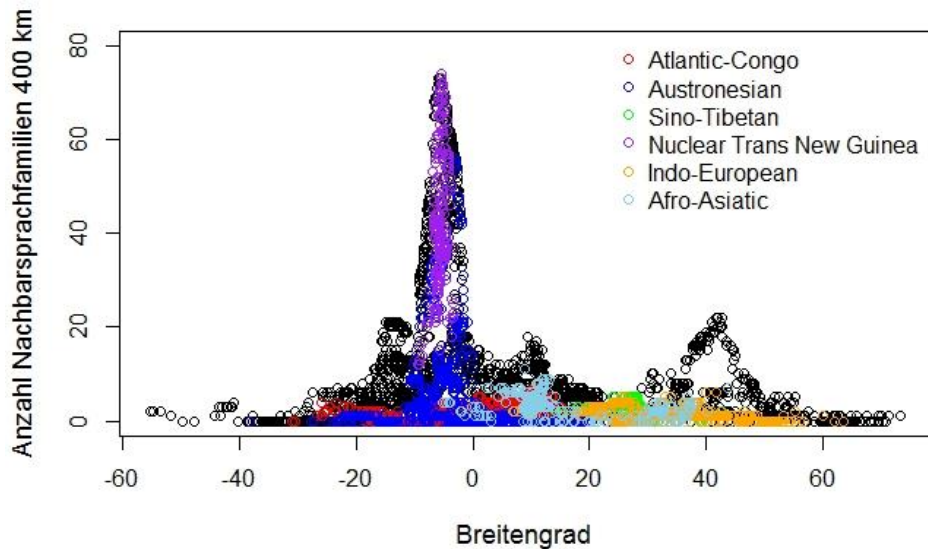


Abbildung 57: Verteilung der lokalen Sprachfamiliendichten über die Breitengrade. Farblich hervorgehoben sind Sprachen der grössten Sprachfamilien.

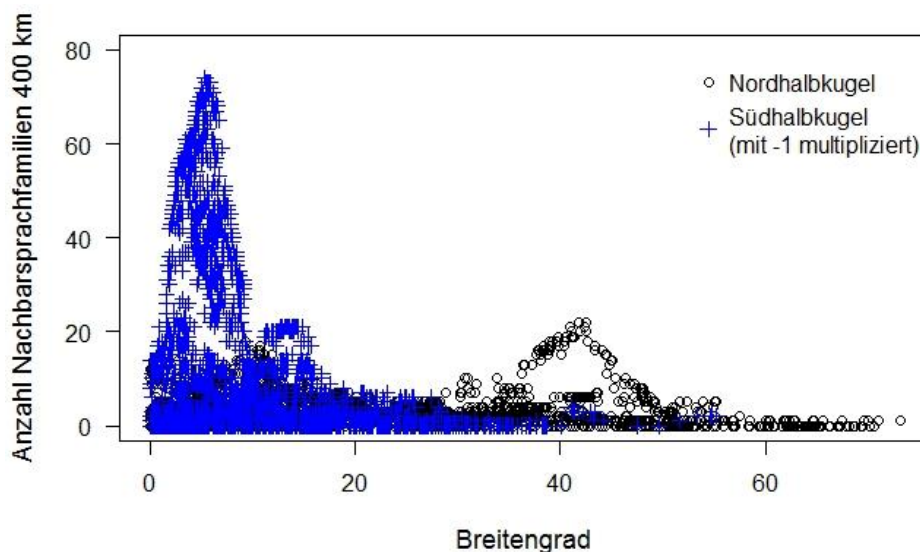


Abbildung 58: Verteilung der lokalen Sprachfamiliendichten über die Breitengrade. In blau die Werte der Südhalbkugel, in schwarze jene der Nordhalbkugel.

16.1.2 Geographisch Nächster nicht-verwandter Nachbar

Die Verteilung der Distanz zum Geographisch Nächsten Nachbarn einer anderen Sprachfamilie auf die Breitengrade ist in Abbildung 59 zu finden. Darin sind deutlich Sprachen zusehen, die zusammen bogenförmige Muster bilden. Diese entstehen dadurch, dass viele Sprachen einer Sprachfamilie denselben GNNnv-Nachbar haben. Oft ist das bei Sprachen der Fall, die sich auf Inseln befinden. Besonders oft kommt dieses Phänomen deshalb bei austronesischen Sprachen vor, was in der Abbildung 59 auch deutlich zu sehen ist. Aber auch Sprachen im südlichen Teil von Afrika, die zu den Niger-Kongo-

Sprachen gehören, haben über ein weites Gebiet hinweg keine Nachbarsprache einer anderen Familie. In Australien befindet sich das dritte Cluster, dieses ist als Bogen zwischen -40 und -20 erkennbar. Dort gibt es über ein weites Gebiet ebenfalls sehr wenige Sprachfamilien. In Abbildung 60 sind die Werte der Nord- und Südhalbkugel übereinandergelegt. Darin ist erkennbar, dass die GNNnv-Distanz mit der geographischen Breite zunimmt. Der Rangkorrelationskoeffizient beträgt 0.14 und 0.42 ohne die grossen Sprachfamilien.

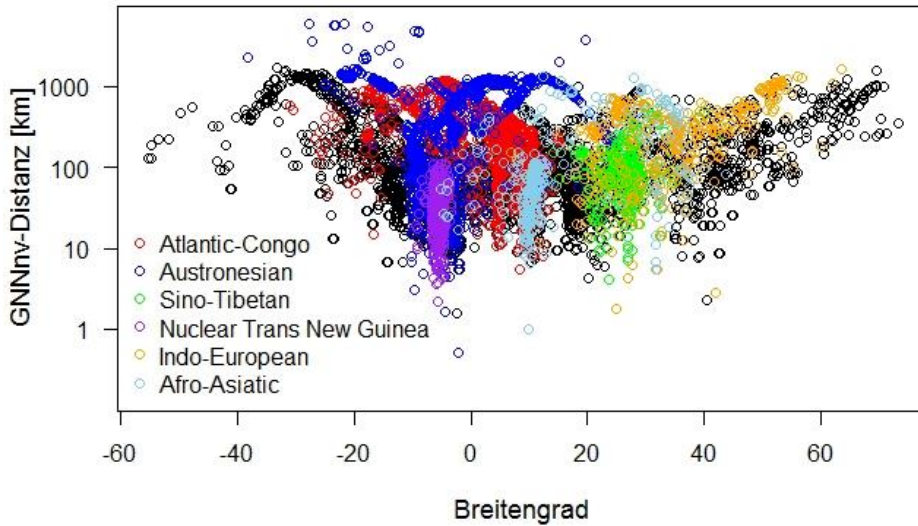


Abbildung 59: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die Breitengrade. Farblich hervorgehoben sind Sprachen der grössten Sprachfamilien.

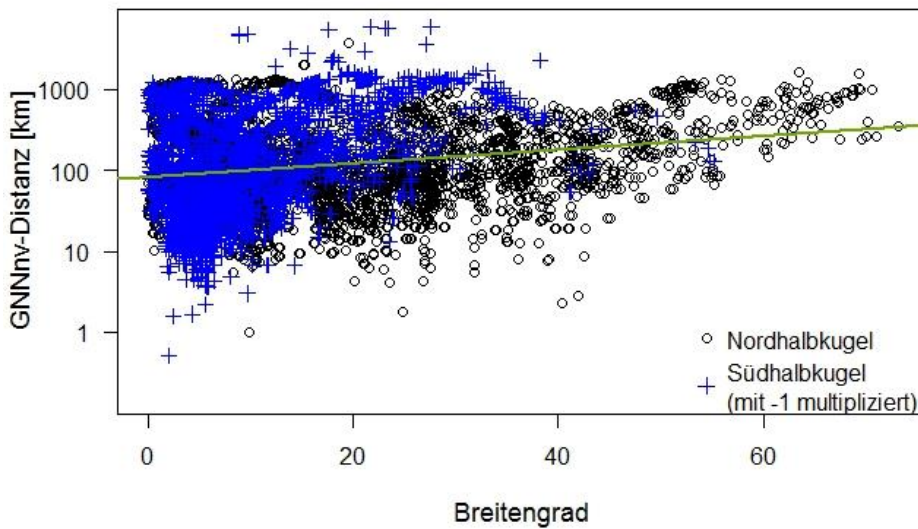


Abbildung 60: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die Breitengrade. In blau die Werte der Südhalbkugel, in schwarze jene der Nordhalbkugel.

16.2 MITTLERE JÄHRLICHE TEMPERATUR

16.2.1 Lokale Sprachfamiliendichte

Wie in Abbildung 61 gezeigt werden kann, treten hohe Sprachfamiliendichten erst ab mittleren jährlichen Temperaturen von 10 °C auf. Wobei es auch bei hohen Temperaturen viele Sprachen hat, die in ihrer Nachbarschaft kaum Sprachen anderer Sprachfamilien aufweisen. Die Korrelation zwischen der lokalen Sprachfamiliendichte und der Temperatur beträgt 0.13 und 0.35 ohne die grossen Sprachfamilien.

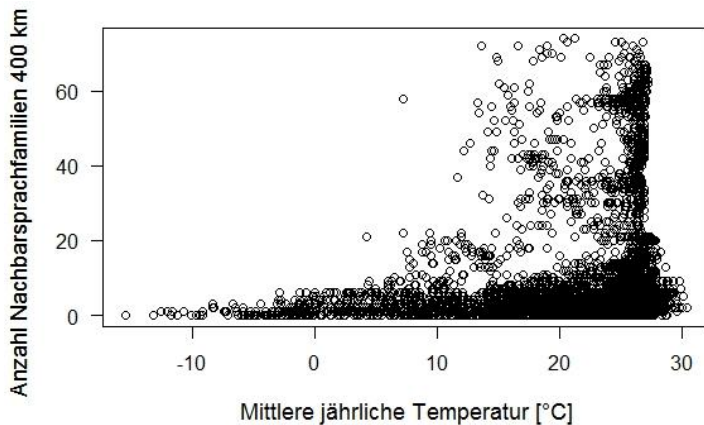


Abbildung 61: Verteilung der lokalen Sprachfamiliendichten über die mittlere jährliche Temperatur.

16.2.2 Geographisch Nächster nicht-verwandter Nachbar

In Abbildung 62 ist erkennbar, dass die Distanz zum Geographisch Nächsten nicht-verwandten Nachbar mit steigender Temperatur abnimmt. Der Rangkorrelationskoeffizient beträgt -0.15 und ist damit nicht viel tiefer als bei der lokalen Sprachfamiliendichte (0.13). Ohne die grossen Sprachfamilien ist der Zusammenhang stärker, denn das *rho* beträgt -0.32. Bei hohen Temperaturen ist zu beobachten, dass sehr viele unterschiedliche Distanzen auftreten.

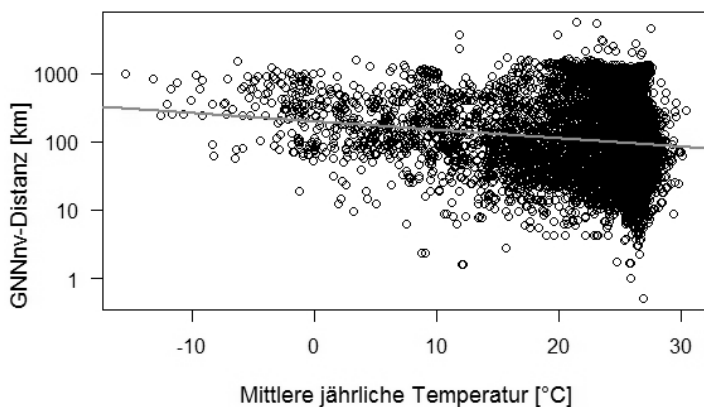


Abbildung 62: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die mittlere jährliche Temperatur.

16.3 JÄHRLICHER TEMPERATURSCHWANKUNGSBEREICH

16.3.1 Lokale Sprachfamiliendichte

In Abbildung 63 ist zu sehen, wie sich die lokalen Sprachfamiliendichten auf die jährlichen Temperaturschwankungsbereiche verteilen. Ähnlich wie schon bei der Sprachendichte lassen sich drei Stufen ausmachen. Die Sprachen mit hohen Sprachfamiliendichten mit einem Schwankungsbereich um 10 °C setzen sich hauptsächlich aus Sprachen auf Neuguinea zusammen. Die hohen Dichten der zweiten Stufen sind vor allem auf die Sprachen im Sudan und im Norden von Australien und ab 30 °C auf die Sprachen an der Westküste Amerikas zurückzuführen. Der Rangkorrelationskoeffizient beträgt -0.16 und zeigt somit einen deutlicheren Zusammenhang an als es bei der jährlichen Temperatur der Fall war. Ohne die grossen Sprachfamilien beträgt er -0.49.

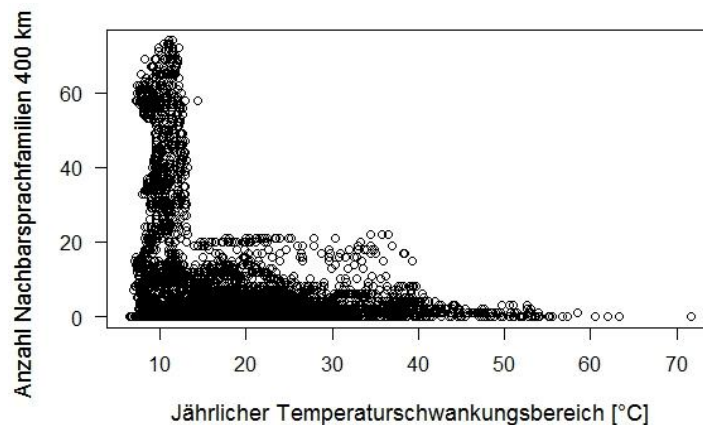


Abbildung 63: Verteilung der lokalen Sprachfamiliendichte über den jährlichen Temperaturschwankungsbereich.

16.3.2 Geographisch Nächster nicht-verwandter Nachbar

In Abbildung 64 ist zu beobachten, wie sich die GNNnv-Distanz auf die jährlichen Temperaturschwankungsbereiche verteilen. Die Anhäufung von Werten bei ca. 1000 km und unter 10 °C ist auf austronesische Sprachen mit demselben Geographisch Nächsten nicht-verwandten Nachbar zurückzuführen. Die Lücke von tiefen Distanzen um 15 °C ergibt sich dadurch, dass sich vor allem Niger-Kongo-Sprachen in diesem Bereich befinden. Diese befinden sich, wie auch schon bei der Verteilung auf die Breitengrade beobachtet werden konnte, in einem Gebiet, wo der nächste nicht-verwandte Nachbar weit entfernt ist. Die Korrelation beträgt 0.06 und ist damit deutlich tiefer als der Betrag des ρ mit 0.16 bei der Sprachfamiliendichte. Ohne die grossen Sprachfamilien ist der Korrelationskoeffizient deutlich höher und beträgt 0.42.

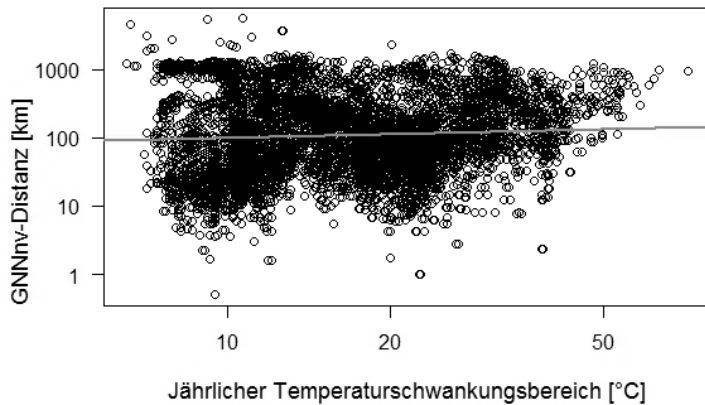


Abbildung 64: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über den jährlichen Temperaturschwankungsbereich in logarithmierter Darstellung.

16.4 JÄHRLICHER NIEDERSCHLAG

16.4.1 Lokale Sprachfamiliendichte

Die hohen Sprachfamiliendichten treten mit zunehmendem jährlichem Niederschlag auf wie in Abbildung 65 zu sehen ist. Regionen mit vielen Sprachfamilien sind eher an Orten mit viel Niederschlag zu finden. Die Korrelation beträgt 0.21 und ist somit gleich stark wie jene mit der geographischen Breite. Ohne grosse Sprachfamilien beträgt das ρ 0.39.

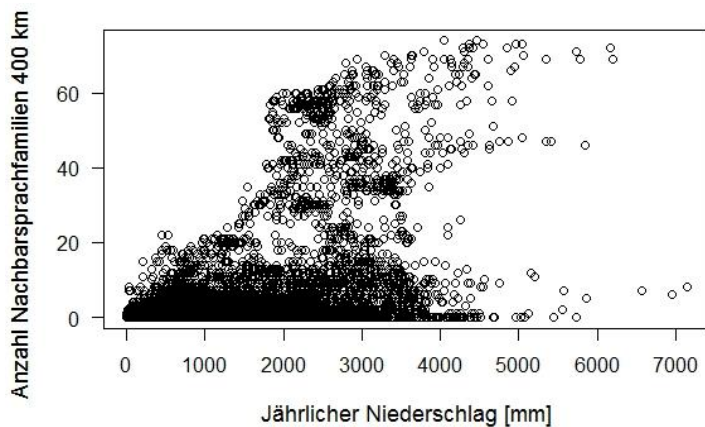


Abbildung 65: Verteilung der lokalen Sprachfamiliendichten über den jährlichen Niederschlag.

16.4.2 Geographisch Nächster nicht-verwandter Nachbar

In Abbildung 66 ist die Verteilung der GNNnv-Distanz auf den jährlichen Niederschlag zu sehen. Wiederum fallen hier einige austronesische Sprachen mit einer Distanz von circa 1'000 km und einem jährlichen Niederschlag bis 4'500 mm auf. Auch wenn sich inklusive dieser Daten ein negativer Zusammenhang zwischen den beiden Variablen erkennen lässt, so ist klar, dass dieser viel stärker wäre ohne diese Daten. Das zeigt sich auch im Rangkorrelationskoeffizient. Dieser beträgt für alle Daten -0.13 und ohne die grossen Sprachfamilien -0.35.

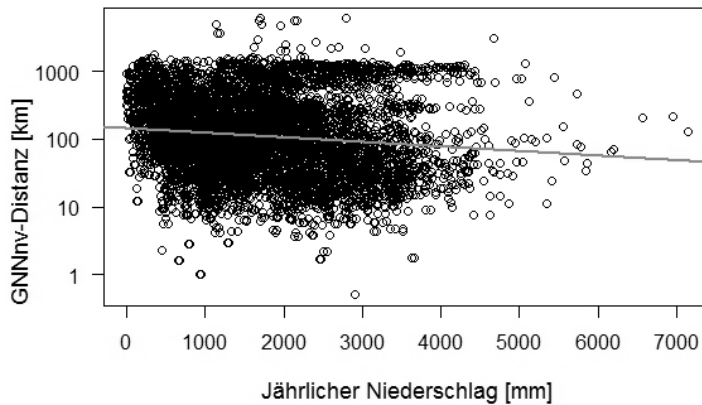


Abbildung 66: Verteilung der GNNv-Distanz in logarithmierter Darstellung über den jährlichen Niederschlag.

16.5 SAISONALITÄT DES NIEDERSCHLAGS

16.5.1 Lokale Sprachfamiliendichte

Abbildung 67 zeigt den Zusammenhang zwischen der Sprachfamiliendichte und der Saisonalität des Niederschlags. Je weniger der Niederschlag während eines Jahres variiert, desto eher treten hohe Sprachfamiliendichten auf. Die Korrelation beträgt -0.1 und ist damit stärker als bei der lokalen Sprachdichte. Ohne die grossen Sprachfamilien ist das ρ -0.29 .

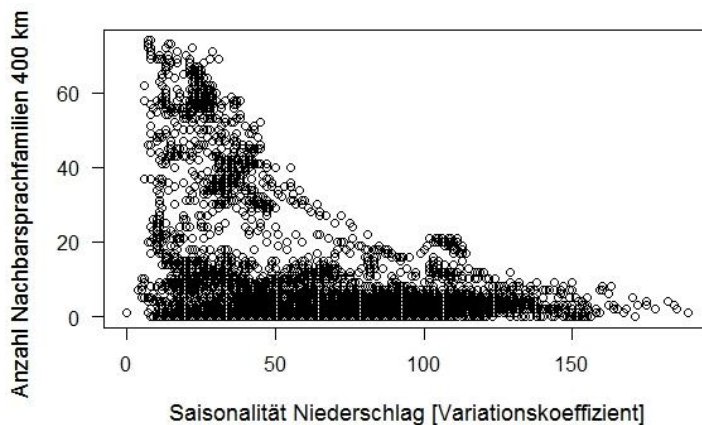


Abbildung 67: Verteilung der lokalen Sprachfamiliendichten über die Saisonalität des Niederschlags.

16.5.2 Geographisch Nächster nicht-verwandter Nachbar

Mit zunehmender Saisonalität des Niederschlags sinkt die GNNv-Distanz nur sehr leicht (Abbildung 68). Bei einer Distanz von 1'000 km ist abermals eine Anhäufung von austronesischen Sprachen, die in der Abbildung jedoch schwer erkennbar ist, zu finden. Der Rangkorrelationskoeffizient beträgt -0.06 und 0.12 ohne die grossen Sprachfamilien.

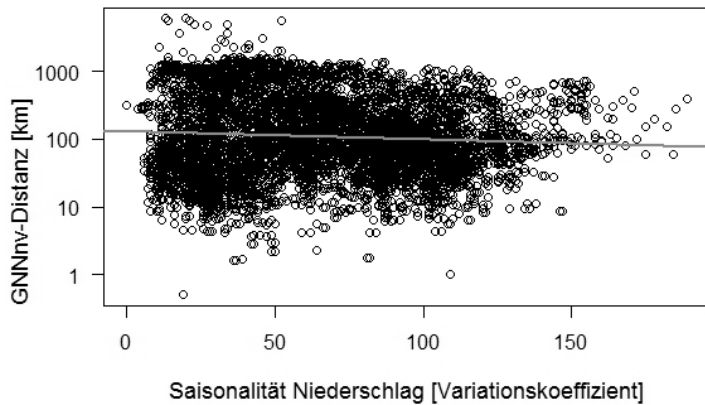


Abbildung 68: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die Saisonalität des Niederschlags.

16.6 NETTOPRIMÄRPRODUKTION

16.6.1 Lokale Sprachfamiliendichte

Der Zusammenhang zwischen der Sprachfamiliendichte und der Nettoprimärproduktion ist in Abbildung 69 zu sehen. Hohe Dichten treten erst bei höherer Nettoprimärproduktion auf. Der Rangkorrelationskoeffizient beträgt 0.16 und 0.31 ohne die grossen Sprachfamilien.

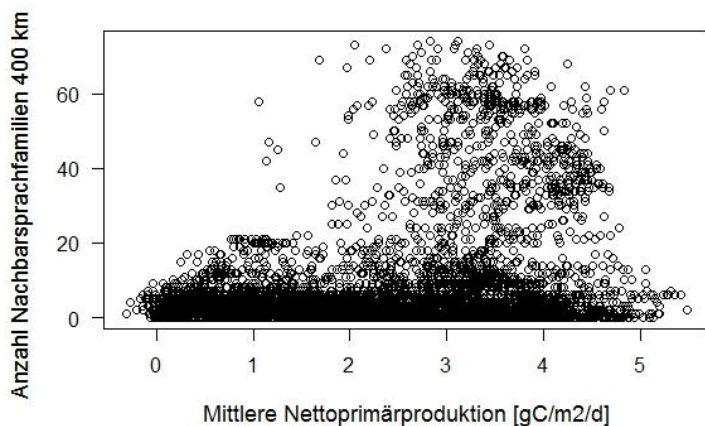


Abbildung 69: Verteilung der lokalen Sprachfamiliendichte über die mittlere Nettoprimärproduktion.

16.6.2 Geographisch Nächster nicht-verwandter Nachbar

In Abbildung 70 ist der Zusammenhang zwischen der GNNnv-Distanz und der mittleren Nettoprimärproduktion abgebildet. Zwischen einem Wert von 2.5-5 $\text{gC}/\text{m}^2/\text{d}$ und einer Distanz von 1'000 km tritt wieder eine Anhäufung von austronesischen Sprachen auf. Der Rangkorrelationskoeffizient beträgt -0.07 mit und -0.3 ohne grosse Sprachfamilien.

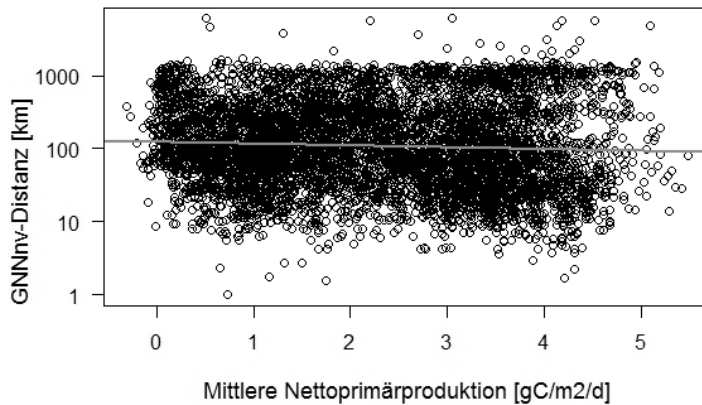


Abbildung 70: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die mittlere Nettoprimärproduktion.

16.7 LÄNGE DER WACHSTUMSPHASE

16.7.1 Lokale Sprachfamiliendichte

Der Zusammenhang zwischen der Sprachfamiliendichte und der Länge der Wachstumsphase pro Jahr ist in Abbildung 71 zu sehen. Die meisten der Sprachen mit vielen Nachbarsprachfamilien liegen in einer Regionen, wo die Anbauphase länger als 364 Tage dauert. Der Korrelationskoeffizient beträgt 0.25 und 0.42 ohne die grossen Sprachfamilien. Demzufolge ist der Zusammenhang in etwa gleich gross wie beim Niederschlag (0.21 und 0.39).

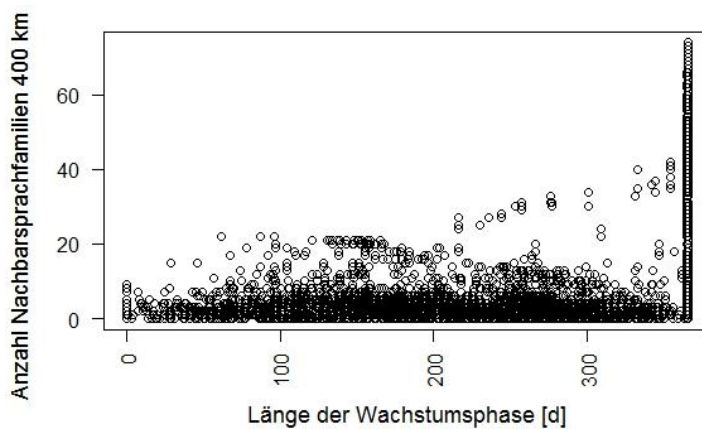


Abbildung 71: Verteilung der lokalen Sprachfamiliendichten über die Länge der Wachstumsphase.

16.7.2 Geographisch Nächster nicht-verwandter Nachbar

Die logarithmierte GNNnv-Distanz sinkt leicht, je länger die Wachstumsphase dauert (Abbildung 72). Der Rangkorrelationskoeffizient beträgt -0.11. Ohne die grossen Sprachfamilien ist der Zusammenhang deutlicher, denn das ρ sinkt auf -0.37.

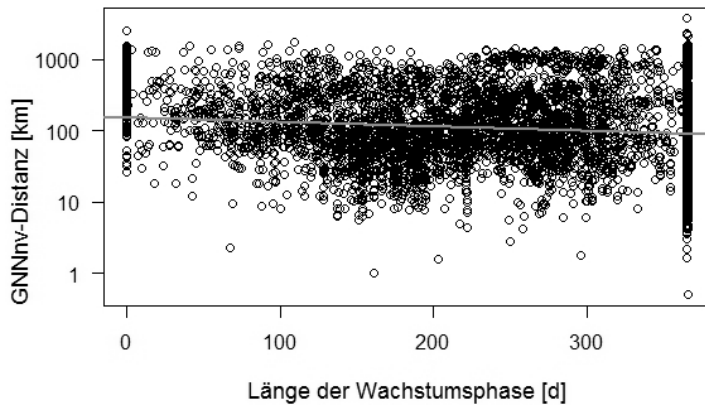


Abbildung 72: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die Länge der Wachstumsphase.

16.8 KÜSTENDISTANZ

16.8.1 Lokale Sprachfamiliendichte

In Abbildung 73 ist die Verteilung der Sprachfamiliendichte auf die Küstendistanz zu sehen. Die Form des auffälligen Peaks, welcher sich aus Sprachen auf Neuguinea zusammensetzt, ist darauf zurückzuführen, dass dort mit einem Radius von 400 km im Landesinnern fast die ganze Inselbreite abgedeckt ist. Somit resultieren dort höhere Werte als an der Küste. Der Rangkorrelationskoeffizient beträgt -0.02 , wobei er nicht signifikant unterschiedlich von Null ist. Ohne die grossen Sprachfamilien resultiert mit -0.11 ein signifikanter Zusammenhang.

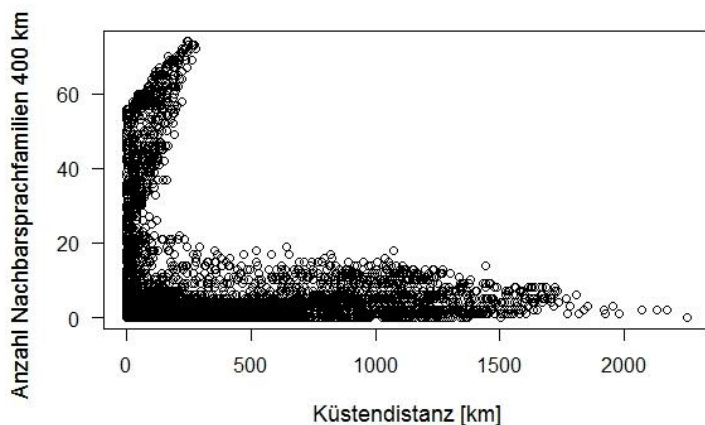


Abbildung 73: Verteilung der lokalen Sprachfamiliendichten über die Küstendistanz.

16.8.2 Geographisch Nächster nicht-verwandter Nachbar

Aus Abbildung 74 lässt sich nur eine leichte Tendenz ausmachen, wonach die GNNnv-Distanz mit zunehmender Entfernung zur Küste abnimmt. Die Rangkorrelation beträgt -0.07 . Ohne grosse Sprachfamilien geht der Zusammenhang in die umgekehrte Richtung, denn das ρ hat dann einen Wert von 0.15 . Demzufolge würde ohne diese Sprachen eine Zunahme der GNNnv-Distanz ins Landesinnere erwartet. Dies würde sich mit den Beobachtungen bei der Sprachendiversität decken, wo der Zusammenhang ebenfalls in diese Richtung geht.

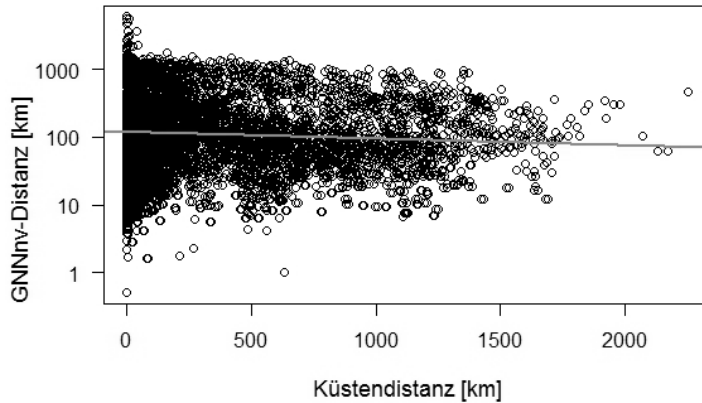


Abbildung 74: Verteilung der GNNnv-Distanz in logarithmierter Darstellung über die Küstendistanz.

16.9 HÖHENMODELL

16.9.1 Lokale Sprachfamiliendichte

In Abbildung 75 ist zu sehen, auf welcher Höhen über Meeresspiegel, welche Sprachfamiliendichten auftreten. Die meisten hohen Dichten treten in tiefen Lagen auf, was nicht verwundert, da in Abschnitt 16.8.1 festgestellt wurde, dass hohe Dichten vor allem in Küstennähe auftreten. Der Rangkorrelationskoeffizient beträgt 0.046 und -0.16 ohne die grossen Sprachfamilien.

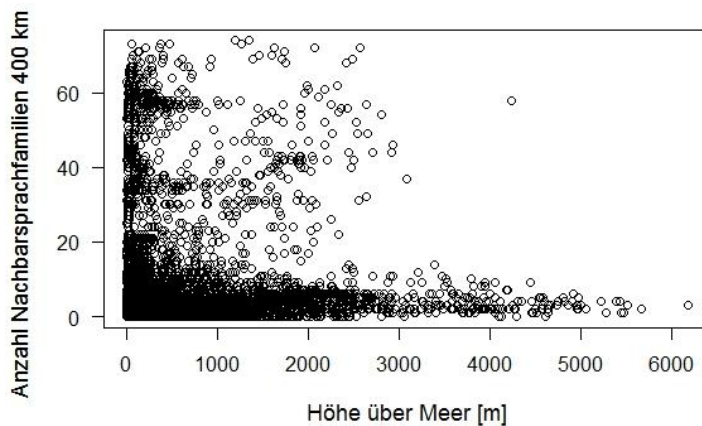


Abbildung 75: Verteilung der lokalen Sprachfamiliendichten über die Höhe über Meer.

16.9.2 Geographisch Nächster nicht-verwandter Nachbar

Wie sich in Abbildung 76 erkennen lässt, nimmt die GNNnv-Distanz mit zunehmender Höhe über Meeresspiegel ab. Der Rangkorrelationskoeffizient beträgt -0.06. Bezieht man die grossen Sprachfamilien nicht mit ein, ergibt sich ein Zusammenhang in die andere Richtung (0.06).

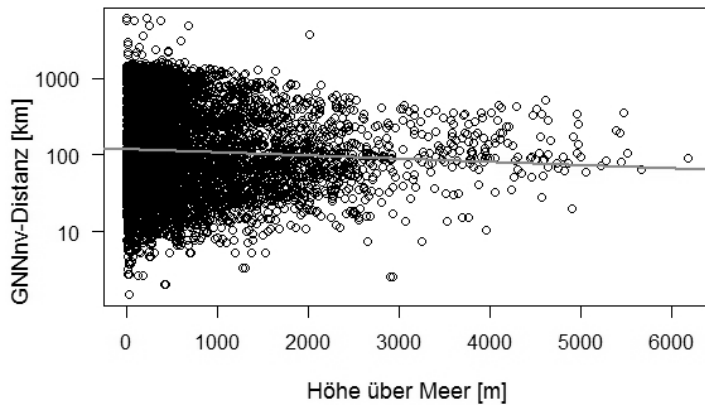


Abbildung 76: Verteilung der GNN-Distanz in logarithmierter Darstellung über die Höhe über Meer.

16.10 HÖHENDIFFERENZ

16.10.1 Lokale Sprachfamiliendichte

Wie sich die Sprachfamiliendichten auf die Höhendifferenz verteilen ist in Abbildung 77 zu sehen. Hohe Dichten treten von 0 bis circa 4'000 m auf, danach sind fast nur noch tiefe Dichten zu finden. Ab 4'000 m treten zwar keine hohen Dichten mehr auf, aber es hat auch keine ganz tiefen Werte mehr. Der Rangkorrelationskoeffizient beträgt 0.19 und 0.065 ohne die grossen Sprachfamilien.

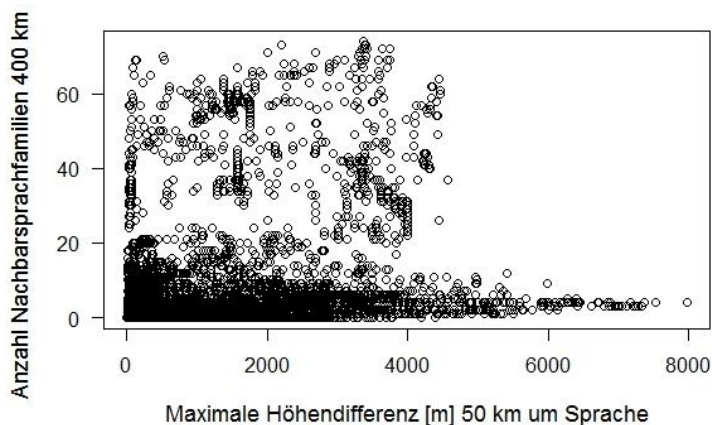
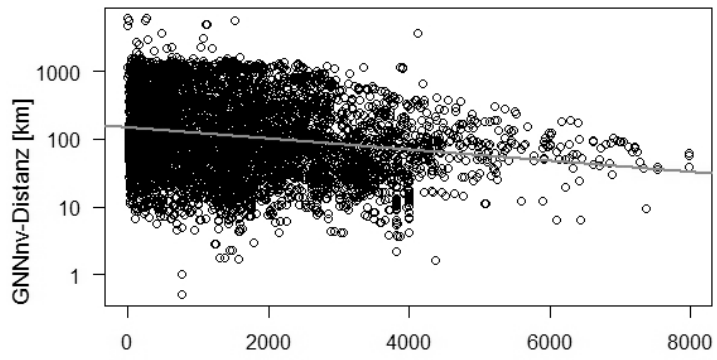


Abbildung 77: Verteilung der lokalen Sprachfamiliendichten über die maximale Höhendifferenz 50 km um die jeweilige Sprache.

16.10.2 Geographisch Nächster nicht-verwandter Nachbar

Wie Abbildung 78 zu entnehmen ist, sinkt die GNN-Distanz, je grösser die Höhendifferenz in einem Umkreis von 50 km um eine Sprache ist. Mit einem ρ von -0.16 ist die Korrelation die höchste, die in Zusammenhang mit der GNN-Distanz auftritt. Ohne grosse Sprachfamilien ist der Koeffizient mit -0.14 weniger stark ausgeprägt.



Maximale Höhendifferenz [m] 50 km um Sprache

Abbildung 78: Verteilung der GNNv-Distanz in logarithmierter Darstellung über die maximale Höhendifferenz 50 km um die jeweilige Sprache.

16.11 ZUSAMMENFASSUNG EINZELNE PRÄDIKTOREN

In der folgenden Tabelle 17 sind nochmals alle Rangkorrelationskoeffizienten der beiden Sprachfamiliendiversitätsmasse mit den geographischen Faktoren einmal mit und einmal ohne grosse Sprachfamilien aufgeführt.

Tabelle 17: Übersicht über die Rangkorrelationskoeffizienten ρ der lokalen Sprachfamiliendichte und der GNNnv-Distanz (mit und ohne grosse Sprachfamilien), für jeden einzelnen geographischen Faktor. Die grössten Werte sind fett hervorgehoben. Kursive Werte sind nicht signifikant. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 ". "<0.1 " "<1.

Geo. Faktor	Lokale Sprachfamiliendichte		Lokale Sprachfamiliendichte ohne grosse Sprachfamilien		GNNnv-Distanz		GNNnv-Distanz ohne grosse Sprachfamilien	
	ρ	Signifikanz	ρ	Signifikanz	ρ	Signifikanz	ρ	Signifikanz
Breitengrad	-0.208	***	-0.509	***	0.144	***	0.419	***
Temperatur	0.127	***	0.347	***	-0.147	***	-0.320	***
Temperaturschwankungsbereich	-0.164	***	-0.494	***	0.062	***	0.423	***
Niederschlag	0.213	***	0.392	***	-0.134	***	-0.353	***
Saisonalität Niederschlag	-0.101	***	-0.290	***	-0.062	***	0.115	***
NPP	0.159	***	0.314	***	-0.066	***	-0.300	***
Wachstumsphase	0.247	***	0.418	***	-0.114	***	-0.370	***
Küstendistanz	-0.020	.	-0.109	***	-0.067	***	0.152	***
Höhenmodell	0.046	***	-0.159	***	-0.059	***	0.064	***
Höhendifferenz	0.190	***	0.065	***	-0.164	***	-0.140	***

Bei der lokalen Sprachfamiliendichte bedeutet eine positive Korrelation, dass die Sprachfamiliendiversität mit dem entsprechenden Faktor zunimmt. Hingegen bedeutet bei der GNNnv-Distanz eine positive Korrelation, dass die Sprachfamiliendiversität mit grösser werdendem geographischem Faktor kleiner wird. Ohne die grossen Sprachfamilien weisen bei der Saisonalität des Niederschlags die beiden Masse in unterschiedliche Richtungen. Die lokale Sprachfamiliendichte zeigt an, dass die Sprachfamiliendiversität mit steigender Saisonalität kleiner wird. Während die GNNnv mit ihr sinkt respektive die Sprachfamiliendiversität steigt. Diese Diskrepanz ist möglicherweise auf den Einfluss der Cluster in den grossen Sprachfamilien zurückzuführen, denn die GNNnv-Distanz ohne die grossen Sprachfamilien weist mit 0.115 in die gleiche Richtung wie die Sprachfamiliendichte ebenfalls ohne die grossen Sprachfamilien. Sonst zeigen die beiden Masse, was die Sprachfamiliendiversität anbelangt, immer in die gleiche Richtung, insofern die Werte signifikant sind. Die beiden Masse ohne die grossen Sprachfamilien zeigen bei allen Koeffizienten in die entgegengesetzte, bezüglich Sprachfamiliendiversität in die gleiche Richtung. Die Stärken der Korrelationen sind recht ähnlich.

Der Zusammenhang der beiden verschiedenen lokalen Sprachendichten zeigt bei zwei geographischen Faktoren (Höhenmodell, Küstendistanz) in unterschiedliche Richtungen. Der Zusammenhang ist, ausser bei der Höhendifferenz, ohne die grossen Sprachfamilien ausgeprägter. Werden die Korrelationskoeffizienten der beiden GNNnv-Distanzen verglichen, lassen sich ebenfalls deutliche Unterschiede erkennen. Bei drei geographischen Faktoren (Saisonalität Niederschlag, Küstendistanz, Höhenmodell) gehen die Zusammenhänge in unterschiedliche Richtungen. Für die restlichen Faktoren gilt, ausser bei der Höhendifferenz, dass die Zusammenhänge ohne die grossen Sprachfamilien jeweils stärker sind. Die stärksten Korrelationen treten bei der lokalen Sprachfamiliendichte (ohne grosse Sprachfamilien) in Zusammenhang mit dem Temperaturschwankungsbereiches, der Länge der

Wachstumsphase und des Niederschlags auf. Die GNNv-Distanz (ohne grosse Sprachfamilien) korreliert ebenfalls mit diesen drei Faktoren am stärksten.

16.12 MULTIPLE REGRESSIONSANALYSE

Nachdem die Korrelation mit allen geographischen Faktoren einzeln geprüft wurde, wird nun eine multiple Regressionsanalyse durchgeführt. So soll analysiert werden, wie viel die geographischen Faktoren insgesamt an der Sprachfamiliendiversität erklären können. Dargestellt werden nur die Auswertungen ohne die grossen Sprachfamilien, jedoch werden auch die Resultate mit allen Sprachfamilien erwähnt. Es wird jeweils noch ein Modell ohne Temperaturschwankungsbereich und Länge der Wachstumsphase berechnet, in dem die Multikollinearität kleiner ist.

16.12.1 Lokale Sprachfamiliendichte

In Tabelle 18 ist das Resultat der multiplen Regressionsanalyse der lokalen Sprachfamiliendichte ohne die grossen Sprachfamilien zu sehen. Für diese Analyse wurde, nebst einigen Prädiktoren, auch die abhängige Variable transformiert, genauer gesagt logarithmiert. Die geographischen Faktoren können insgesamt 35.9 % (respektive 14.86 % mit allen Sprachfamilien) der Streuung der Sprachfamiliendichte erklären. Wird das Modell um die geographische Breite ergänzt, steigt das korrigierte Bestimmtheitsmass um einen respektive einen halben Prozentpunkt. Die Breitengrade liefern kaum zusätzliche Erklärungen zum Modell, sondern ihre Korrelation mit der Sprachfamiliendichte kann auf die geographischen Faktoren zurückgeführt werden. Die im Zusammenhang mit dem Klima stehenden Faktoren (die ersten sechs in der Tabelle) weisen in einem separaten Modell ein korrigiertes R^2 von 0.333 (respektive 0.096) und die lagebedingten Faktoren (die restlichen) ein R^2 von 0.07 (respektive 0.049). Diese beiden Typen von Faktoren erklären grösstenteils unterschiedliche Einflüsse, auch wenn einige Faktoren leicht untereinander korrelieren. Wird für die geographischen Faktoren noch eine Interaktion mit den Subsistenztypen gerechnet, lassen sich 59.5 % (32.7 % mit den grossen Sprachfamilien) der Varianz der lokalen Sprachfamiliendichten erklären. Je nach Subsistenztyp fallen die Einflüsse der geographischen Faktoren also unterschiedlich aus. Darauf wird nochmals genauer in Abschnitt 16.14.1 eingegangen.

Tabelle 18: Multiple Regressionsanalyse mit der lokalen Sprachfamiliendichte (ohne grosse Sprachfamilien) als abhängige Variable und den geographischen Faktoren als unabhängigen Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 " " <0.1 " " <1.

Geo. Faktor	Transformation	Koeffizienten	Standardisierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		2.109	0.845	< 2e-16	***
Temperatur		0.016	0.130	< 2e-16	***
Temperaturschwankungsbereich	log10	-0.811	-0.163	< 2e-16	***
Niederschlag	√	-0.001	-0.011	4.74E-01	
Saisonalität Niederschlag	√	-0.057	-0.127	< 2e-16	***
NPP		-0.023	-0.029	1.82E-02	*
Wachstumsphase		0.000	-0.046	3.99E-03	**
Küstendistanz	√	0.004	0.047	1.37E-06	***
Höhenmodell+1	log10	-0.097	-0.062	2.46E-08	***
Höhendifferenz	√	0.007	0.112	< 2e-16	***

p-Wert	<2E-16	***
Bestimmtheitsmass R^2	0.359	

In Tabelle 19 sind die Resultate der multiplen Regressionsanalyse mit einer kleineren Multikollinearität respektive ohne die geographischen Faktoren Temperaturschwankungsbereich und Länge der Wachstumsphase aufgeführt. Eine schrittweise Analyse der AIC-Werte hat ergeben, dass das beste Modell erreicht wird, wenn der jährliche Niederschlag ebenfalls nicht mit in das Modell einbezogen wird. Den höchsten Einfluss auf die Sprachfamiliendichte hat die Temperatur, gefolgt von der Saisonalität des Niederschlags. Die NPP und die Küstendistanz haben keinen signifikanten Einfluss. Die Küstendistanz ist nur so lange signifikant, wie der Temperaturschwankungsbereich im Modell ist. Zudem zeigte der Koeffizient einen Zusammenhang in die umgekehrte Richtung, als wenn der Zusammenhang zwischen Sprachfamiliendichte und Küstendistanz separat betrachtet wird. Dies ist wahrscheinlich auf eine *net suppression* zurückzuführen (Wuensch 2012).

Tabelle 19: Multiple Regressionsanalyse mit der logarithmierten lokalen Sprachfamiliendichte als abhängige Variable und sechs geographischen Faktoren als unabhängige Variablen Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 " " "<0.1 " "<1.

Geo. Faktor	Trans-formation	Koeffizienten	Standardisierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		0.810	0.845	< 2e-16	***
Temperatur		0.027	0.218	< 2e-16	***
Saisonalität Niederschlag	√	-0.073	-0.162	< 2e-16	***
NPP		-0.012	-0.015	1.16E-01	
Küstendistanz	√	0.001	0.015	1.20E-01	
Höhenmodell+1	log10	-0.075	-0.048	0.0000196	***
Höhendifferenz	√	0.007	0.117	< 2e-16	***

p-Wert	<2E-16	***
Bestimmtheitsmass R ²	0.331	

16.12.2 Geographisch Nächster nicht-verwandter Nachbar

In Tabelle 20 sind die Resultate der multiplen Regressionsanalyse der GNNnv-Distanz aufgeführt. Nebst einigen geographischen Faktoren, wie in der Spalte Transformation zu sehen ist, wurde auch die GNNnv-Distanz logarithmiert. Die geographischen Faktoren vermögen 28.1 % (11.4 % mit allen Sprachfamilien) der Streuung der Distanz zu erklären. Wird als Faktor noch die geographische Breite ergänzt, steigt das korrigierte Bestimmtheitsmass nur leicht von 0.2784 auf 0.2807. Somit erklärt sich auch hier die Korrelation mit den Breitengraden fast vollständig durch die geographischen Faktoren. Die klimatischen Faktoren (die ersten drei) alleine erklären 22.9 % (5.4 %) und die lagebedingten Faktoren 4.2 % der Varianz der GNNnv-Distanz. Das ist praktisch gleich viel wie mit den grossen Sprachfamilien (4.7 %). Wird im Modell in Tabelle 20 noch die Subsistenzstrategie als Interaktion mit den einzelnen Faktoren einbezogen, steigt die erklärable Varianz auf 35.72 %.

Tabelle 20: Multiple Regressionsanalyse mit der GNNv-Distanz (ohne grosse Sprachfamilien) als abhängige Variable und den geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "***"<0.01 "**"<0.05 "."<0.1 ""<1.

Geo. Faktor	Transformation	Koeffizienten	Standardisierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		1.828	1.809	< 2e-16	***
Temperatur		-0.022	-0.181	< 2e-16	***
Temperaturschwankungsbereich	log10	0.420	0.085	2.71E-07	***
Niederschlag	√	0.003	0.042	9.74E-03	**
Saisonalität Niederschlag	√	0.009	0.020	0.090921	.
NPP		0.033	0.041	1.69E-03	**
Wachstumsphase		-0.001	-0.058	7.25E-04	***
Küstendistanz	√	-0.002	-0.019	6.52E-02	.
Höhenmodell+1	log10	0.033	0.021	0.079833	.
Höhendifferenz	√	-0.008	-0.134	< 2e-16	***

p-Wert	<2E-16	***
Bestimmtheitsmass R ²	0.281	

Da zwischen den geographischen Faktoren eine Multikollinearität besteht, wurden die beiden Faktoren Temperaturschwankungsbereich und Wachstumsphase aus dem Modell entfernt, was dazu führt, dass diese deutlich sinkt. Mit diesem Modell wurde dann schrittweise der AIC-Wert überprüft. Daraus resultierte das Modell in Tabelle 21. Wie darin zu sehen ist, erklären lediglich drei geographische Faktoren rund 27 % der Varianz der GNNv-Distanz. Das ist nur rund ein Prozentpunkt weniger als im Modell mit allen Faktoren. Den grössten Einfluss hat die Temperatur. Der Einfluss der Saisonalität des Niederschlags wurde erst mit dem Entfernen des Temperaturschwankungsbereiches und der Wachstumsphase signifikant.

Tabelle 21: Multiple Regressionsanalyse mit der logarithmierten GNNv-Distanz als abhängige Variable und drei geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "***"<0.01 "**"<0.05 "."<0.1 ""<1.

Geo. Faktor	Transformation	Koeffizienten	Standardisierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		2.452	1.809	< 2e-16	***
Temperatur		-0.029	-0.234	< 2e-16	***
Saisonalität Niederschlag	√	0.026	0.058	3.24E-14	***
Höhendifferenz	√	-0.007	-0.122	< 2e-16	***

p-Wert	<2E-16	
Bestimmtheitsmass R ²	0.269	

16.13 MEHREBENENANALYSE

Mithilfe einer Mehrebenenanalyse soll herausgefunden werden, ob innerhalb von Sprachfamilien, beispielsweise aufgrund von kulturellen Einflüssen, *per se* höhere oder tiefere Sprachfamiliendiversitäten auftreten. Wiederum werden hier nur die Auswertungen ohne die grossen Sprachfamilien dargestellt. Ebenfalls werden noch die Resultate einer Mehrebenenanalyse pro Subsistenztyp Landwirtschaft und Jäger und Sammler aufgeführt. Die Auswertungen zu den Typen sesshafte Jäger und Sammler und Hirte sind im Anhang zu finden.

16.13.1 Lokale Sprachfamiliendichte

In Tabelle 22 sind die Resultate der Mehrebenenanalyse der lokalen Sprachfamiliendichte aufgeführt. Eine schrittweise Überprüfung der AIC-Werte hat ergeben, dass das beste Modell erreicht wird, wenn alle geographischen Faktoren einbezogen werden. Die Temperatur ist jener Faktor, der mit Abstand den grössten Einfluss auf die Anzahl an Nachbarsprachen anderer Sprachfamilien hat. Die Saisonalität des Niederschlags und die Höhendifferenz verlieren deutlich an Einfluss mit der Berücksichtigung der Sprachfamilien-Gruppen.

Tabelle 22: Mehrebenenanalyse der lokalen Sprachfamiliendichte mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 ". "<0.1 " "<1.

Geo. Faktor	Transformation	Koeffizienten	Standardisierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		0.401	1.030	0	***
Temperatur		0.016	0.133	0	***
Niederschlag	√	0.003	0.034	0	***
Saisonalität Niederschlag	√	-0.006	-0.012	0.0736	.
NPP		-0.009	-0.012	0.068	.
Küstendistanz	√	0.007	0.075	0	***
Höhenmodell+1	log10	0.028	0.018	0.0062	**
Höhendifferenz	√	0.002	0.034	0	***

Anzahl Sprachen	2635
Anzahl Gruppen	414

In der Tabelle 23 sind die einzelnen Mehrebenenanalysen nach Subsistenztyp aufgeführt. Zwischen der Landwirtschaft und den Jäger-und-Sammler-Gesellschaften bestehen die grössten Unterschiede bei der Saisonalität des Niederschlags und der Küstendistanz, da dort die Einflüsse in entgegengesetzte Richtungen gehen. Die Distanz zur Küste hat bei der Landwirtschaft sogar den grössten Einfluss, während es bei den Jäger und Sammlern die Temperatur ist.

Tabelle 23: Mehrebenenanalyse der Sprachfamiliendichte aufgeteilt nach Subsistenztyp Landwirtschaft und Jäger und Sammler. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 ". "<0.1 " "<1.

Geo. Faktor	Landwirtschaft			Jäger und Sammler		
	Koeffizienten	Standardisierte Koeffizienten	Signifikanz	Koeffizienten	Standardisierte Koeffizienten	Signifikanz
(Achsenabschnitt)	0.537	0.879	***	-0.364	0.775	***
Temperatur	0.009	0.066	***	0.028	0.304	***
Niederschlag	0.001	0.016	*	0.003	0.035	.
Saisonalität Niederschlag	-0.017	-0.036	***	0.039	0.082	***
NPP	-0.018	-0.023	**	0.044	0.047	*
Küstendistanz	0.011	0.121	***	-0.006	-0.071	***
Höhenmodell+1	-0.003	-0.002		0.100	0.062	***
Höhendifferenz	0.002	0.041	***	0.004	0.052	**

Anzahl Sprachen	1783			477		
Anzahl Gruppen	182			145		

16.13.2 Geographisch Nächster nicht-verwandter Nachbar

Bei einer schrittweisen Analyse der AIC-Werte ergibt sich, nach dem Entfernen des Temperaturschwankungsbereiches und der Länge der Wachstumsphase, ein Modell, welches nur noch vier geographische Faktoren beinhaltet. Die Resultate der Auswertung sind in Tabelle 24 zu finden. Der Faktor mit dem grössten Einfluss ist, wie schon bei der lokalen Sprachfamiliendichte, die Temperatur.

Tabelle 24: Mehrebenenanalyse der GNNnv-Distanz mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 * <0.1 " <1.

Geo. Faktor	Transformation	Koeffizienten	Standardisierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		2.553	1.720	0	***
Temperatur		-0.021	-0.169	0	***
Niederschlag	√	-0.004	-0.050	0	***
Küstendistanz	√	-0.004	-0.041	0.0001	***
Höhendifferenz	√	-0.006	-0.093	0	***

Anzahl Sprachen	2635
Anzahl Gruppen	414

In Tabelle 25 wurde für die Subsistenztypen Landwirtschaft und Jäger und Sammler je einzeln eine Mehrebenenanalyse mit den Sprachfamiliengruppen durchgeführt. Bei beiden hat die Temperatur den grössten Einfluss auf die Distanz zum nächsten nicht-verwandten Nachbarn. Ebenfalls einen grossen Einfluss hat die Höhendifferenz. Die Saisonalität des Niederschlags hat unterschiedliche Effekte auf die GNNnv-Distanz, je nach Subsistenztyp. Je regelmässiger verteilt der Niederschlag über das Jahr fällt, desto höher ist die Sprachfamiliendichte bei der Landwirtschaft und desto tiefer bei den Jäger-und-Sammler-Gesellschaften. Jedoch muss angemerkt werden, dass der Effekt nicht sehr gross ist und bei der Landwirtschaft nur das Signifikanzniveau von 0.01 erreicht.

Tabelle 25: Mehrebenenanalyse der GNNnv-Distanz aufgeteilt nach Subsistenztyp Landwirtschaft und Jäger und Sammler. Die Zeichen für die Signifikanz sind: **** <0.001 *** <0.01 ** <0.05 * <0.1 " <1.

Geo. Faktor	Landwirtschaft			Jäger und Sammler		
	Koeffizienten	Standardisierte Koeffizienten	Signifikanz	Koeffizienten	Standardisierte Koeffizienten	Signifikanz
(Achsenabschnitt)	2.294	1.748	***	3.208	2.008	***
Temperatur	-0.016	-0.114	***	-0.025	-0.265	***
Niederschlag	-0.002	-0.020		-0.004	-0.050	
Saisonalität Niederschlag	0.018	0.038	**	-0.038	-0.080	***
NPP	0.019	0.024	.	-0.068	-0.073	*
Küstendistanz	-0.010	-0.102	***	0.005	0.056	*
Höhenmodell+1	0.010	0.006		-0.053	-0.032	
Höhendifferenz	-0.005	-0.092	***	-0.009	-0.120	***

Anzahl Sprachen	1783			477		
Anzahl Gruppen	182			145		

16.14 MODELLVERGLEICH

In den beiden folgenden Abschnitten wird ein Modellvergleich mit den Modellen aus den vorangegangenen Unterkapiteln durchgeführt, um herauszufinden, welches die Daten am besten erklärt.

16.14.1 Lokale Sprachfamiliendichte

In Tabelle 26 ist eine Zusammenstellung der AIC-Werte der verschiedenen Modelle zu finden. Bei einer Varianzanalyse konnte festgestellt werden, dass sie alle signifikant (p -Wert < 0.001) unterschiedlich voneinander sind. Mit dem deutlich tiefsten AIC-Wert ist die Mehrebenenanalyse mit den Sprachfamilien-Gruppen inklusive der Interaktion der Subsistenztypen mit den geographischen Faktoren, das beste Modell. Beide gemischten Modelle sind deutlich besser als die beiden multiplen Regressionsmodelle, die den zufälligen Effekt der Sprachfamilien nicht berücksichtigen. Die Zugehörigkeit zu einer Sprachfamilie hat also *per se* schon einen Einfluss auf die Höhe der lokalen Sprachfamiliendichte. Ebenso entscheidet die Subsistenzstrategie darüber, wie sich einzelne geographische Faktoren auf die Sprachfamiliendichte auswirken.

Tabelle 26: Vergleich verschiedener Modelle zur lokalen Sprachfamiliendichte: Zwei multiple Regressionsmodelle und zwei Mehrebenenanalysen jeweils mit und ohne Subsistenzfaktor als Interaktion.

AIC	Multiples Regressionsmodell		Mehrebenenanalyse	
	Multiples Regressionsmodell	Multiples Regressionsmodell mit Faktor Subsistenzstrategie	Sprachfamilie	Sprachfamilie, Faktor Subsistenzstrategie
AIC	2179	1052	-1117	-1697

Es verwundert nicht, dass das gemischte Modell, welches Zufallseffekte aufgrund der Sprachfamilien-Zugehörigkeit berücksichtigt, das beste Modell ist. Erstens gibt es möglicherweise kulturelle Gründe, dass Sprachen von gewissen Sprachfamilien mehr Nachbarsprachen anderer Familien haben. Zweitens bewegt sich die Sprachfamiliendichte von Sprachen der gleichen Sprachfamilie eher auf gleichem Niveau, da sie räumlich gehäuft vorkommen und so oft die gleichen Nachbarsprachen anderer Familien aufweisen. Ein dritter Punkt trifft vor allem auf grosse Sprachfamilien zu, von denen die meisten in diesen Modellen nicht verwendet wurden. Denn Sprachen grosser Sprachfamilien befinden sich oftmals in Gebieten, wo es kaum andere Sprachfamilien gibt und weisen dadurch alle sehr ähnliche, tiefe Sprachfamiliendiversitäten auf. Dies ist darauf zurückzuführen, dass diese sich in der Vergangenheit zum Teil massiv ausgebreitet haben und dabei andere Sprachen verdrängt haben (Diamond 1997).

16.14.2 Geographisch Nächster nicht-verwandter Nachbar

Alle berechneten Modelle wurden miteinander verglichen, wobei sich zeigte, dass alle signifikant (p -Wert < 0.001) unterschiedlich voneinander sind (Tabelle 27). Das Modell, welches die Daten am besten beschreibt, ist wiederum das gemischte Modell mit den Sprachfamilien-Gruppen und der Interaktion der Subsistenzstrategie mit den geographischen Faktoren. Somit hat auch hier, wie schon bei der lokalen Sprachfamiliendichte, die Sprachfamilie einen Einfluss auf das Niveau der GNNv-Distanz. Auch hier gilt wiederum anzumerken, dass dies einerseits auf kulturelle und andererseits auf Cluster-Effekte zurückzuführen ist.

Tabelle 27: Vergleich verschiedener Modelle zur GNNv-Distanz: Zwei multiple Regressionsmodelle und zwei Mehrebenenanalysen jeweils mit und ohne Subsistenzfaktor als Interaktion.

	Multiples Regressionsmodell		Mehrebenenanalyse	
	Multiples Regressionsmodell	Multiples Regressionsmodell inkl. Interaktion Subsistenzstrategie	Sprachfamilie	Sprachfamilie inkl. Interaktion Subsistenzstrategie
AIC	2440	2205	1812	1677

16.15 VERGLEICH DER SPRACHFAMILIENDIVERSITÄTS-MASSE

Die lokale Sprachfamiliendichte und die GNNnv-Distanz stellen beide ein Mass für die Sprachfamiliendiversität dar. Um festzustellen wie stark die beiden Masse miteinander korrelieren, wurden sie verglichen. In Abbildung 79 ist ein Vergleich zwischen der GNNnv-Distanz und der Sprachfamiliendichte dargestellt. Der Rangkorrelationskoeffizient beträgt -0.76 . Bei einer Anzahl von null Nachbarsprachfamilien treten erst ab 400 km GNNnv-Werte auf, da dies dem Suchradius der Nachbarsprachfamilien entspricht.

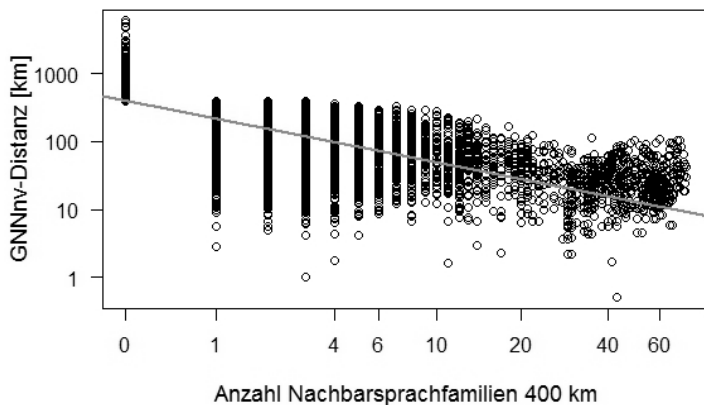


Abbildung 79: Vergleich zwischen der GNNnv-Distanz in logarithmierter Darstellung und der Anzahl an Nachbarsprachfamilien in logarithmierter Darstellung.

In Abbildung 80 ist der gleiche Zusammenhang nochmals dargestellt, diesmal jedoch ohne die grossen Sprachfamilien. Der Rangkorrelationskoeffizient beträgt mit -0.58 weniger als über alle Sprachen gesehen. Bei beiden Abbildungen ist deutlich zu sehen, dass bei wenigen Nachbarsprachen die Streuung der GNNnv-Distanz grösser ist. Das kommt daher, dass diese eine exaktere Beschreibung erlaubt. Dieser Effekt ist bei mehr Nachbarsprachfamilien kleiner.

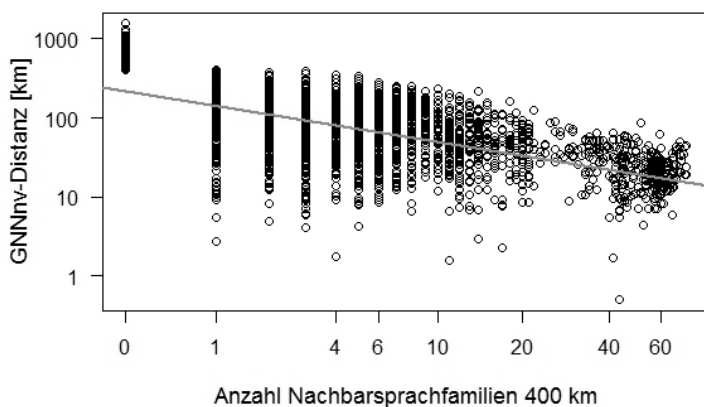


Abbildung 80: Vergleich zwischen der GNNnv-Distanz in logarithmierter Darstellung und der Anzahl an Nachbarsprachfamilien in logarithmierter Darstellung ohne grosse Sprachfamilien.

In der nachfolgenden Tabelle 28 ist ein Vergleich zwischen der Mehrebenenanalyse der lokalen Sprachfamiliendichte und der GNNnv-Distanz dargestellt. Damit die beiden miteinander verglichen werden konnten, wurde für beide das Modell ohne den Temperaturschwankungsbereich und die Länge der Wachstumsphase (niedrige Multikollinearität) berechnet. Bei beiden Modellen hat die Temperatur den grössten Einfluss auf die jeweiligen Masse. Bei allen signifikanten Werten sind die Richtungen der Sprachfamiliendiversität dieselbe, einzig die Stärke des Einflusses variiert.

Tabelle 28: Vergleich der Mehrebenenanalyse der lokalen Sprachfamiliendichte und der GNNv-Distanz. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 ". "<0.1 " "<1.

Geo. Faktor	Mehrebenenanalyse - Sprachfamilien					
	Lokale Sprachfamiliendichte			GNNv-Distanz		
	Koeffizienten	Standardisierte Koeffizienten	Signifikanz	Koeffizienten	Standardisierte Koeffizienten	Signifikanz
(Achsenabschnitt)	0.401	1.030	***	2.515	1.723	***
Temperatur	0.016	0.133	***	-0.022	-0.175	***
Niederschlag	0.003	0.034	***	-0.004	-0.049	***
Saisonalität Niederschlag	-0.006	-0.012	.	0.007	0.016	
NPP	-0.009	-0.012	.	0.008	0.010	
Küstendistanz	0.007	0.075	***	-0.004	-0.040	***
Höhenmodell+1	0.028	0.018	**	-0.008	-0.005	
Höhendifferenz	0.002	0.034	***	-0.006	-0.093	***
Anzahl Sprachen	2635			2635		
Anzahl Gruppen	414			414		

16.16 INTERPRETATION

16.16.1 Masse für Sprachfamiliendiversität

Es ist schwierig ein geeignetes Mass für die Sprachfamiliendiversität zu finden, da diese sehr unregelmässig auf der Welt verteilt ist. Die beiden verwendeten Methoden stellen eine Möglichkeit dar, wie diese quantifiziert werden kann. Jedoch weisen sowohl das Mass der lokalen Sprachfamiliendichte wie auch die Distanz zum Geographisch Nächsten nicht-verwandten Nachbarn einige Schwachpunkte auf.

Wie schon bei der lokalen Sprachendichte (siehe *Abschnitt 15.16.1*), stellt sich auch bei der Sprachfamiliendichte die Frage nach der Grösse des Radius. Mit 400 km wurde ein guter Kompromiss ausgewählt, jedoch gibt es mit diesem Radius 1'304 Sprachen, die keine Nachbarsprachfamilie haben. Wie im *Abschnitt 10.1.2* gesehen werden konnte, gibt es Sprachen, die erst in einer Entfernung von über 1'000 km eine Nachbarsprache einer anderen Familie haben. Würde ein noch grösserer Radius verwendet, würde immer mehr der Bezug zu den geographischen Faktoren verloren gehen, da diese schon auf relativ kurze Distanzen stark variieren können. Ein anderes Problem dieser Methode ist, dass sie nur die Anzahl der Nachbarsprachfamilien ausgibt, aber keine weiteren Informationen. Egal wie weit oder nah die nächsten Sprachen anderer Sprachfamilien innerhalb des 400 km Radius liegen. Ein Problem, welches die Sprachen an den Küsten betrifft ist, dass dort mit einem Puffer von vorhinein weniger Sprachen erfasst werden, da ein grosser Teil des Umkreises auf Meeresgebiet liegt. Im *Abschnitt 16.8.1*, in der die Küstendistanz analysiert wird, ist dies sehr schön zu sehen. Denn dort ist für die Insel Neuguinea deutlich erkennbar, dass die Sprachen im Landesinnern viel mehr Nachbarsprachfamilien aufweisen als jene an den Küsten. Bei der lokalen Sprachfamiliendichte kann zwar der Effekt der grossen Sprachfamilien, ausser für *Nuclear-Trans-New-Guinea*, nicht so deutlich gesehen werden wie bei der GNNnv-Distanz, dennoch macht es auch hier Sinn die Interpretation mit den Auswertungen ohne die grossen Sprachfamilien zu machen. Denn einige dieser Familien haben sich mit der Landwirtschaft stark ausgebreitet und dabei andere Sprachen verdrängt (Diamond & Bellwood 2003). Da dies hauptsächlich aufgrund von kulturellen und historischen Prozessen geschah, ist dies logischerweise mit geographischen Faktoren gar nicht zu erfassen.

Bei den GNNnv-Distanzen lassen sich zwei Phänomene aus den Daten erkennen, die die Resultate verfälschen können. Erstens weisen Sprachen auf Inseln oftmals sehr grosse Werte auf. Dies liegt einerseits daran, dass dort oft nur eine Sprache gesprochen wird und die nächste Nachbarsprachfamilie dementsprechend erst ein paar Inseln weiter liegt. Andererseits liegt das daran, dass auf vielen Inselgruppen im Pazifik nur austronesische Sprachen vorkommen und die dann alle meist die gleiche nächste Nachbarsprachfamilie haben, die sehr weit entfernt ist. Diese Sprachen waren in den Abbildungen mit den geographischen Faktoren immer wieder als Anhäufungen zu erkennen. Das zweite Phänomen, welches durch diese Methode abgebildet wird, ist die räumliche Dominanz von einzelnen grossen Sprachfamilien, die schon im vorherigen Abschnitt erläutert wurde. Die Sprachen, welche zu solchen Sprachfamilien gehören, weisen meist sehr grosse oder ähnliche GNNnv-Distanzen auf. Damit die Resultate möglichst nicht von kulturellen oder historischen Prozessen überlagert werden, wurden die Modellanalysen ohne die grossen Sprachfamilien durchgeführt.

Dass die grossen Sprachfamilien bei der Analyse der Sprachfamiliendiversität ausgeschlossen wurden, muss natürlich bei der Interpretation bedacht werden. So bezieht sich diese nur auf circa ein Drittel der Sprachen und ist auch nur für diese gültig.

16.16.2 Einordnung der Resultate

Nichols (1990) formulierte in einem Artikel Beobachtungen, wonach die Sprachfamiliendichten besonders entlang von Küsten, in Bergen und bei viel Niederschlag hoch sind. Sie quantifizierte diese Aussagen jedoch nicht. In dieser Arbeit wurde hingegen versucht mithilfe von Analysen und Berechnungen einen Zusammenhang zwischen geographischen Faktoren und der Sprachfamiliendiversität zu finden. In der folgenden Interpretation der Resultate wird, da alle Daten zusammen zu stark historisch und kulturell geprägt sind (siehe *Abschnitt 10.1.2*), nur auf die Daten ohne die grossen Sprachfamilien eingegangen. Zur Analyse mit allen Daten kann jedoch festgehalten werden, dass die Zusammenhänge zu den einzelnen geographischen Faktoren, wie erwartet eher tief ausfallen. In einem multiplen Regressionsmodell konnte zudem nur 14.86 % der Varianz der lokalen Sprachfamiliendichten und nur 11.4 % der Varianz der GNNv-Distanz durch geographische Faktoren erklärt werden. Es verwundert deshalb auch nicht, dass das Mehrebenenmodell mit den Sprachfamilien bei beiden Massen am deutlich besten abschneidet, da so die grossen Unterschiede in den Sprachfamilien ein wenig abgefangen werden konnten.

Nun zur Interpretation der Analysen und Resultate ohne die Sprachen von grossen Sprachfamilien. Bei der Analyse wurde festgestellt, dass alle geographischen Faktoren einzeln betrachtet einen signifikanten Zusammenhang zur Sprachfamiliendiversität aufweisen. Der Temperaturschwankungsbereich und die Wachstumsphase weisen die grössten Zusammenhänge mit den beiden Massen auf. Insgesamt konnten die geographischen Faktoren 38.2 % der Varianz der lokalen Sprachfamiliendichte und 29.2 % der Varianz der GNNv-Distanz erklären. Wird zusätzlich noch die Subsistenzstrategie im Modell ergänzt steigt die erklärbare Streuung der Sprachfamiliendiversität auf 59.5 % (Sprachfamiliendichte) respektive 35.7 % (GNNv-Distanz). Die geographische Breite liefert keine zusätzliche Erklärung zu diesen Modellen, weshalb die Korrelation mit ihr vollständig durch die verwendeten geographischen Faktoren erklärt werden kann. Somit scheint der von Nichols (1990) beobachtete Breitengrad-Gradient geklärt zu sein. Hingegen kann die von ihr gemachte Beobachtung, dass die Sprachfamiliendichte an Küsten höher ist, nur teilweise bestätigt werden. Einzeln betrachtet gibt es zwar einen signifikanten negativen Zusammenhang zwischen der Sprachfamiliendiversität und der Küstendistanz, im auf Multikollinearität kontrollierten multiplen Regressionsmodell hat die Küstendistanz allerdings keinen signifikanten Einfluss mehr. Eine weitere Beobachtung von Nichols (1990) war, dass der Niederschlag, vor allem auch in Zusammenhang mit der Küste, erhöhte Sprachfamiliendiversität erklärt. Dieser hat sowohl einzeln betrachtet wie auch im Mehrebenenmodell einen signifikant positiven Einfluss. Noch wichtiger scheint jedoch die Temperatur zu sein, wie aus dem gemischten Modell hervorgeht. Die letzte Beobachtung von Nichols (1990) war, dass in bergigen Gebieten die Sprachfamiliendiversität hoch ist, da dort Gesellschaften natürlich separiert werden, keine Integration in ein grösseres Wirtschaftssystem erfolgt und das dort Zufluchtsorte gebildet werden. Die Resultate dieser Arbeit unterstützen ihre Vermutungen. Denn je grösser die Höhendifferenz ist, desto höher ist die Sprachfamiliendiversität, auch unter Berücksichtigung der Sprachfamilien.

Mittels einer Mehrebenenanalyse konnte gezeigt werden, dass je nach Sprachfamilie ein anderes Niveau von Sprachfamiliendiversität zu erwarten ist. Dies ist natürlich nicht überraschend, da Mitglieder einer Sprachfamilie, da sie meist nahe beieinander liegen, ähnliche Sprachfamiliendichten aufweisen. Andererseits gibt es auch kulturelle Gründe dafür, dass einige Sprachen unter den gleichen Umweltbedingungen andere Diversitätsniveaus aufweisen. Dieser Effekt war auch schon bei der Sprachendiversität feststellbar. Allgemein sind sich die beiden Resultate sehr ähnlich. Werden die grossen Sprachfamilien weggelassen, lassen sich mit den geographischen Faktoren fast gleich viel der Varianz der Sprachfamiliendiversität erklären (35.9 % und 28.1 %) wie bei der Sprachendiversität über alle Sprachen (37.3 und 30.9 %). Sowohl die Sprachfamilien- als auch die Sprachendiversität können zu etwa einem Drittel mit geographischen Faktoren erklärt werden. Diesbezüglich könnte nun argumentiert werden, dass die Sprachfamilien das gleiche Bild abgeben wie die Sprachen, jedoch einen älteren Zeitstand davon, da es sehr lange geht bis diese nicht mehr als Sprachen einer gemeinsamen Familie erkennbar sind. Da bei der Berechnung der Sprachfamiliendiversität die meisten der weggelassenen Sprachfamilien auf historische und kulturelle Gründe zurückzuführende Ausbreitungsschübe aufweisen, widerspiegeln die

restlichen womöglich besser noch den ursprünglichen Einfluss der geographischen Faktoren. Dieser ist anscheinend sehr ähnlich wie bei der Sprachdiversität. Dass sich einige Sprachfamilien mit der Erfindung der Landwirtschaft ausgebreitet haben, lässt sich auch bei einem Vergleich der Resultate der Sprachen- und Sprachfamiliendiversität aufgeteilt nach Subsistenztyp erkennen. Es ist interessant, dass die erklärte Varianz der Sprachen mit Subsistenztyp Landwirtschaft bei der Sprachfamiliendiversität deutlich tiefer ist (13.6 und 10.9 %) als bei der Sprachdiversität (29.4 und 23.1 %). Während die Sprachdiversität auch bei der Landwirtschaft zu einem guten Teil mit den geographischen Faktoren erklärt werden kann, ist dies bei der Sprachfamiliendiversität weniger der Fall. Dies spricht dafür, dass einige Sprachfamilien mit Subsistenzstrategie Landwirtschaft unabhängig von Umwelteinflüssen andere Sprachfamilien verdrängt haben. Werden die grossen Sprachfamilien nämlich weggelassen, kann wieder deutlich mehr der Varianz der Sprachdiversität (bei Subsistenztyp Landwirtschaft) mit geographischen Faktoren erklärt werden (28.2 und 16.8 %). Wird im multiplen Regressionsmodell, ohne die grossen Sprachfamilien, noch eine Interaktion mit der Subsistenzstrategie ergänzt, lassen sich sogar 59.5 % respektive 35.72 % der Varianz der Sprachfamiliendiversität erklären. Der Subsistenztyp spielt somit eine wesentliche Rolle, wenn es um die Einflüsse von Umweltfaktoren auf die Sprachfamiliendiversität geht.

Zusammenfassend lässt sich festhalten, dass die Sprachfamiliendiversität, wie auch schon die Sprachdiversität, etwa zu einem Drittel durch geographische Faktoren erklärt werden kann. Da die Sprachfamilien ältere Zeitstände widerspiegeln, scheinen diese Einflüsse schon lange zu wirken. Dies ist wahrscheinlich auch der Grund dafür, dass andere geographische Faktoren mehr oder weniger Einfluss haben, da sich diese selbst auch über die Zeit veränderten. Dies trifft allerdings nur dann zu, wenn die grössten Sprachfamilien, deren Anzahl Nachbarsprachen offenbar unabhängiger von der Umwelt ist, weggelassen werden.

17 GENEALOGISCHE DIVERSITÄT

Wie bereits bei der Sprachen- und Sprachfamiliendiversität wird bei der genealogischen Diversität auch zuerst die globale Verteilung betrachtet und danach der Zusammenhang zu jedem einzelnen geographischen Faktor. Da die Zusammenhänge und Verteilungen zwischen den beiden Massen, mittlere Verwandtschaftsdistanz und PNN-Distanz, auf die geographischen Faktoren sehr ähnlich ist, wird jeweils nur die Abbildung mit der mittleren Verwandtschaftsdistanz dargestellt. Die Grafiken zur PNN-Distanz finden sich auf der beigelegten CD. Oftmals zeigen sich in den Abbildungen Muster, die schon bei der GNN-Distanz sichtbar waren. Wenn diese Muster auf dieselben Gründe zurückgehen wie dort, was meistens mit der globalen Verteilung des geographischen Faktors zu tun hat, werden diese hier nicht nochmals erläutert. Nach der Auswertung der einzelnen Faktoren sind die Resultate der multiplen Regressionsanalysen und die Mehrebenenanalyse dargestellt. Am Ende des Kapitels folgen die Interpretation und die Einordnung der Resultate. In den Zusammenfassungen sind jeweils auch die Korrelationen mit den PNN-Distanzen der AUTOTYP-Daten aufgeführt. So soll überprüft werden, ob diese Methode auch auf andere Sprachdatenbanken übertragen werden kann oder ob sie zu stark von der Taxonomie von den Hammarström-Daten abhängt.

17.1 GLOBALE VERTEILUNG

Um einen Überblick über die globale Verteilung der beiden Masse für die genealogische Diversität zu erhalten, wird der Zusammenhang zur geographischen Breite untersucht. In Abbildung 81 ist die Verteilung der mittleren Verwandtschaftsdistanz über die geographische Breite zu sehen. Wie darin erkennbar ist, bilden Sprachen einer Sprachfamilie vertikale Cluster. Um die Tendenzen vom Äquator hin zu den Polen besser erkennen zu können, sind in der Abbildung 82 die Werte der Nord- und Südhalbkugel übereinander gelegt. Darin wird deutlich, dass in Richtung der Pole die mittlere Distanz zu den nächsten Verwandten grösser wird. Der Rangkorrelationskoeffizient beträgt 0.29 und für die PNN-Distanz 0.27.

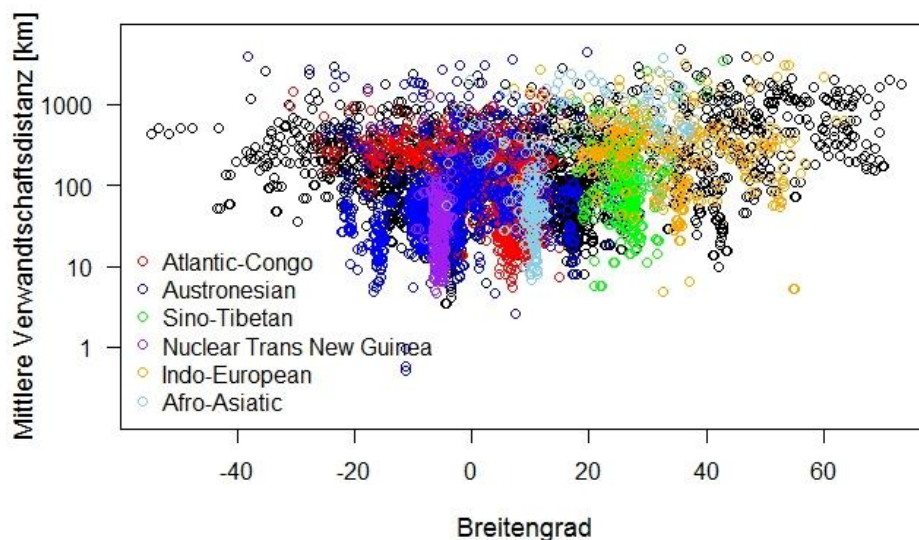


Abbildung 81: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die Breitengrade. Farblich hervorgehoben sind die grössten Sprachfamilien.

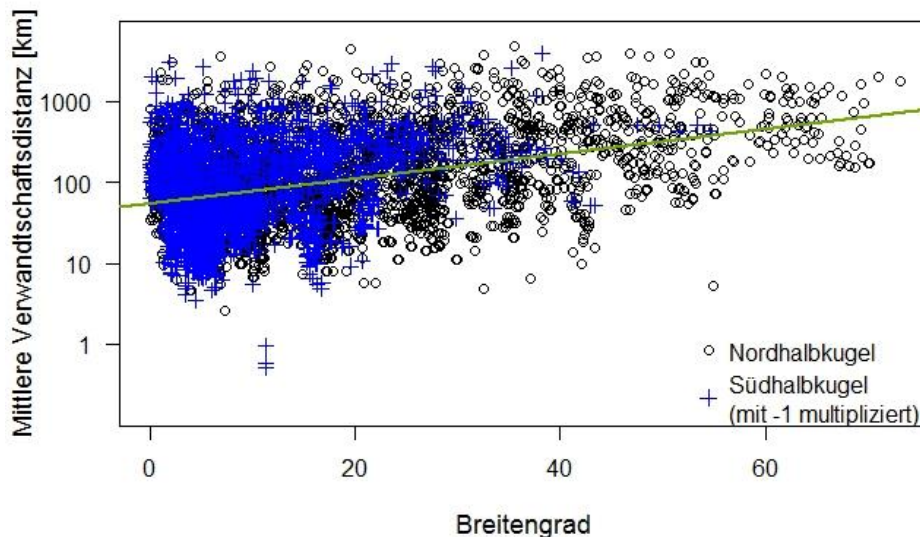


Abbildung 82: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die Breitengrade. In blau die Werte der Südhalbkugel, in schwarze jene der Nordhalbkugel.

In beiden Abbildungen sind die Tendenzen klar ersichtlich. Wenn diese jedoch mit jenen der GNN-Distanz in Abschnitt 15.1.2 verglichen werden, fällt auf, dass die mittlere Verwandtschaftsdistanz eine stärkere Streuung der Werte aufweist. Während kurze Distanzen mit höherer geographischer Breite deutlich abnehmen, treten grosse Distanzen auf fast allen Breitengraden auf. Dies ist vor allem auf Sprachen auf kleinen Inseln im Pazifik zurückzuführen. Dort hat es meist nur eine Sprache pro Insel und die nächsten verwandten Sprachen liegen auf den nächsten Inseln. Aber auch Sprachen im Amazonas weisen zum Teil grosse Distanzen zu ihren nächsten Verwandten auf. Die säulenartigen Strukturen, die sich in Abbildung 81 zeigen, waren schon bei der GNN-Distanz sichtbar. Cluster von Sprachen führen dazu, dass die dazugehörigen Sprachen sowohl was die Breitengrade, als auch was die Verwandtschaftsdistanz betrifft, recht ähnliche Werte aufweisen. Dieses Muster lässt sich auch in den folgenden Unterkapiteln im Zusammenhang mit den geographischen Faktoren immer wieder finden.

17.2 MITTLERE JÄHRLICHE TEMPERATUR

Abbildung 83 zeigt, dass die logarithmierte mittlere Verwandtschaftsdistanz mit steigender Temperatur abnimmt. Die nächsten verwandten Sprachen sind bei höheren mittleren jährlichen Temperaturen näher beieinander als bei tieferen Temperaturen. Der Rangkorrelationskoeffizient beträgt -0.15 für die mittlere Verwandtschaftsdistanz und -0.13 für die PNN-Distanz.

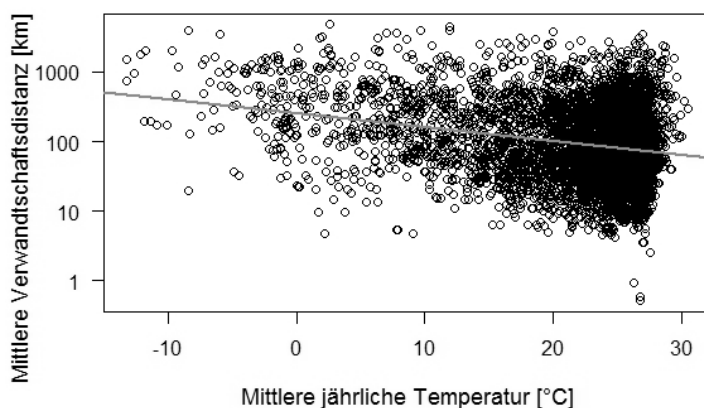


Abbildung 83: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die mittlere jährliche Temperatur.

17.3 JÄHRLICHER TEMPERATURSCHWANKUNGSBEREICH

Wie aus Abbildung 84 sichtbar wird steigt die logarithmierte Verwandtschaftsdistanz mit dem jährlichen Temperaturschwankungsbereich an. Sprachen, die sich an Orten befinden wo es während eines Jahres extreme Temperaturunterschiede gibt, haben ihre nächsten Verwandten tendenziell in weiterer Entfernung als Sprachen, bei welchen es das ganze Jahr hindurch in etwa gleich warm ist. Das ρ beträgt 0.41 für die mittlere Verwandtschaftsdistanz und 0.4 für die PNN-Distanz.

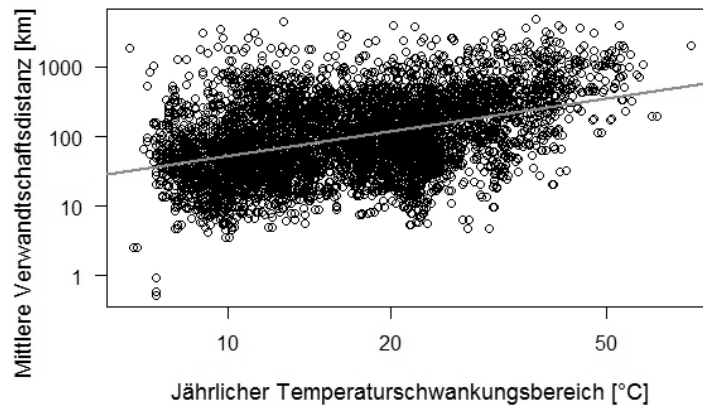


Abbildung 84: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über den jährlichen Temperaturschwankungsbereich in logarithmierter Darstellung.

Wird auf das Bestimmtheitsmass zurückgegriffen, kann auch gesagt werden, dass 19 % der Streuung der mittleren Verwandtschaftsdistanz mit dem Temperaturschwankungsbereich erklärt werden können. Das ist ein beträchtlicher Anteil, wenn bedacht wird, dass es sich lediglich um einen einzigen Faktor handelt. Die beiden Häufungen, die sich um 10 und 20°C zeigen, waren schon bei den beiden Sprachdiversitätsmassen sichtbar und wurden dort auch schon begründet (siehe *Unterkapitel 15.3*).

17.4 JÄHRLICHER NIEDERSCHLAG

Die logarithmierte mittlere Distanz zu den nächsten verwandten Sprachen nimmt mit dem jährlichen Niederschlag ab (Abbildung 85). Der Rangkorrelationskoeffizient der mittleren Verwandtschaftsdistanz beträgt -0.38 und jener der PNN-Distanz -0.37.

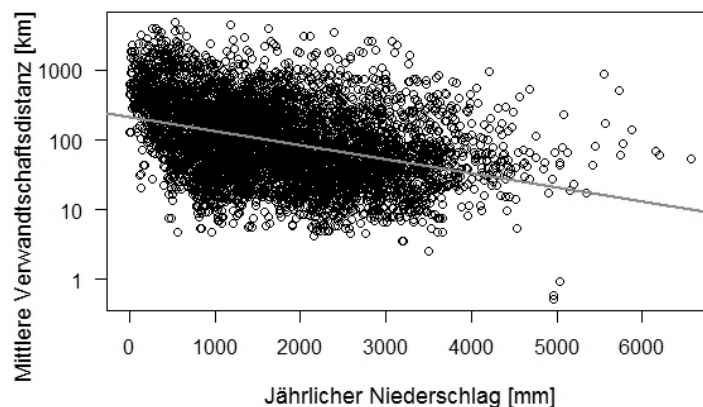


Abbildung 85: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über den jährlichen Niederschlag.

17.5 SAISONALITÄT DES NIEDERSCHLAGS

Wie in Abbildung 86 zu sehen ist nimmt die mittlere Verwandtschaftsdistanz mit der Saisonalität des Niederschlags zu. Je unregelmässiger verteilt der Niederschlag über das Jahr fällt, desto grösser ist die mittlere Distanz zu den nächsten verwandten Sprachen. Der Rangkorrelationskoeffizient beträgt 0.14 und für die PNN-Distanz 0.11.

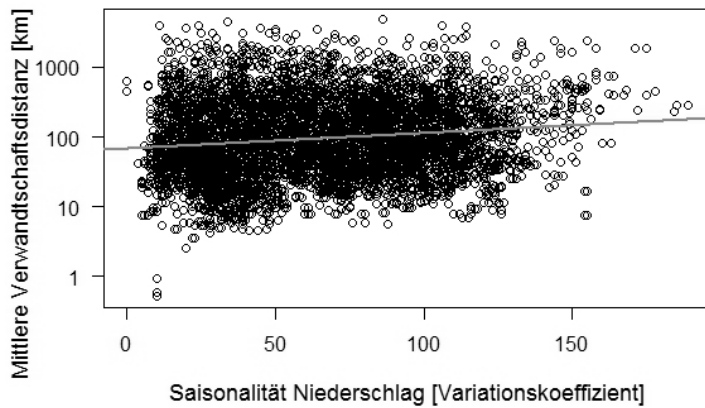


Abbildung 86: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die Saisonalität des Niederschlags.

17.6 NETTOPRIMÄRPRODUKTION

Die Distanz von einer Sprache zu den ihr am nächsten verwandten Sprachen sinkt mit zunehmender Nettoprimärproduktion (Abbildung 87). Der Rangkorrelationskoeffizient der mittleren Verwandtschaftsdistanz hat einen Wert von -0.34 und jener der PNN-Distanz einen Wert von -0.33.

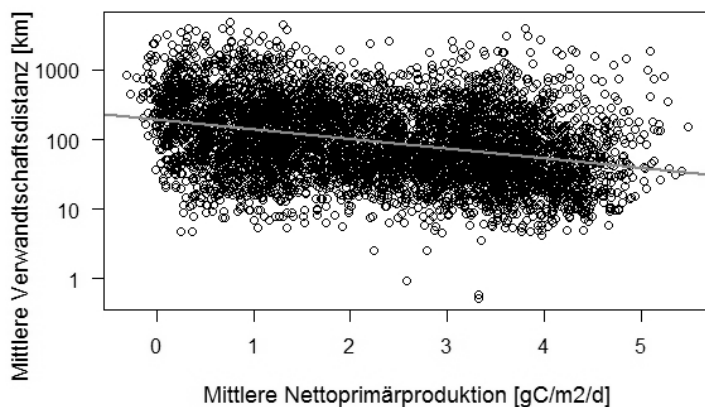


Abbildung 87: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die mittlere Nettoprimärproduktion.

17.7 LÄNGE DER WACHSTUMSPHASE

Abbildung 88 veranschaulicht, dass die mittlere Verwandtschaftsdistanz mit der Länge der Wachstumsphase zunimmt. Wie schon bei der Sprachendiversität konzentrieren sich bei null sowie 365 und 366 Tagen viele Sprachen. Kurz danach respektive kurz davor hat es dagegen weniger Sprachen. Dass das darauf zurückzuführen ist, dass es grosse Gebiete mit 0 und 365/366 Tagen Länge hat, wurde bereits in Unterkapitel 15.7 besprochen. Der Rangkorrelationskoeffizient der mittleren Verwandtschaftsdistanz beträgt -0.36 und jener der PNN-Distanz -0.34.

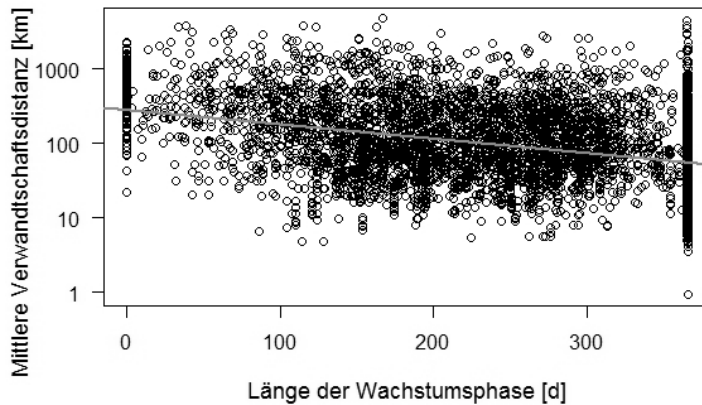


Abbildung 88: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die Länge der Wachstumsphase.

17.8 KÜSTENDISTANZ

Wie in Abbildung 89 zu sehen ist sind die nächsten Verwandten einer Sprache mit zunehmender Entfernung von der Küste weiter entfernt. Der Rangkorrelationskoeffizient der mittleren Verwandtschaftsdistanz hat einen Wert von 0.29 und jener der PNN-Distanz 0.3.

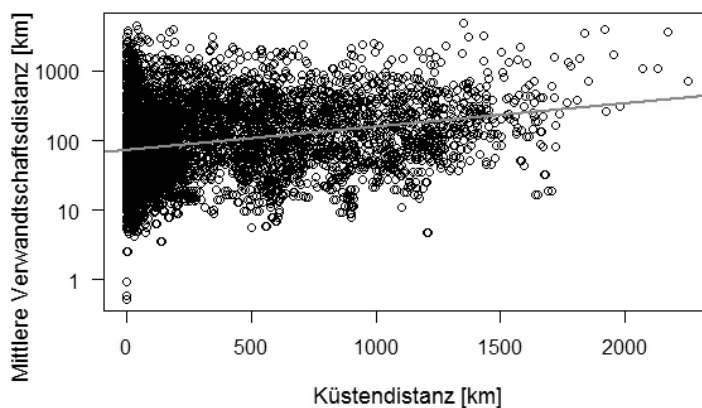


Abbildung 89: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die Küstendistanz.

In der Abbildung 89 lassen sich zwei Peaks hin zu tiefen Verwandtschaftsdistanzen bei circa 500 und 800 km Entfernung zur Küste erkennen. Der erste lässt sich mit Niger-Kongo-Sprachen erklären, welche in dieser Entfernung in Kamerun ein Cluster bilden und darin relativ kleine Verwandtschaftsdistanzen aufweisen. Der zweite Peak mit tiefen Werten setzt sich hauptsächlich aus afroasiatischen Sprachen zusammen, welche in Nigeria ein Cluster in circa 800 km Entfernung zum Meer bilden. Diese Peaks zeigten sich auch schon bei der lokalen Sprachendichte in Abschnitt 15.8.1.

17.9 HÖHENMODELL

Zwischen der mittleren Verwandtschaftsdistanz und der Höhe über Meeresspiegel lässt sich in der Abbildung 90 kein Zusammenhang feststellen. Dies zeigt sich auch in den Resultaten der Korrelation. Der Rangkorrelationskoeffizient für die mittlere Verwandtschaftsdistanz beträgt -0.008 und jener der PNN-Distanz -0.0001 . Beide Werte erreichen das Signifikanzniveau von 0.1% nicht.

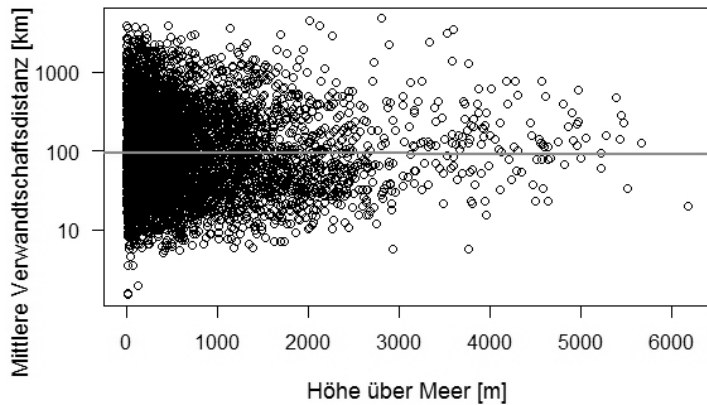


Abbildung 90: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die Höhe über Meer.

17.10 HÖHENDIFFERENZ

In Abbildung 91 ist klar zu erkennen, dass die mittlere Distanz zu den nächsten verwandten Sprachen abnimmt, wenn die Höhendifferenz um eine Sprache zunimmt. Bei einer topographischen Homogenität sind also eher grosse Distanzen zu finden, bei Heterogenität eher kleinere Distanzen. Der Rangkorrelationskoeffizient der mittleren Verwandtschaftsdistanz und der PNN-Distanz beträgt rund -0.28.

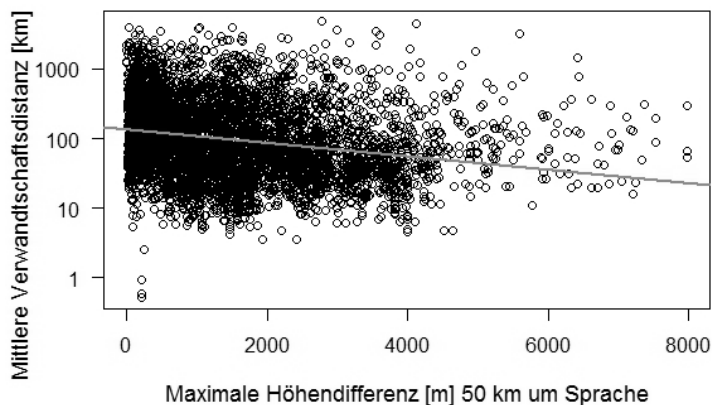


Abbildung 91: Verteilung der mittleren Verwandtschaftsdistanz in logarithmierter Darstellung über die maximale Höhendifferenz 50 km um die jeweilige Sprache.

17.11 ZUSAMMENFASSUNG EINZELNE PRÄDIKTOREN

In Tabelle 29 sind nochmals alle Rangkorrelationskoeffizienten und Bestimmtheitsmasse aufgeführt. Darin enthalten sind sowohl die Analysen mit der mittleren Verwandtschaftsdistanz als auch die PNN-Distanz als abhängige Variable. Zudem sind die Auswertungen der PNN-Distanz ebenfalls mit dem Datensatz AUTOTYP durchgeführt worden. Die Resultate dazu sind ebenso in dieser Tabelle zu finden.

Tabelle 29: Übersicht über die Rangkorrelationskoeffizienten ρ und deren Signifikanz der mittleren Verwandtschaftsdistanz und der PNN-Distanz für jeden einzelnen geographischen Faktor (sowohl für Hammarström- als auch AUTOTYP-Daten). Die grössten Werte sind fett hervorgehoben. Kursive Werte sind nicht signifikant. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 "****"<0.1 "****"<1.

Geo. Faktor	Hammarström-Daten				AUTOTYP-Daten	
	Mittlere Verwandtschaftsdistanz		PNN-Distanz		PNN-Distanz	
	ρ	Signifi- kanz	ρ	Signifi- kanz	ρ	Signifi- kanz
Breitengrad	0.294	***	0.274	***	0.250	***
Temperatur	-0.150	***	-0.134	***	-0.075	**
Temperaturschwankungsbereich	0.413	***	0.397	***	0.330	***
Niederschlag	-0.381	***	-0.366	***	-0.360	***
Saisonalität Niederschlag	0.139	***	0.115	***	0.032	
NPP	-0.337	***	-0.326	***	-0.334	***
Wachstumsphase	-0.362	***	-0.338	***	-0.337	***
Küstendistanz	0.290	***	0.303	***	0.186	***
Höhenmodell	-0.008		-0.018		-0.192	***
Höhendifferenz	-0.278	***	-0.280	***	-0.421	***

Generell fällt an den Zahlen in Tabelle 29 auf, dass die Korrelation der mittleren Verwandtschaftsdistanz meistens leicht höher ist als die der PNN-Distanz der Hammarström-Daten. Die PNN-Distanz der AUTOTYP-Daten zeigen meist deutlich tiefere Werte als die beiden genealogischen Diversitätsmasse der Hammarström-Daten. Die Richtung der Zusammenhänge ist bei allen die gleiche. Die mittlere Verwandtschaftsdistanz und die PNN-Distanz der Hammarström-Daten korrelieren eindeutig mit dem Temperaturschwankungsbereich am besten. Bei vier Faktoren (Temperaturschwankungsbereich, Niederschlag, Wachstumsphase, NPP) ist die Korrelation mit der mittlere Verwandtschaftsdistanz grösser als 0.3. Da diese jedoch untereinander korrelieren, erklären sie zum Teil ähnliche Anteile wie sich bei der multiplen Regressionsanalyse zeigen wird. Dass die Zusammenhänge der PNN-Distanz jeweils leicht tiefer sind als jene der Verwandtschaftsdistanz, lässt sich mit der Berechnungsmethode der Werte erklären. Im Gegensatz zur PNN-Distanz handelt es sich bei den mittleren Verwandtschaftsdistanzen um einen Durchschnittswert. Solche Werte korrelieren automatisch stärker.

Weshalb die PNN-Distanzen der AUTOTYP-Daten deutlich weniger stark mit den geographischen Faktoren korrelieren ist schwieriger zu erklären. Dabei muss bedacht werden, dass nur für rund 1'800 Sprachen die Distanzen berechnet werden konnten. Aus diesem Grund, wie in Abschnitt 11.2.2 in der Häufigkeitsverteilung gesehen werden konnte, ist der Median mit 94 km deutlich höher als bei den Hammarström-Daten, wo er 56 km beträgt. Was wahrscheinlich den Unterschied in den Zusammenhängen am stärksten beeinflusst ist die ungleiche Verteilung der Sprachen der AUTOTYP- und der Hammarström-Daten auf der Welt. Während beide in Europa und Nordamerika etwa gleich viele Sprachen aufweisen, hat es beispielsweise in Zentralafrika in den Hammarström-Daten vier Mal mehr Sprachen. Ungefähr das gleiche Bild zeigt sich auch bei der Sprachenanzahl in Südostasien. Das heisst, dass zwar in Europa und Nordamerika die PNN-Distanzen in etwa dieselben sind, in Regionen nahe dem

Äquator sich diese in den beiden Datensätzen jedoch deutlich unterscheiden. Wenn nun davon ausgegangen wird, dass geographische Faktoren einen Effekt auf die Distanz zwischen verwandten Sprachen haben, ist es zentral, dass der Anteil der erfassten Sprachen überall auf der Welt gleich ist. Wäre dem so, würden sich wohl trotz der unterschiedlichen Taxonomie und den grösseren PNN-Distanzen, in etwa gleich starke oder schwache Korrelation wie bei den Hammarström-Daten ergeben.

17.12 MULTIPLE REGRESSIONSANALYSE

In diesem Unterkapitel sollen nun nicht mehr nur die Zusammenhänge zwischen einem einzelnen geographischen Faktor und eines der genealogischen Diversitätsmasse betrachtet werden, sondern es wird analysiert, wie viel alle Faktoren zusammen erklären können.

17.12.1 Mittlere Verwandtschaftsdistanz

In Tabelle 30 sind die Resultate der multiplen Regressionsanalyse mit der mittleren Verwandtschaftsdistanz als abhängige Variable aufgeführt. Wie schon bei der Sprachendiversität mussten einige Faktoren sowie die Verwandtschaftsdistanz selbst transformiert werden damit sie ungefähr normalverteilt sind.

Tabelle 30: Multiple Regressionsanalyse mit der mittleren Verwandtschaftsdistanz als abhängige Variable und den geographischen Faktoren als unabhängigen Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 "****"<0.1 "****"<1.

Geo. Faktor	Trans-formation	Koeffizienten	Skalierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		2.008	1.972	< 2e-16	***
Temperatur		-0.013	-0.088	< 2e-16	***
Temperaturschwankungsbereich	log10	0.789	0.154	< 2e-16	***
Niederschlag	√	-0.003	-0.032	1.32E-02	*
Saisonalität Niederschlag	√	-0.027	-0.060	5.5E-10	***
NPP		0.028	0.036	1.16E-03	**
Wachstumsphase		0.000	-0.044	2.87E-03	**
Küstendistanz	√	0.006	0.068	2.53E-14	***
Höhenmodell+1	log10	-0.079	-0.051	7.57E-08	***
Höhendifferenz	√	-0.008	-0.127	< 2e-16	***

p-Wert	<2E-16	***
Bestimmtheitsmass R ²	0.268	

Das Modell kann 26.8 % der Streuung der Distanzen erklären. Wird dieses multiple Regressionsmodell um die geographische Breite ergänzt, steigt das korrigierte Bestimmtheitsmass kaum an. Die Breitengrade tragen somit kaum Neues zur Erklärung der mittleren Verwandtschaftsdistanz bei. Die Korrelation mit der geographischen Breite kann auf die gewählten geographischen Faktoren im Modell zurückgeführt werden. Alle Faktoren die in Zusammenhang mit dem Klima stehen (die ersten sechs) erklären in einem separaten multiplen Regressionsmodell 20.52 % der Streuung. Die Lage-Faktoren erreichen einen Wert von 11.86 %. Gewisse Erklärungsanteile überschneiden sich dementsprechend, was nicht überraschend ist, da die beiden Typen von Faktoren zum Teil leicht korrelieren. Wird zusätzlich die Interaktion der Subsistenzstrategie mit den geographischen Faktoren berücksichtigt, steigt das korrigierte Bestimmtheitsmass lediglich um 0.016.

Aufgrund der Multikollinearität wurden der Temperaturschwankungsbereich und die Länge der Wachstumsphase aus dem Modell entfernt. Eine schrittweise Analyse der AIC-Werte ergab, dass die NPP und die Saisonalität des Niederschlags keine zusätzliche Erklärung zum Modell liefern. So hat es im Modell in Tabelle 31 nur noch fünf geographische Faktoren, die zusammen dennoch 25 % der Varianz der mittleren Verwandtschaftsdistanz erklären können. Den grössten Einfluss hat die Temperatur, gefolgt von der Höhendifferenz. Je höher diese beiden Faktoren und auch der Niederschlag sowie die Höhe über Meeresspiegel sind, desto näher liegen die nächsten verwandten Sprachen beieinander. Je weiter entfernt von der Küste eine Sprache liegt, desto grösser ist die mittlere Verwandtschaftsdistanz.

Tabelle 31: Multiple Regressionsanalyse mit der logarithmierten mittleren Verwandtschaftsdistanz als abhängige Variable und fünf geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <math><0.001</math> "****" <math><0.01</math> "****" <math><0.05</math> " " <math><0.1</math> " " <math><1</math>.

Geo. Faktor	Transformation	Koeffizienten	Skalierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		3.175	1.972	<math>< 2e-16</math>	***
Temperatur		-0.023	-0.159	<math>< 2e-16</math>	***
Niederschlag	√	-0.008	-0.097	<math>< 2e-16</math>	***
Küstendistanz	√	0.009	0.095	<math>< 2e-16</math>	***
Höhenmodell+1	log10	-0.103	-0.067	8.85E-13	***
Höhendifferenz	√	-0.008	-0.128	<math>< 2e-16</math>	***

p-Wert	<math><2E-16</math>	***
Bestimmtheitsmass R ²	0.250	

17.12.2 PNN-Distanz – Hammerström-Daten

Die Resultate des multiplen Regressionsmodells mit der Distanz zur nächsten verwandten Sprache sind in Tabelle 32 zu finden. Die PNN-Distanz wurde für diese Auswertung mit log10 transformiert. Das Modell erklärt 26.3 % der gesamten Streuung der PNN-Distanz. Das ist praktisch gleich viel wie bei der mittleren Verwandtschaftsdistanz (26.8 %). Wird die geographische Breite im Modell ergänzt, ändert sich das korrigierte Bestimmtheitsmass kaum. Die Breitengrade liefern somit keine zusätzlichen Erklärungen. Alle mit dem Klima im Zusammenhang stehenden Faktoren erklären 19.6 % und alle Lagefaktoren 12.9 % der Varianz der PNN-Distanz. Wird im Modell noch eine Interaktion der geographischen Faktoren mit der Subsistenzstrategie ergänzt, steigt das korrigierte Bestimmtheitsmass lediglich um 1.5 Prozentpunkte.

Tabelle 32: Multiple Regressionsanalyse mit der logarithmierten PNN-Distanz als abhängige Variable und den geographischen Faktoren als unabhängigen Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 " " <0.1 " " <1.

Geo. Faktor	Trans-formation	Koeffizienten	Skalierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		1.800	1.772	< 2e-16	***
Temperatur		-0.012	-0.081	3.24E-13	***
Temperaturschwankungsbereich	log10	0.816	0.160	< 2e-16	***
Niederschlag	√	-0.003	-0.031	1.96E-02	*
Saisonalität Niederschlag	√	-0.036	-0.081	2.32E-15	***
NPP		0.025	0.033	5.15E-03	**
Wachstumsphase		0.000	-0.038	1.42E-02	*
Küstendistanz	√	0.008	0.088	< 2e-16	***
Höhenmodell+1	log10	-0.083	-0.054	6.32E-08	***
Höhendifferenz	√	-0.008	-0.131	< 2e-16	***

p-Wert	<2E-16	***
Bestimmtheitsmass R ²	0.263	

In Tabelle 33 ist das Modell mit der reduzierten Multikollinearität dargestellt. Zusätzlich zu den beiden Faktoren Temperaturschwankungsbereich und Länge der Wachstumsphase, wurde nach einer schrittweisen Überprüfung der AIC-Werte auch noch die NPP entfernt. Den stärksten Einfluss auf die PNN-Distanz hat, wie schon bei der mittleren Verwandtschaftsdistanz, die Temperatur, gefolgt von der Höhendifferenz.

Tabelle 33: Multiple Regressionsanalyse mit der logarithmierten PNN-Distanz als abhängige Variable und sechs geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 " " <0.1 " " <1.

Geo. Faktor	Trans-formation	Koeffizienten	Skalierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		3.026	1.772	< 2e-16	***
Temperatur		-0.022	-0.153	< 2e-16	***
Niederschlag	√	-0.008	-0.099	< 2e-16	***
Saisonalität Niederschlag	√	-0.010	-0.023	0.0067	**
Küstendistanz	√	0.011	0.118	< 2e-16	***
Höhenmodell+1	log10	-0.108	-0.070	7.96E-13	***
Höhendifferenz	√	-0.008	-0.131	< 2e-16	***

p-Wert	<2E-16	***
Bestimmtheitsmass R ²	0.245	

17.12.3 PNN-Distanz – AUTOTYP-Daten

In der nachfolgenden Tabelle 34 sind die Resultate der multiplen Regressionsanalyse der PNN-Distanz der AUTOTYP-Daten aufgeführt. Dieses Modell beinhaltet lediglich 1'780 Sprachen im Gegensatz zum Modell der PNN der Hammarström-Daten, welches über 5'500 berücksichtigt. Die geographischen Faktoren können insgesamt 25 % der Streuung der PNN-Distanz erklären. Das ist leicht weniger als bei der PNN-

Distanz der Hammarström-Daten (26.3 %). Die Faktoren, welche im Zusammenhang mit dem Klima stehen (die ersten sechs in Tabelle 34), erklären alleine 14.3 % und die Lagefaktoren (die restlichen drei) 17.5 %. Sowohl bei der PNN-Distanz und auch der mittleren Verwandtschaftsdistanz der Hammarström-Daten, waren es jeweils die Klima-Faktoren, die einen grösseren Anteil erklären konnten.

Tabelle 34: Multiple Regressionsanalyse mit der logarithmierten PNN-Distanz der AUTOTYP-Daten als abhängige Variable und den geographischen Faktoren als unabhängigen Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 ". "<0.1 " "<1.

Geo. Faktor	Transformation	Koeffizienten	Skalierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		2.162	1.949	< 2e-16	***
Temperatur		-0.005	-0.042	0.03824	*
Temperaturschwankungsbereich	log10	0.397	0.082	0.00165	**
Niederschlag	√	-0.003	-0.033	1.59E-01	
Saisonalität Niederschlag	√	-0.009	-0.019	0.21051	
NPP		0.000	0.000	9.91E-01	
Wachstumsphase		0.000	-0.019	4.57E-01	
Küstendistanz	√	0.003	0.034	2.10E-02	*
Höhenmodell+1	log10	-0.038	-0.025	0.12885	
Höhendifferenz	√	-0.011	-0.188	< 2e-16	***

p-Wert	<2E-16	***
Bestimmtheitsmass R ²	0.250	

In der nachfolgenden Tabelle 35 wurden die Faktoren Temperaturschwankungsbereich und Länge der Wachstumsphase entfernt, da diese hohe VIF-Werte aufweisen, wodurch die Multikollinearität reduziert werden konnte. Zusätzlich wurde eine schrittweise Analyse der AIC-Werte gemacht. Dabei wurde festgestellt, dass sich ohne die Faktoren NPP und Saisonalität des Niederschlags das beste Modell ergibt. In diesem korrigierten Modell haben nun auch die Küstendistanz und der Niederschlag einen signifikanten Einfluss auf die PNN-Distanz.

Tabelle 35: Multiple Regressionsanalyse mit der logarithmierten PNN-Distanz der AUTOTYP-Daten als abhängige Variable und sechs geographischen Faktoren als unabhängige Variablen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 ". "<0.1 " "<1.

Geo. Faktor	Transformation	Koeffizienten	Skalierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		2.825	1.949	< 2e-16	***
Temperatur		-0.009	-0.083	7.19E-10	***
Niederschlag	√	-0.007	-0.086	5.20E-10	***
Küstendistanz	√	0.005	0.049	2.13E-04	***
Höhenmodell+1	log10	-0.055	-0.036	0.022276	*
Höhendifferenz	√	-0.011	-0.188	< 2e-16	***

p-Wert	<2E-16	***
Bestimmtheitsmass R ²	0.244	

17.13 MEHREBENENANALYSE

In diesem Unterkapitel werden zwei gemischte Modelle einmal mit und einmal ohne die Interaktion mit der Subsistenzstrategie gerechnet. Diese Modelle sollen zeigen, ob die Sprachfamilie *per se* einen Einfluss darauf hat auf welchem Niveau sich die genealogische Diversität bewegt und welchen Einfluss die Subsistenzstrategie hat. Damit die Einflüsse der einzelnen geographischen Faktoren auf die genealogische Diversität der verschiedenen Subsistenzgruppen interpretiert werden können, werden die Resultate für die Landwirtschaft und die Jäger und Sammler noch einzeln in einer Tabelle abgebildet. Die Auswertungen für die sesshaften Jäger und Sammler und die Hirten sind im Anhang zu finden.

17.13.1 Mittlere Verwandtschaftsdistanz

In Tabelle 36 sind die Resultate der Mehrebenenanalyse der mittleren Verwandtschaftsdistanz aufgeführt. Nachdem der Temperaturschwankungsbereich und die Länge der Wachstumsphase wegen erhöhten VIF-Werten entfernt wurden, wurde eine schrittweise Überprüfung der AIC-Werte durchgeführt. Dabei zeigte sich, dass das Modell besser ist ohne den Faktor NPP. Den grössten Einfluss auf die mittlere Verwandtschaftsdistanz, unter Berücksichtigung der Zufallseffekte der Sprachfamilien, haben der Niederschlag und die Höhendifferenz.

Tabelle 36: Mehrebenenanalyse der mittleren Verwandtschaftsdistanz mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 "****"<0.1 "****"<1.

Geo. Faktor	Transformation	Koeffizienten	Skalierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		2.822	1.917	0	***
Temperatur		-0.009	-0.060	0	***
Niederschlag	√	-0.008	-0.100	0	***
Saisonalität Niederschlag	√	-0.019	-0.042	0	***
Küstendistanz	√	0.005	0.059	0	***
Höhenmodell+1	log10	-0.046	-0.029	0.0018	**
Höhendifferenz	√	-0.006	-0.105	0	***

Anzahl Sprachen	5676
Anzahl Gruppen	230

In Tabelle 37 sind die Resultate der einzelnen Mehrebenenanalysen für die Subsistenztypen Landwirtschaft und Jäger und Sammler aufgeführt. Die Einflüsse der verschiedenen geographischen Faktoren wirken bei allen signifikanten Koeffizienten in dieselbe Richtung. Während bei der Landwirtschaft die Höhendifferenz und der Niederschlag den höchsten Einfluss haben, ist es bei den Jäger- und-Sammler-Gesellschaften die Temperatur. Der Niederschlag hat hingegen keinen signifikanten Einfluss auf die mittlere Verwandtschaftsdistanz.

Tabelle 37: Mehrebenenanalyse der mittleren Verwandtschaftsdistanz aufgeteilt nach Subsistenztyp Landwirtschaft und Jäger und Sammler. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "***"<0.05 " "<0.1 " "<1.

Geo. Faktor	Landwirtschaft			Jäger und Sammler		
	Koeffizienten	Standardisierte Koeffizienten	Signifikanz	Koeffizienten	Standardisierte Koeffizienten	Signifikanz
(Achsenabschnitt)	2.846	1.945	***	2.655	2.191	***
Temperatur	-0.007	-0.045	***	-0.013	-0.131	***
Niederschlag	-0.008	-0.098	***	0.000	-0.004	
Saisonalität Niederschlag	-0.019	-0.042	***	-0.031	-0.069	***
NPP	0.013	0.017		-0.051	-0.059	*
Küstendistanz	0.005	0.055	***	0.007	0.067	**
Höhenmodell+1	-0.059	-0.037	***	0.044	0.027	
Höhendifferenz	-0.006	-0.103	***	-0.004	-0.051	*

Anzahl Sprachen	4834			533		
Anzahl Gruppen	105			73		

17.13.2 PNN-Distanz – Hammarström-Daten

Die Resultate einer Mehrebenenanalyse für die PNN-Distanz sind in Tabelle 38 abgebildet. Die NPP wurde entfernt, da eine schrittweise Analyse der AIC-Werte ergeben hat, dass so das beste Modell erreicht wird. Wie schon bei der mittleren Verwandtschaftsdistanz haben die Höhendifferenz und der Niederschlag den grössten Einfluss auf die PNN-Distanz. Durch die Berücksichtigung der Sprachfamilie ist der Einfluss der Temperatur gesunken.

Tabelle 38: Mehrebenenanalyse der PNN-Distanz mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "***"<0.05 " "<0.1 " "<1.

Geo. Faktor	Transformation	Koeffizienten	Skalierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		2.773	1.801	0	***
Temperatur		-0.010	-0.072	0	***
Niederschlag	√	-0.009	-0.110	0	***
Saisonalität Niederschlag	√	-0.021	-0.046	0	***
Küstendistanz	√	0.007	0.081	0	***
Höhenmodell+1	log10	-0.045	-0.029	0.0034	**
Höhendifferenz	√	-0.007	-0.112	0	***

Anzahl Sprachen	5616
Anzahl Gruppen	230

Damit die Unterschiede der Einflüsse der geographischen Faktoren je nach Subsistenztyp analysiert werden können, sind in Tabelle 39 je ein Modell für die Landwirtschaft und ein Modell für die Jäger und Sammler aufgeführt. Während bei der Landwirtschaft die Höhendifferenz und der Niederschlag grossen Einfluss haben, sind es bei den Jägern und Sammlern die Küstendistanz und die Saisonalität des Niederschlags.

Tabelle 39: Mehrebenenanalyse der PNN-Distanz aufgeteilt nach Subsistenztyp Landwirtschaft und Jäger und Sammler. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 ". "<0.1 " "<1.

Geo. Faktor	Landwirtschaft			Jäger und Sammler		
	Koeffizienten	Standardisierte Koeffizienten	Signifikanz	Koeffizienten	Standardisierte Koeffizienten	Signifikanz
(Achsenabschnitt)	2.831	1.815	***	2.437	2.100	***
Temperatur	-0.010	-0.062	***	-0.007	-0.070	*
Niederschlag	-0.009	-0.103	***	-0.005	-0.054	.
Saisonalität Niederschlag	-0.021	-0.046	***	-0.035	-0.078	***
NPP	0.004	0.005		-0.010	-0.011	
Küstendistanz	0.007	0.076	***	0.010	0.099	***
Höhenmodell+1	-0.064	-0.041	***	0.078	0.049	*
Höhendifferenz	-0.007	-0.111	***	-0.003	-0.044	.

Anzahl Sprachen	4778			531		
Anzahl Gruppen	105			73		

17.13.3 PNN-Distanz – AUTOTYP-Daten

In Tabelle 40 sind die Resultate einer Mehrebenenanalyse mit Sprachfamilien-Gruppen abgebildet. Die meisten Koeffizienten erreichen das Signifikanzniveau von 0.001 nicht. Einzig die Höhendifferenz und der Niederschlag haben einen signifikanten Einfluss auf die PNN-Distanz. Da in der AUTOTYP-Datenbank jeweils nur vermerkt ist, ob die Subsistenzstrategie von Sprechern einer Sprache *Jäger und Sammler* ist oder nicht und dieses Attribut längst nicht bei allen ergänzt ist, wurde die Subsistenzstrategie nicht in die Mehrebenenanalyse einbezogen.

Tabelle 40: Mehrebenenanalyse der PNN-Distanz der AUTOTYP-Daten mit den Sprachfamilien-Gruppen. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 ". "<0.1 " "<1.

Geo. Faktor	Transformation	Koeffizienten	Skalierte Koeffizienten	p-Wert	Signifikanz
(Achsenabschnitt)		2.477	1.928	0	***
Temperatur		-0.003	-0.024	0.239	
Niederschlag	√	-0.008	-0.098	0	***
Saisonalität Niederschlag	√	0.012	0.025	0.089	.
NPP		-0.019	-0.025	0.171	
Küstendistanz	√	0.003	0.029	0.062	.
Höhenmodell+1	log10	0.011	0.007	0.672	
Höhendifferenz	√	-0.009	-0.166	0	***

Anzahl Sprachen	1780
Anzahl Gruppen	68

17.14 MODELLVERGLEICH

17.14.1 Mittlere Verwandtschaftsdistanz

In Tabelle 41 sind die AIC-Werte der verschiedenen Modelle der mittleren Verwandtschaftsdistanz dargestellt. Alle Modelle wurden miteinander verglichen (ANOVA-Test) und dabei zeigte sich, dass alle signifikant unterschiedlich voneinander sind. Das Modell, welches die Daten am besten beschreibt, ist wiederum das Mehrebenenmodell mit den Sprachfamilien-Gruppen inklusive der Interaktion zwischen den Subsistenztypen und der geographischen Faktoren. Somit beeinflussen zufällige Faktoren, die im Zusammenhang mit der Zugehörigkeit zu einer Sprachfamilie stehen, die mittlere Verwandtschaftsdistanz.

Tabelle 41: Vergleich verschiedener Modelle zur mittleren Verwandtschaftsdistanz: Zwei multiple Regressionsmodelle und zwei Mehrebenenanalysen jeweils mit und ohne Subsistenzfaktor als Interaktion.

AIC	Multiples Regressionsmodell		Mehrebenenanalyse	
	Multiples Regressionsmodell	Multiples Regressionsmodell inkl. Interaktion Subsistenzstrategie	Sprachfamilie	Sprachfamilie inkl. Interaktion Subsistenzstrategie
	7369	7272	6550	6520

17.14.2 PNN-Distanz – Hammarström-Daten

Ein Vergleich zwischen den verschiedenen Modellen in Tabelle 42 zeigt, dass alle Modelle signifikant (p -Wert < 0.0001) unterschiedlich voneinander sind. Das Modell, welches die Daten am besten beschreibt, ist auch hier jenes gemischte Modell mit den Sprachfamilien-Gruppen und der Interaktion der Subsistenzstrategie mit den geographischen Faktoren.

Tabelle 42: Vergleich verschiedener Modelle zur PNN-Distanz: Zwei multiple Regressionsmodelle und zwei Mehrebenenanalysen jeweils mit und ohne Subsistenzfaktor als Interaktion.

AIC	Multiples Regressionsmodell		Mehrebenenanalyse	
	Multiples Regressionsmodell	Multiples Regressionsmodell mit Faktor Subsistenzstrategie	Sprachfamilie	Sprachfamilie, Faktor Subsistenzstrategie
	7752	7665	7126	7092

17.14.3 PNN-Distanz – AUTOTYP-Daten

Wie in Tabelle 43 gesehen werden kann, hat auch bei der PNN-Distanz der AUTOTYP-Daten das Mehrebenenmodell den tieferen AIC-Wert. Ein ANOVA-Test zeigte, dass dieses Modell signifikant (p -Wert < 0.0001) besser ist. Somit spielen auch bei der PNN-Distanz der AUTOTYP-Daten zufällige Effekte, die in Zusammenhang mit den Sprachfamilien stehen, eine entscheidende Rolle.

Tabelle 43: Vergleich zwischen dem multiplen Regressionsmodell und der Mehrebenenanalyse der PNN-Distanzen der AUTOTYP-Daten.

	Multiples Regressionsmodell	Mehrebenenanalyse
AIC	2238	2002

17.15 VERGLEICH DER GENEALOGISCHEN DIVERSITÄTS-MASSE

Die mittlere Verwandtschaftsdistanz und die *Phylogenetisch Nächster Nachbar* Distanz stellen ein Mass für die genealogische Diversität dar. Sie untersuchen leicht andere Aspekte. Die PNN-Distanz beschreibt die kürzeste Distanz auf welche sich zwei Sprachen separiert haben. Die mittlere Verwandtschaftsdistanz sagt aus, wie nahe die nächsten Verwandten einer Sprache durchschnittlich sind. Oder anders formuliert, in durchschnittlich welcher Entfernung haben sich Sprachen einer gemeinsamen Protosprache separiert. In Abbildung 92 ist ein Vergleich zwischen diesen beiden genealogischen Sprachmassen zu finden.

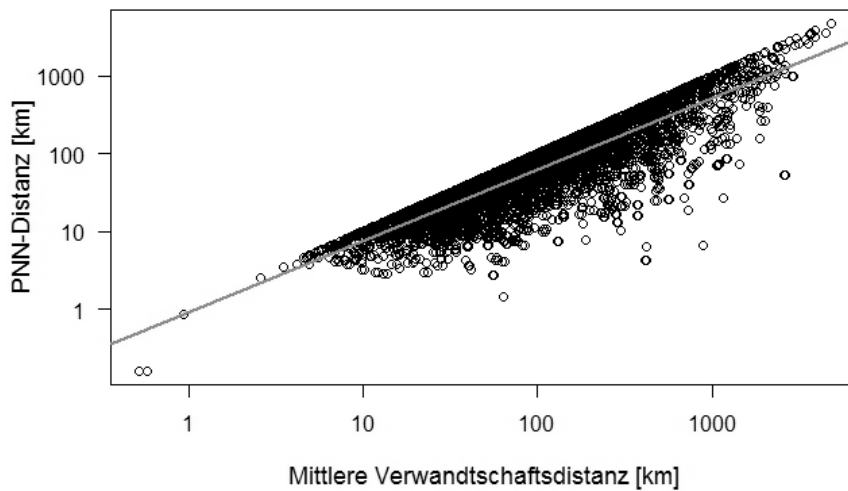


Abbildung 92: Vergleich zwischen der PNN-Distanz und mittleren Verwandtschaftsdistanz.

Die scharfe Kante der Punktwolke in Abbildung 92 ergibt sich aus dem Fakt, dass die PNN-Distanz kleiner, aber nicht grösser als die mittlere Verwandtschaftsdistanz sein kann. So gibt es Ausreisser nach unten, welche eine deutlich tiefere Distanz zum nächsten Verwandten haben als die mittlere Distanz zu allen nächsten Verwandten. Der Rangkorrelationskoeffizient beträgt 0.88.

In Tabelle 44 ist ein Vergleich zwischen den gemischten Modellen der beiden Masse zu sehen. Die standardisierten Koeffizienten sind sich sehr ähnlich. Bei beiden hat die NPP keinen signifikanten Einfluss und das Höhenmodell nur mit einem Niveau von 0.01. Sowohl auf die mittlere Distanz zu den nächsten verwandten Sprachen als auch auf die Distanz zur nächsten verwandten Sprache, haben die geographischen Faktoren den gleichen Einfluss.

Tabelle 44: Vergleich der Mehrebenenanalyse der mittleren Verwandtschaftsdistanz und der PNN-Distanz. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 " " <0.1 " " <1.

Geo. Faktor	Mehrebenenanalyse - Sprachfamilien					
	mittlere Verwandtschaftsdistanz			PNN-Distanz		
	Koeffizienten	Standard-isierte Koeffizienten	Signifi-kanz	Koeffizienten	Standard-isierte Koeffizienten	Signifi-kanz
(Achsenabschnitt)	2.812	1.920	***	2.774	1.801	***
Temperatur	-0.009	-0.059	***	-0.010	-0.072	***
Niederschlag	-0.008	-0.104	***	-0.009	-0.108	***
Saisonalität Niederschlag	-0.018	-0.041	***	-0.021	-0.047	***
NPP	0.006	0.008		-0.003	-0.004	
Küstendistanz	0.006	0.060	***	0.007	0.080	***
Höhenmodell+1	-0.045	-0.029	**	-0.044	-0.029	**
Höhendifferenz	-0.006	-0.106	***	-0.007	-0.112	***
Anzahl Sprachen	5616			5616		
Anzahl Gruppen	230			230		

17.16 INTERPRETATION

17.16.1 Masse für genealogische Diversität

Bevor eine Interpretation der Resultate vorgenommen wird, werden noch Überlegungen zu den Massen respektive Methoden, welche für die genealogische Diversität verwendet wurden, angestellt. Wie im vorherigen Abschnitt 17.15 gesehen werden konnte, korrelieren die beiden Masse *mittlere Verwandtschaftsdistanz* und *PNN-Distanz* stark. Es gibt Sprachen, die bei beiden Massen denselben Wert haben, nämlich dann, wenn es auf der untersten Verwandtschaftsebene nur zwei Sprachen hat. Weiterhin gibt es Sprachen, die relativ grosse Unterschiede zwischen den Massen aufweisen. Bezüglich der geographischen Faktoren wurde erwartet, wenn davon ausgegangen wird, dass sie einen Einfluss auf die Sprachendiversifikation haben, dass die PNN-Distanz diesen Einfluss stärker widerspiegeln würde. Dies deshalb, weil die PNN-Distanz ein kleinräumigeres Phänomen abbildet. Werden alle nächsten verwandten Sprachen einer Sprache betrachtet, so erstrecken sich diese möglicherweise über Regionen mit verschiedenen Umweltbedingungen und innerhalb dieser Gruppe sind die Sprachen möglicherweise sehr unterschiedlich dicht beieinander. Dass der Einfluss der geographischen Faktoren auf die PNN-Distanz grösser ist als auf die mittlere Verwandtschaftsdistanz konnte jedoch nicht gezeigt werden. Im multiplen Regressionsmodell erklären die Faktoren ähnlich grosse Anteile der Streuung (26.8 % und 26.3 %) des jeweiligen Masses. Es könnte auch sein, dass durch die mittlere Verwandtschaftsdistanz Spezialfälle und Ausreisser abgeflacht werden, die möglicherweise durch hier nicht berücksichtigte Faktoren entstehen.

Ein Schwachpunkt dieser beiden Masse, der schwierig zu umgehen ist, ist dass die Sprachdaten die Sprachverteilung zu einem bestimmten Zeitpunkt darstellen. Demzufolge basiert die Sprachendiversifikation, die hier interpretiert wird, auf einem statischen Bild. So wären Sprachen denkbar, die zu einem gewissen Zeitpunkt viele verwandte Sprachen in der Nähe hatten, die dann aber fast alle durch andere verdrängt wurden. Nach diesem Prozess gibt es dann vielleicht nur noch drei nah verwandte Sprachen der ersten Sprache und diese sind alle weit voneinander entfernt, da sie sich nur in Randgebieten halten konnten. Einen solchen Prozess beschrieb Diamond (1997) im Artikel mit dem Namen: *the language steamrollers*.

Ein anderer Schwachpunkt der beiden Masse für die genealogische Diversität ist die starke Abhängigkeit von der Taxonomie oder genauer gesagt von der untersten Taxonomiestufe. Diese muss unbedingt überall auf der Welt nach denselben Kriterien erstellt worden sein. Werden irgendwo Sprachen grosszügiger zur untersten Stufe zusammengefasst, verfälscht das sofort die Resultate. Es scheint dann als würden die mittleren Verwandtschaftsdistanzen in diesem Gebiet grösser sein. Aus diesem Grund wurden auch noch die Daten der AUTOTYP-Datenbank herbeigezogen. Dabei zeigte sich, dass die PNN-Distanz dieser Daten zu 25 % durch die geographischen Faktoren erklärbar ist, was nicht viel weniger ist als die 26.5% der Hammarström-Daten. Auch die Richtung der einzelnen Zusammenhänge (siehe Abschnitt 17.11) ist meistens die gleiche und dort, wo sie abweicht, sind die Koeffizienten der AUTOTYP-Daten nicht signifikant unterschiedlich von Null. Das einzige, das variiert, sind die Grössen der standardisierten Koeffizienten. Der Grund dafür, wie in Abschnitt 17.11 schon ausführlich erläutert, ist die ungleiche globale Verteilung der erfassten Sprachen der AUTOTYP- und der Hammarström-Daten. Leider konnte für die AUTOTYP-Datenbank die mittlere Verwandtschaftsdistanz nicht berechnet werden. Die Gründe dafür wurden bereits in Abschnitt 11.1 dargelegt. Immerhin hat der Vergleich mit der PNN-Distanz der Hammarström- und AUTOTYP-Daten gezeigt, dass diese unterschiedlichen Taxonomien sich nicht stark auf die Resultate zum Zusammenhang zwischen der Distanz zum nächsten verwandten Nachbarn und geographischen Faktoren auswirken.

17.16.2 Einordnung der Resultate

Currie und Mace (2012) machten die Aussage, dass, wenn man wüsste, welche geographischen Faktoren die kulturelle Gruppendiversifikation beeinflussen, könnte man feststellen, wie sich die Ethnogeneserate bei einer Migration in eine neue Umwelt verändert. In dieser Arbeit wurde versucht genau dieses zu beantworten. Es galt herauszufinden, ob unter gewissen Umweltbedingungen die nächsten verwandten Sprachen weiter oder kürzer entfernt liegen. Da diese Sprachen von einer gemeinsamen Protosprache abstammen, stellen sie ein Abbild der Sprachendiversifikation in der entsprechenden Regionen dar.

Bisher konnte nur nachgewiesen werden, dass in tropischen Regionen die Sprachen auf kürzeren Distanzen weniger ähnlich sind als in nicht-tropischen Regionen (Holman et al. 2007). In dieser Arbeit konnte das insofern bestätigt werden, als festgestellt wurde, dass die mittlere Distanz zu den nächsten Verwandten in tropischen Regionen tiefer ist (siehe *Unterkapitel 17.1*). Dieser Fakt konnte auf gewisse geographische Faktoren zurückgeführt werden. So zeigte sich in einer Mehrebenenanalyse mit Sprachfamilien-Gruppen (siehe *Unterkapitel 17.13*), dass alle verwendeten Faktoren, ausser das Höhenmodell, einen signifikanten Einfluss (p -Wert < 0.001) auf die mittlere Verwandtschaftsdistanzen haben. In diesem Modell wurde berücksichtigt, dass sich die Distanzen in verschiedenen Sprachfamilien auf verschiedenen Niveaus befinden. Dies ist wichtig, da Sprachen von einer Familie sich kulturell ähnlicher sind, was dazu führen kann, dass sie, kulturell bedingt, näher beieinander liegen (Nettle 2009). Das gemischte Modell zeigt für die Höhendifferenz und den Niederschlag den grössten Einfluss auf. Demnach liegen die nächsten verwandten Sprachen bei heterogener Topographie näher beieinander. Auf die Sprachendiversifikation umformuliert würde das bedeuten, dass sich Sprachen in bergigen Regionen auf kürzere Distanzen voneinander separieren. Unterstützt wird diese Aussage auch durch das Resultat, wonach die nächste verwandte Sprache in topographisch extremeren Gebieten näher zu finden ist. Die Höhendifferenz hat zwar den grössten Einfluss auf die mittlere Verwandtschaftsdistanz und auch auf die Distanz zum nächsten Nachbar, aber die restlichen Faktoren haben zusammen natürlich einen genauso wichtigen Einfluss und deren Einflüsse fallen deshalb nicht so hoch aus, da sie untereinander korrelieren. Werden nämlich die Korrelationen für jeden geographischen Faktor einzeln betrachtet, sind es die klimatischen und vom Klima abhängigen Faktoren, welche die grössten Werte aufweisen. Diese können zusammen auch den grösseren Anteil an der mittleren Verwandtschaftsdistanz und der PNN-Distanz erklären als die Lage-Faktoren. Insgesamt können die Faktoren rund 27 % der Streuung der beiden Masse erklären. Wird noch berücksichtigt, dass je nach Subsistenztyp die verschiedenen geographischen Faktoren einen anderen Einfluss haben können, steigt dieser Anteil leicht an, auf rund 28.5 %. Somit kann gesagt werden, dass die Modelle einen Einfluss der geographischen Faktoren auf die Sprachendiversifikation anzeigen. Zusätzlich haben zufällige Effekte, wie kulturelle und soziale Einflüsse, die sich aus der Zugehörigkeit zur gleichen Sprachfamilie ergeben, *per se* schon einen Einfluss auf die Distanz zu den nächsten Verwandten. Dies konnte mit einem Modellvergleich gezeigt werden. Auch sind es wohl kulturelle und historische Einflüsse, die nicht direkt mit einzelnen Sprachfamilien in Zusammenhang stehen, die weitere Anteile an der Diversifikation von Sprachen erklären könnten.

Im *Unterkapitel 17.12* und *17.13* wurde auch gezeigt, dass je nachdem welche Subsistenzstrategie Sprecher einer Sprache verfolgen, andere geographische Faktoren eine Rolle spielen. Es wäre sehr interessant diese Resultate weiter zu interpretieren, da von Currie und Mace (2012) gezeigt wurde, dass die Gebietsflächen von Sprachen bei nicht-landwirtschaftlichen Gesellschaften eine höhere Korrelation mit Umweltfaktoren aufweisen als bei landwirtschaftlichen Gesellschaften und dies möglicherweise auch auf die Sprachendiversifikation zutrifft. Für diese Aussage stützten sie sich allerdings auf nur vier geographische Faktoren. Aus drei Gründen wird hier nicht mehr in diese Daten interpretiert: erstens sind diese verschiedenen Subsistenztypen viel zu ungleichmässig auf der Welt verteilt und haben eine extrem unterschiedliche Anzahl von zugehörigen Sprachen. Zweitens konnte ein Modell mit einer Interaktion der Subsistenztypen mit den geographischen Faktoren nicht viel mehr erklären als ein multiples Regressionsmodell ohne diese Interaktion. Drittens hat das gemischte Modell mit den Sprachfamilien-Gruppen gezeigt, dass viel entscheidender als die Subsistenzstrategie die Sprachfamilie ist.

18 VERGLEICH DER RESULTATE

UNTEREINANDER

Nun wurden die Resultate der Sprachdiversität, der Sprachfamiliendiversität und der genealogischen Diversität besprochen. Diesbezüglich ist es natürlich wichtig, jene Diversitäten auch untereinander zu vergleichen. Einerseits interessiert dabei, wie stark diese Masse untereinander korrelieren. Andererseits soll eine Schlussfolgerung gemacht werden, welches der untersuchten Phänomene denn nun am meisten durch geographische Faktoren beeinflusst wird und bei welchen möglicherweise nicht untersuchte Faktoren wie kulturelle Einflüsse eine grössere Rolle spielen.

18.1 VERGLEICH DER SPRACHENMASSE

In Tabelle 45 sind die Korrelationskoeffizienten zwischen den verschiedenen Sprachenmassen aufgeführt. Wie zu erwarten, sind die Korrelationen zwischen den Massen, die den gleichen Typ von Diversität (z.B. Sprachfamiliendiversität) beschreiben sollen, jeweils am stärksten. Eine Ausnahme bildet hier die GNNnv-Distanz, welche ohne die grossen Sprachfamilien eine höhere Korrelation mit der Sprachendichte und der GNN-Distanz aufweist als mit der Sprachfamiliendichte. Dies ist natürlich darauf zurückzuführen, dass in diesen reduzierten Daten die Fälle entfernt werden in denen sich die Sprachen- und die Sprachfamiliendiversität extrem unterscheiden. Während die beiden genealogischen Diversitätsmasse eine relativ hohe Korrelation mit der Sprachdiversität aufweisen, korrelieren sie wenig mit der Sprachfamiliendiversität. Dort wo die nächsten verwandten Sprachen nicht weit auseinander liegen, ist logischerweise auch die Sprachdiversität relativ hoch. Die umgekehrte Schlussfolgerung, dass die genealogische Diversität hoch ist, wenn die Sprachdiversität hoch ist, muss jedoch nicht immer stimmen.

Tabelle 45: Korrelation zwischen den verschiedenen Sprachenmasse. Für die lokale Sprachfamiliendichte und die GNNnv-Distanz ist jeweils der erste Wert mit und der zweite ohne die grossen Sprachfamilien. Werte ≥ 0.5 resp. ≤ -0.5 sind fett hervorgehoben.

	Lokale Sprachen-dichte	GNN-Distanz	Lokale Sprachfamilie ndichte	GNNnv-Distanz	Mittlere Verwandtschaftsdistanz	PNN-Distanz
Lokale Sprachendichte		-0.83	0.39 / 0.51	-0.45 / -0.66	-0.74	-0.75
GNN-Distanz	-0.83		-0.3 / -0.42	0.41 / 0.7	0.69	0.75
Lokale Sprachfamiliendichte	0.39 / 0.51	-0.3 / -0.42		-0.76 / -0.58	-0.34 / -0.49	-0.29 / -0.43
GNNnv-Distanz	-0.45 / -0.66	0.41 / 0.7	-0.76 / -0.58		0.28 / 0.44	0.25 / 0.41
Mittlere Verwandtschaftsdistanz	-0.74	0.69	-0.34 / -0.49	0.28 / 0.44		0.88
PNN-Distanz	-0.75	0.75	-0.29 / -0.43	0.25 / 0.41	0.88	

In Tabelle 46 ist ein Vergleich zwischen den standardisierten Koeffizienten der Mehrebenenanalysen mit den Sprachfamilien-Gruppen zu sehen. Da nicht alle Masse gleich viele Sprachen untersucht haben, wurde

die Auswertung so angepasst, dass alle die gleichen Sprachen beinhalten. Für die Sprachfamiliendichte und die GNNv-Distanz wurden deshalb die Sprachen der grossen Sprachfamilien ebenfalls in die Auswertungen mit einbezogen. Die Einflüsse der verschiedenen geographischen Faktoren sind zwar unterschiedlich stark, wirken jedoch in den meisten Fällen in dieselbe Richtung. Die Temperatur, der Niederschlag und die Höhendifferenz haben auf alle Masse einen signifikanten Einfluss. Je höher diese Faktoren ausfallen, desto höher ist die jeweilige Diversität. Während mit steigender Distanz zur Küste die Sprachendiversität und Sprachendiversifikation abnimmt, nimmt die Sprachfamiliendiversität zu. Dies ist jedoch höchstwahrscheinlich auf eine *net suppression* zurückzuführen (Wuensch 2012). Denn in einem direkten Vergleich nur mit der Küstendistanz nimmt auch die Sprachfamiliendichte von der Küste weg ab.

Tabelle 46: Vergleich der standardisierten Koeffizienten der Mehrebenenanalyse zwischen den verschiedenen Sprachenmasse. Kursive Werte erreichen das Signifikanzniveau von 0.001 nicht.

Geo. Faktor	Standardisierte Koeffizienten					
	Lokale Sprachen-dichte	GNN-Distanz	Lokale Sprach-familien-dichte	GNNv-Distanz	Mittlere Verwandtschafts-distanz	PNN-Distanz
(Achsenabschnitt)	0.880	1.498	1.016	1.714	1.920	1.801
Temperatur	0.101	-0.096	0.124	-0.186	-0.059	-0.072
Niederschlag	0.160	-0.092	0.050	-0.061	-0.105	-0.110
Saisonalität Niederschlag	0.085	-0.055	-0.024	-0.026	-0.042	-0.048
NPP	-0.005	-0.015	-0.066	0.073	0.006	-0.005
Küstendistanz	-0.042	0.041	0.037	-0.034	0.060	0.081
Höhenmodell+1	-0.001	0.006	0.011	0.018	-0.029	-0.029
Höhendifferenz	0.200	-0.130	0.071	-0.132	-0.106	-0.112

Anzahl Sprachen	5597
Anzahl Gruppen	230

18.2 FAZIT ZU DEN SPRACHENMASSEN

Geographische Faktoren haben am meisten Einfluss auf die Sprachendiversität. Mit einem multiplen Regressionsmodell lassen sich rund 37 % (lokale Sprachendichte) respektive 31 % (GNN-Distanz) der Varianz der Sprachendiversität erklären. Wird noch eine Interaktion der Faktoren mit der Subsistenzstrategie berücksichtigt, steigt dieser Anteil auf rund 41 % respektive 33 %. Die Streuung der genealogischen Diversität lässt sich hingegen nur zu rund 27 % durch die geographischen Faktoren erklären. Bei Ergänzung mit der Subsistenzstrategie sind es 29 %. Auch die Sprachfamiliendiversität lässt sich nur zu rund 36 % (lokale Sprachfamiliendichte) respektive 28 % durch diese Faktoren erklären. Diese Werte sind zwar nicht viel tiefer als jene der Sprachendiversität, jedoch muss dabei bedacht werden, dass bei der Sprachfamiliendiversität die grossen Sprachfamilien ausgeschlossen wurden. Wird für dieselben Sprachen das Bestimmtheitsmass für die beiden Sprachendiversitätsmasse berechnet, sind diese mit 42 % (Sprachendichte) respektive 37 % (GNN-Distanz) deutlich höher.

Die Sprachfamiliendiversitäten widerspiegeln offenbar ähnliche Auswirkungen der geographischen Faktoren wie die Sprachendiversität, jedoch zu einem anderen Zeitpunkt. So zeigen die Sprachfamilien auf, welche Prozesse vor über 10'000 Jahren die Separation von Sprachen beeinflusst haben, die bis heute zur Separation von Sprachfamilien geführt haben. Jedoch ist das nur der Fall, wenn grosse Sprachfamilien, die sich in den letzten 10'000 Jahren stark ausgebreitet haben, weggelassen werden. Die geographischen Faktoren haben wahrscheinlich deshalb einen grösseren Einfluss auf die Sprachendiversität als auf die

genealogische Diversität, da bei diesem Mass nicht die komplette Nachbarschaft berücksichtigt wird. So kann die Sprachendichte an einem bestimmten Ort hoch sein, aber die Distanzen zu den nächsten Verwandten dennoch relativ weit. Die Gesamtsituation respektive die Sprachendiversität widerspiegeln die Einflüsse der geographischen Faktoren anscheinend besser. Dies bedeutet jedoch nicht, dass Umweltfaktoren sich nicht auch auf die genealogische Diversität auswirken, sondern nur, dass das Modell weniger erklären kann, da nicht die komplette Nachbarschaft berücksichtigt wird.

19 FORSCHUNGSFRAGEN UND ANTWORTEN

Im fünften Teil der Arbeit wurden die Resultate präsentiert und anschliessend interpretiert. So wurden die am Anfang der Arbeit aufgestellten Forschungsfragen (*Unterkapitel 2.2*) indirekt bereits beantwortet. In diesem Kapitel werden diese nun nochmals einzeln sowie kurz und konzise beantwortet um einen Überblick über die Resultate zu erhalten. Zur Erinnerung, die Frage die über allen Auswertungen stand, lautet:

- Gibt es einen Zusammenhang zwischen der weltweiten Verteilung von Sprachen und geographischen Faktoren?

Diese Frage wurde anschliessend in zwei weiteren Fragen konkretisiert:

- Bei welchen geographischen Faktoren oder Konstellationen von Faktoren kann eine hohe Sprachendiversität respektive Sprachfamiliendiversität erwartet werden?
- Gibt es einen Zusammenhang zwischen den geographischen Umweltbedingungen und der genealogischen Diversität?

Die erstere der beiden Fragen wurde mittels der Analyse der Sprachendiversität und der Sprachfamiliendiversität (*Kapitel 15 und 16*) beantwortet. Es konnte gezeigt werden, dass geographische Faktoren rund ein Drittel der Streuung der Sprachendiversität erklären können. Dementsprechend gibt es einen Zusammenhang zwischen der Verteilung der Sprachen und gewissen Umweltbedingungen. Besonders stark beeinflusst wird die Sprachendiversität dabei von der Temperatur, dem Niederschlag und der Höhendifferenz. Je höhere Werte diese Faktoren aufweisen, desto höher ist jeweils auch die Sprachendiversität. Wird zusätzlich noch berücksichtigt, dass je nach Subsistenzstrategie die geographischen Faktoren die Diversität unterschiedlich beeinflussen können, kann diese nochmals besser erklärt werden. Zudem konnte gezeigt werden, dass die Zugehörigkeit zu einer bestimmten Sprachfamilie *per se* die Höhe der Sprachendiversität beeinflusst. Die Sprachfamiliendiversität kann deutlich schlechter mit geographischen Faktoren erklärt werden. Werden jedoch die sieben grössten Sprachfamilien weggelassen, lässt sich wiederum ungefähr ein Drittel der Diversität mit den geographischen Faktoren erklären. Dies ist darauf zurückzuführen, dass sich die meisten grossen Sprachfamilien aufgrund von kulturellen oder historischen Gründen, meist in Zusammenhang mit der Erfindung der Landwirtschaft, so stark ausgebreitet haben (Diamond 1997, Diamond & Bellwood 2003). Somit kann die Sprachfamiliendiversität in der Nähe dieser Sprachen kaum mit geographischen Faktoren erklärt werden. Faktoren, welche in Zusammenhang mit dem Klima stehen, insbesondere die Temperatur, konnten am meisten von der Streuung der Sprachfamiliendiversität erklären. Wird noch die Subsistenzstrategie ins Modell einbezogen, kann nochmals deutlich mehr erklärt werden. Die Sprachfamiliendiversität lässt sich schlechter durch geographische Faktoren erklären als die Sprachendiversität. Dies ist wahrscheinlich darauf zurückzuführen, dass die Sprachfamilien eine länger zurückliegende Zeitperiode widerspiegeln und diese deshalb schlechter durch heutige Klimafaktoren erklärt werden können. Zudem wirkten durch die längere Zeitdauer möglicherweise mehr kulturelle und historische Prozesse.

Die zweite Forschungsfrage legt den Fokus auf die genealogische Diversität, die bisher im Zusammenhang mit geographischen Faktoren noch kaum untersucht wurde. Dabei wurde die Nähe zu den nächsten verwandten Sprachen untersucht, um Aussagen über die Sprachendiversifikation unter bestimmten Umweltbedingungen machen zu können. Mit den gewählten geographischen Faktoren konnte gut ein Viertel der genealogischen Diversität erklärt werden. Wiederum wurde eine Zunahme mit steigender

Temperatur, Niederschlag und Höhendifferenz festgestellt. Eine Abnahme war hingegen mit zunehmender Distanz zur Küste feststellbar. Wurde zusätzlich noch die Subsistenzstrategie berücksichtigt, konnte leicht mehr der Streuung der genealogischen Diversität erklärt werden.

Somit können alle drei untersuchten Phänomene teilweise durch geographische Faktoren erklärt werden. Die nicht erklärten Anteile sind wohl auf kulturelle und historische Einflüsse zurückzuführen. Von allen drei Massen kann die Sprachendiversität am besten durch geographische Faktoren erklärt werden. Die geographischen Faktoren haben wahrscheinlich einen grösseren Einfluss auf die Sprachendiversität als auf die Distanz über die sich zwei Sprachen separieren, da bei diesem Mass nicht die komplette Nachbarschaft berücksichtigt wird.

20 GRENZEN DER GEWÄHLTEN METHODIK

Obwohl sich in dieser Arbeit interessante Resultate zeigten, ist es auch wichtig, dass die Grenzen der gewählten Methode nicht vergessen werden. So musste für die drei untersuchten Phänomene (die Sprachendiversität, die Sprachfamiliendiversität und die genealogische Diversität) ein Mass zur Beschreibung der jeweiligen Diversität gefunden werden, welches oftmals Kompromisse und Annahmen beinhaltet. So stützt sich sowohl die Wahl des Radius der Sprachendiversität wie auch jener der Sprachfamiliendiversität auf Überlegungen und Konsultationen der Häufigkeitsverteilungen. Werden zukünftig regionale Analysen mit diesen Methoden gemacht, ist es unabdingbar, den Radius auf die lokalen Gegebenheiten und den regionalen Massstab anzupassen.

Alle Klimafaktoren, die in dieser Arbeit verwendet wurden, beschreiben die heutige Situation. Untersucht wurde jedoch die Sprachenverteilung vor der Kolonialisierung. Von vielen Faktoren standen keine älteren Datengrundlagen zur Verfügung, weshalb mit heutigen Daten gearbeitet wurde. Bei der Untersuchung der Sprachfamiliendiversität sollten eigentlich noch ältere Klimadaten verwendet werden, da dort Prozesse, die sich über 10'000 Jahre vollzogen haben, untersucht werden.

Obwohl viele Korrelationen der verschiedenen Sprachenmasse mit geographischen Faktoren nachgewiesen werden konnten, muss dabei natürlich bedacht werden, dass diese keine Kausalitäten beweisen.

Viele der verwendeten Daten sind räumlich autokorreliert. Das heisst, dass Daten von Orten, die nahe beieinander liegen, sich ähnlicher sind als Daten die weit entfernt voneinander liegen und somit die Daten nicht zufällig sind (O'Sullivan & Unwin 2010). In dieser Arbeit wurden keine speziell für räumliche Autokorrelation entwickelte statistische Analysen verwendet, was bei der Interpretation der Resultate bedacht werden muss.

21 AUSBLICK

Neben der Optimierung des gewählten Vorgehens hinsichtlich der soeben erwähnten Limiten könnte in einem weiteren Schritt untersucht werden, in welchen Regionen der Welt aufgrund der Umweltfaktoren eine höhere oder tiefere Sprachendiversität erwartet werden würde. Diese Abweichungen würden darauf hindeuten, dass dort möglicherweise historische Ereignisse und kulturelle Faktoren einen entscheidenderen Einfluss genommen haben. Auch könnte analysiert werden, wie sich die Sprachfamiliendiversität, bevor sich einzelne Sprachfamilien stark ausgebreitet und andere Sprachen verdrängt haben, verhielt. Beispielsweise könnte so die mögliche Verteilung der Familien in Zentralafrika vor der Expansion der Niger-Kongo Sprachen rekonstruiert werden. Zudem liefern die Erkenntnisse dieser Arbeit auch Informationen für die Rekonstruktion der Ursprungsregionen von Sprachfamilien.

22 LITERATURVERZEICHNIS

- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R., 2011. *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. 13 Hrsg. Heidelberg: Springer.
- Beyer, H. L., 2009-2012. *Geospatial Modelling Environment (Version 0.7.2.1)*. [Software]. [Online] <http://www.spataleecology.com/gme/>
[Zugriff am 22.07.2013].
- Birdsell, J. B., 1953. Some Environmental and Cultural Factors Influencing the Structuring of Australian Aboriginal Populations. *The American Naturalist*, 87 (834), 171-207.
- Cashdan, E., 2001. Ethnic diversity and its environmental determinants: effects of climate, pathogens, and habitat diversity. *American Anthropologist*, 103 (4), 968-991.
- CGIAR-CSI, 2008. *SRTM 90m Digital Elevation Data*. [Online]. Consortium for Spatial Information. <http://srtm.csi.cgiar.org/>
[Zugriff am 16.07.2013].
- Clark, P. J. & Evans, F. C., 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, 35 (4), 445-453.
- Collard, I. F. & Foley, R. A., 2002. Latitudinal patterns and environmental determinants of recent human cultural diversity: do humans follow biogeographical rules? *Evolutionary Ecology Research*, 4 (3), 371-383.
- Currie, T. E. & Mace, R., 2009. Political complexity predicts the spread of ethnolinguistic groups. *Proceedings of the National Academy of Sciences*, 106(18), pp. 7339-7344.
- Currie, T. E. & Mace, R., 2012. The evolution of ethnolinguistic diversity. *Advances in Complex Systems*, 15 (01&02).
- Diamond, J. & Bellwood, P., 2003. Farmers and their languages: the first expansions. *Science*, 300 (5619), 597-603.
- Diamond, J. M., 1997. Linguistics: the language steamrollers. *Nature*, 389 (6651), 544-546.
- Dolić, D., 2004. *Statistik mit R: Einführung für Wirtschafts- und Sozialwissenschaftler*. Oldenbourg Wissenschaftsverlag.
- Everitt, B. & Hothorn, T., 2011. *An Introduction to Applied Multivariate Analysis with R*. Heidelberg: Springer.
- FAO, 2000. *GeoNetwork: GLOBAL NUMBER OF GROWING PERIOD DAYS*. [Online]. Food and Agriculture Organization of the United Nations. <http://www.fao.org/geonetwork/srv/en/metadata.show?uuid=c45725c0-88fd-11da-a88f-000d939bc5d8&currTab=simple>
[Zugriff am 24.07.2013].
- FAO, 2010. *Land Area*. [Online]. Food and Agriculture Organization of the United Nations. <http://geodata.grid.unep.ch/options.php?selectedID=1685&selectedDatasettype=1>
[Zugriff am 05.07.2013].
- Fischer, G., Nachtergaele, F. O., Prieler, S., van Velthuisen, H., Verelst, L. & Wiberg, D., 2008. *Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)*. [Online] <http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/global-terrain->

[slope.html](#)

[Zugriff am 16.07.2013].

- Hammarström, H., 2012. *The Language Families of the World: A critical synopsis*. (Unveröffentlichtes Manuskript).
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International journal of climatology*, 25 (15), 1965-1978.
- Holman, E. W., Schulze, C., Stauffer, D. & Wichmann, S., 2007. On the relation between structural diversity and geographical distance among languages: observations and computer simulations. *Linguistic Typology*, 11 (2), 393-421.
- Imhoff, M. L., Bounoua, L., Ricketts, T., Loucks, C., Harriss, R. & Lawrence, W.T., 2004. *HANPP Collection: Global Patterns in Net Primary Productivity (NPP)*. [Online]
<http://sedac.ciesin.columbia.edu/data/set/hanpp-net-primary-productivity>
[Zugriff am 07.08.2013].
- Lewis, P. M., Simons, G. F. & Fennig, C. D., 2013. *Ethnologue: Languages of the World, Seventeenth edition*. [Online]
<http://www.ethnologue.com>
[Zugriff am 01.09.2013].
- Mace, R. & Holden, C. J., 2005. A phylogenetic approach to cultural evolution. *Trends in Ecology and Evolution*, 20 (3), 116-121.
- Mace, R. & Jordan, F. M., 2011. Macro-evolutionary studies of cultural diversity: a review of empirical studies of cultural transmission and cultural adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366 (1563), 402-411.
- Mace, R. & Pagel, M., 1994. The Comparative Method in Anthropology. *Current Anthropology*, 35 (5), 549-564.
- Mace, R. & Pagel, M., 1995. A latitudinal gradient in the density of human languages in North America. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 261 (1360), 117-121.
- McGregor, W. B., 2009. *Linguistics: an introduction*. London: Continuum International Publishing Group Ltd.
- Moore, J. L., Manne, L., Brooks, T., Burgess, N. D., Davies, R., Rahbek, C., Williams, P., & Balmford, A., 2002. The distribution of cultural and biological diversity in Africa. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269 (1501), 1645-1653.
- Murdock, G. P. & White, D. R., 2006. *Standard Cross-Cultural Sample: on-line edition*. [Online]
<http://escholarship.org/uc/item/62c5c02n>
[Zugriff am 16.07.2013].
- NASA, 2000-2009. *NEO: Net Primary Productivity (1 month - Terra/MODIS)*. [Online]. National Aeronautics and Space Administration, Earth Observations.
http://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD17A2_M_PSN
[Zugriff am 23.07.2013].
- NASA, 2013. *NEO: Net Primary Productivity (1 month - Terra/MODIS)*. [Online]. National Aeronautics and Space Administration, Earth Observations.
http://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD17A2_M_PSN
[Zugriff am 23.07.2013].

- Nettle, D., 1998. Explaining global patterns of language diversity. *Journal of anthropological archaeology*, 17 (4), 354-374.
- Nettle, D., 1999. *Linguistic Diversity*. New York: Oxford University Press.
- Nettle, D., 2009. Ecological influences on human behavioural diversity: a review of recent findings. *Trends in Ecology and Evolution*, 24 (11), 618-624.
- Nichols, J., 1990. Linguistic diversity and the first settlement of the New World. *Language*, 66 (3), 475-521.
- Nichols, J. & Bickel, B., 2009. *The AUTOTYP Genealogy and Geography Database*. [Online] <http://www.spw.uzh.ch/autotyp/> [Zugriff am 30.08.2013].
- Nordhoff, S., Hammarström, H., Forkel, R. & Haspelmath, M., 2013. *Glottolog 2.0*. [Online] <http://glottolog.org/> [Zugriff am 17.09.2013].
- Openshaw, S., 1983. *The modifiable areal unit problem*. Norwich: Geo books.
- O'Sullivan, D. & Unwin, D. J., 2010. *Geographic information analysis*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & the R Development Core Team, 2013. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-109. [Software] [Online] <http://cran.r-project.org/web/packages/nlme> [Zugriff am 30.08.2013].
- Press, F. & Siever, R., 2003. *Allgemeine Geologie*. 3 Hrsg. München: Elsevier GmbH.
- R Core Team, 2013. *R: A language and environment for statistical computing*. [Software] Wien: R Foundation for Statistical Computing.
- Stepp, J. R., Castaneda, H. & Cervone, S., 2005. Mountains and biocultural diversity. *Mountain Research and Development*, 25 (3), 223-227.
- USGS, 2012. *Global 30 Arc-Second Elevation (GTOPO30)*. [Online]. United States Geological Survey. <https://lta.cr.usgs.gov/GTOPO30> [Zugriff am 16.07.2013].
- Wuensch, K. L., 2012. Redundancy and Suppression in Trivariate Correlation/Regression Analysis. [Online]. <http://core.ecu.edu/psyc/wuenschk/MV/multReg/Suppress.docx> [Zugriff am 25.09.2013].

ANHANG

ANHANG A: FORSCHUNGSANSATZ – ZUSÄTZE

Berechnung der lokalen Sprachendichte 100 km

SQL-Abfrage

```
UPDATE [dbo].[Hammerstroem_Buffer]
SET
    [LanguagesInBuffer_100] =
        ([dbo].[CountLanguagesInBuffer]((select OBJECTID from Daten_Hammerstroem where
OBJECTID = Hammerstroem_Buffer.ID_Hammerstroem)))
```

Dazugehörige Skalarwertfunktion

```
-- =====
-- Author:          Köhli, Martina
-- Create date:     10.04.2013
-- Description:     Counts the Languages which are nearer than 100km from the language
                   with @ID. Doesn't count itself, no new languages and no languages
                   which are not in the Hammarström Taxonomie.
-- =====
CREATE FUNCTION [dbo].[CountLanguagesInBuffer]
(
    -- Add the id of the language
    @ID int
)
RETURNS int
AS
BEGIN
    DECLARE @returnvalue int
    DECLARE @Buffer geography
    DECLARE @LowestFamilyID int

    SET @Buffer = (SELECT geog.STBuffer(100000) FROM Daten_Hammerstroem WHERE OBJECTID =
@ID)
    SET @LowestFamilyID = (SELECT LowestFamilyID FROM Daten_Hammerstroem WHERE OBJECTID
= @ID)

    IF @LowestFamilyID is null
        BEGIN
            SET @returnvalue = null
        END
    ELSE
        SET @returnvalue = (SELECT count(*)
                           FROM Daten_Hammerstroem
                           WHERE @ID != OBJECTID and LowestFamilyID is not null and
                           (SELECT [Status] FROM dbo.[Language] WHERE
                           dbo.Daten_Hammerstroem.Code = dbo.[Language].ShortName) is
                           null and @Buffer.STContains(geog) = 1
                           )

    RETURN @returnvalue
END
```

Berechnung Distanz zum Geographisch Nächsten Nachbar (GNN)

SQL Cursor

```
DECLARE @ID int
DECLARE @ID2 int
DECLARE @ID3 int
DECLARE @geog geography
```

```

DECLARE @geog2 geography
DECLARE @distance float

DECLARE neighbor_cursor CURSOR FOR
SELECT dbo.Daten_Hammerstroem.OBJECTID,
       dbo.Daten_Hammerstroem.geog
FROM   dbo.Daten_Hammerstroem INNER JOIN dbo.[Language] ON dbo.Daten_Hammerstroem.Code =
       dbo.[Language].ShortName
WHERE  dbo.[Language].[Status] is null and dbo.Daten_Hammerstroem.HighestFamilyName is not
       null and dbo.Daten_Hammerstroem.OBJECTID

OPEN neighbor_cursor
FETCH NEXT FROM neighbor_cursor INTO @ID, @geog
WHILE @@FETCH_STATUS = 0
BEGIN
    DECLARE draw_cursor CURSOR FOR
    SELECT OBJECTID,
           geog
    FROM   Daten_Hammerstroem INNER JOIN dbo.[Language] ON dbo.Daten_Hammerstroem.Code
           = dbo.[Language].ShortName
    WHERE OBJECTID != @ID and dbo.[Language].[Status] is null and
           dbo.Daten_Hammerstroem.HighestFamilyName is not null

    SET @distance = 999999999999999999

    OPEN draw_cursor
    FETCH NEXT FROM draw_cursor INTO @ID2, @geog2
    WHILE @@FETCH_STATUS = 0
        BEGIN
            IF(@geog.STDistance(@geog2) < @distance)
            BEGIN
                set @distance = @geog.STDistance(@geog2)
                set @ID3 = @ID2
            END
            FETCH NEXT FROM draw_cursor INTO @ID2, @geog2
        END
    INSERT INTO Language_Network_GNN
           (IDFrom, IDTo, distance)
    VALUES (@ID, @ID3, @distance)
    CLOSE draw_cursor
    DEALLOCATE draw_cursor
    FETCH NEXT FROM neighbor_cursor INTO @ID, @geog
END
CLOSE neighbor_cursor
DEALLOCATE neighbor_cursor

```

Berechnung der lokalen Sprachfamiliendichte 400 km

SQL-Abfrage

```

UPDATE [dbo].[Hammerstroem_Buffer_NeighbourFamilies]
SET [CountLanguageFamily_400] =
    ([dbo].[CountLanguageFamilies]((SELECT OBJECTID FROM Daten_Hammerstroem WHERE
OBJECTID = [Hammerstroem_Buffer_NeighbourFamilies].ID_Hammerstroem)))

```

Dazugehörige Skalarwertfunktion

```

-- =====
-- Author:          Köhli, Martina
-- Create date:     12.05.2013
-- Description:     Counts the language families which are nearer than 400km from the
--                  language with @ID.
--                  No null, not itself and not itselfs family.
-- =====
CREATE FUNCTION [dbo].[CountLanguageFamilies]
(
    --Add the ID of the language
    @ID int
)

```

```

RETURNS int
AS
BEGIN
    DECLARE @returnvalue int
    DECLARE @Buffer geography
    DECLARE @HighestFamilyName varchar(50)

    SET @Buffer = (SELECT geog.STBuffer(400000) FROM Daten_Hammerstroem WHERE OBJECTID =
@ID)
    SET @HighestFamilyName = (SELECT HighestFamilyName FROM Daten_Hammerstroem WHERE
OBJECTID = @ID)

    IF @HighestFamilyName is null
        BEGIN
            SET @returnvalue = null
        END
    ELSE
        SET @returnvalue = (
            SELECT COUNT(DISTINCT HighestFamilyName)
            FROM Daten_Hammerstroem
            WHERE @ID != OBJECTID and HighestFamilyName is not null and
HighestFamilyName != @HighestFamilyName and @Buffer.STContains(geog)
= 1
        )
    RETURN @returnvalue
END

```

Berechnung mittlere Verwandtschaftsdistanz und PNN-Distanz

SQL Cursor zur Berechnung der Linien zwischen den Verwandten

```

Declare @ID int
Declare @LanguageFamilyID int
Declare @x1 float
Declare @y1 float
Declare @x2 float
Declare @y2 float
Declare @ShortName varchar(500)
Declare @ShortName1 varchar(500)

DECLARE sibling_cursor CURSOR FOR
SELECT  dbo.Daten_Hammerstroem.OBJECTID,
        dbo.Daten_Hammerstroem.Code,
        dbo.Language.LanguageFamilyID,
        dbo.Daten_Hammerstroem.x,
        dbo.Daten_Hammerstroem.y
FROM    dbo.Daten_Hammerstroem INNER JOIN
        dbo.Language ON dbo.Daten_Hammerstroem.Code = dbo.Language.ShortName
WHERE   dbo.Language.Status is null --no new languages

OPEN sibling_cursor
FETCH NEXT FROM sibling_cursor INTO @ID, @ShortName, @LanguageFamilyID, @x1, @y1
WHILE @@FETCH_STATUS = 0
BEGIN
    SET @ShortName1 = @ShortName

    DECLARE draw_cursor CURSOR FOR
    SELECT Code, x, y
    FROM    Daten_Hammerstroem INNER JOIN
            dbo.Language ON dbo.Daten_Hammerstroem.Code = dbo.Language.ShortName
    WHERE   LowestFamilyID = @LanguageFamilyID and OBJECTID != @ID and Language.Status
is null

    OPEN draw_cursor
    FETCH NEXT FROM draw_cursor INTO @ShortName, @x2, @y2
    WHILE @@FETCH_STATUS = 0
    BEGIN
        INSERT INTO Language_Network
            (geog,

```

```

relation,
CodeFrom,
CodeTo,
ID_Hammerstroem)
VALUES
    (geography::STLineFromText('LINESTRING('
CONVERT(NVARCHAR(50), @x1) + ' ' + CONVERT(NVARCHAR(50), @y1) + ', ' + CONVERT(NVARCHAR(50),
@x2) + ' ' + CONVERT(NVARCHAR(50), @y2) + ' )', 4326),
    1,
    @ShortName1,
    @ShortName,
    @ID)
    FETCH NEXT FROM draw_cursor INTO @ShortName, @x2, @y2
END
CLOSE draw_cursor
DEALLOCATE draw_cursor

    FETCH NEXT FROM sibling_cursor INTO @ID, @ShortName, @LanguageFamilyID, @x1, @y1
END
CLOSE sibling_cursor
DEALLOCATE sibling_cursor

```

Select-Statement zur Anzeige der mittleren Distanz

```

SELECT      CodeFrom
            ,AVG(geog.STLength()) AS avg_DistanceNextRelatedNeighbours
            ,ID_Hammerstroem
FROM        dbo.Language_Network
GROUP BY   CodeFrom, ID_Hammerstroem

```

Select-Statement zur Anzeige der PNN-Distanz

```

SELECT      CodeFrom, MIN(geog.STLength()) AS PNN, ID_Hammerstroem
FROM        dbo.Language_Network
GROUP BY   CodeFrom, ID_Hammerstroem

```

ANHANG B: RESULTATE – ZUSÄTZE

Sprachendiversität – Mehrebenenanalyse, restliche Subsistenztypen

Tabelle XLVII: Mehrebenenanalyse der lokalen Sprachendichte aufgeteilt nach Subsistenztyp sesshafte Jäger und Sammler und Hirten. Die Zeichen für die Signifikanz sind: "****" <0.001 "****" <0.01 "****" <0.05 ". "<0.1 " "<1.

Geo. Faktor	Sesshafte Jäger und Sammler			Hirten		
	Koeffizienten	Standard-isierte Koeffizienten	Signifikanz	Koeffizienten	Standard-isierte Koeffizienten	Signifikanz
(Achsenabschnitt)	0.515	1.415	***	-1.141	0.497	***
Temperatur	0.020	0.067	***	0.012	0.114	**
Niederschlag	0.005	0.039	**	0.028	0.192	***
Saisonalität Niederschlag	-0.016	-0.021		0.017	0.025	
NPP	0.025	0.018	*	-0.026	-0.022	
Küstendistanz	0.005	0.019	.	0.001	0.006	
Höhenmodell+1	0.072	0.042	***	0.066	0.036	
Höhendifferenz	0.000	-0.006		0.011	0.171	***
Anzahl Sprachen	299			76		
Anzahl Gruppen	83			4		

Tabelle XLVIII: Mehrebenenanalyse der GNN-Distanz aufgeteilt nach Subsistenztyp sesshafte Jäger und Sammler und Hirten. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "***"<0.05 " "<0.1 " "<1.

Geo. Faktor	Sesshafte Jäger und Sammler			Hirten		
	Koeffizienten	Standard-isierte Koeffizienten	Signifi-kanz	Koeffizienten	Standard-isierte Koeffizienten	Signifi-kanz
(Achsenabschnitt)	1.831	1.248	***	3.354	1.682	***
Temperatur	-0.023	-0.078	**	-0.022	-0.205	**
Niederschlag	-0.001	-0.009		-0.030	-0.204	**
Saisonalität Niederschlag	0.008	0.011		-0.014	-0.020	
NPP	-0.001	0.000		0.046	0.040	
Küstendistanz	0.020	0.075	***	-0.002	-0.016	
Höhenmodell+1	0.006	0.004		-0.005	-0.003	
Höhendifferenz	-0.004	-0.058	*	-0.010	-0.165	***

Anzahl Sprachen	299			76		
Anzahl Gruppen	83			4		

Sprachfamiliendiversität – Mehrebenenanalyse, restliche Subsistenztypen

Tabelle XLIX: Mehrebenenanalyse der lokalen Sprachfamiliendichte aufgeteilt nach Subsistenztyp sesshafte Jäger und Sammler und Hirten. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "***"<0.05 " "<0.1 " "<1.

Geo. Faktor	Sesshafte Jäger und Sammler			Hirten		
	Koeffizienten	Standard-isierte Koeffizienten	Signifi-kanz	Koeffizienten	Standard-isierte Koeffizienten	Signifi-kanz
(Achsenabschnitt)	0.612	1.590	***	0.464	0.598	***
Temperatur	0.034	0.114	***	0.018	0.172	***
Niederschlag	0.000	0.003		-0.007	-0.049	
Saisonalität Niederschlag	0.001	0.001		0.013	0.019	
NPP	-0.034	-0.024	***	0.032	0.027	
Küstendistanz	0.017	0.063	***	0.002	0.020	
Höhenmodell+1	0.051	0.030	***	-0.069	-0.038	
Höhendifferenz	0.000	-0.001		-0.001	-0.022	

Anzahl Sprachen	299			76		
Anzahl Gruppen	83			4		

Tabelle L: Mehrebenenanalyse der GNNv-Distanz aufgeteilt nach Subsistenztyp sesshafte Jäger und Sammler und Hirten. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 ". "<0.1 " "<1.

Geo. Faktor	Sesshafte Jäger und Sammler			Hirten		
	Koeffizienten	Standardisierte Koeffizienten	Signifikanz	Koeffizienten	Standardisierte Koeffizienten	Signifikanz
(Achsenabschnitt)	1.967	1.337	***	3.306	1.836	***
Temperatur	-0.016	-0.054	**	-0.024	-0.226	***
Niederschlag	-0.008	-0.065	**	-0.028	-0.191	**
Saisonalität Niederschlag	0.008	0.010		-0.007	-0.010	
NPP	0.015	0.010		0.085	0.074	
Küstendistanz	0.029	0.112	***	0.000	-0.001	
Höhenmodell+1	-0.015	-0.009		0.008	0.004	
Höhendifferenz	-0.002	-0.037	.	-0.010	-0.165	***

Anzahl Sprachen		299			76	
Anzahl Gruppen		83			4	

Genealogische Diversität – Mehrebenenanalyse, restliche Subsistenztypen

Tabelle LI: Mehrebenenanalyse der mittleren Verwandtschaftsdistanz aufgeteilt nach Subsistenztyp sesshafte Jäger und Sammler und Hirten. Die Zeichen für die Signifikanz sind: "****" <0.001 "****"<0.01 "****"<0.05 ". "<0.1 " "<1.

Geo. Faktor	Sesshafte Jäger und Sammler			Hirten		
	Koeffizienten	Standardisierte Koeffizienten	Signifikanz	Koeffizienten	Standardisierte Koeffizienten	Signifikanz
(Achsenabschnitt)	3.517	1.517	***	2.974	2.396	***
Temperatur	-0.057	-0.163	***	-0.025	-0.218	**
Niederschlag	0.001	0.011		-0.013	-0.084	
Saisonalität Niederschlag	-0.025	-0.033		0.048	0.067	
NPP	-0.049	-0.035		-0.093	-0.083	
Küstendistanz	-0.003	-0.011		0.011	0.096	
Höhenmodell+1	-0.133	-0.080	**	-0.111	-0.062	
Höhendifferenz	-0.002	-0.024		0.000	0.004	

Anzahl Sprachen	242			67		
Anzahl Gruppen	49			3		

Tabelle LII: Mehrebenenanalyse der PNN-Distanz aufgeteilt nach Subsistenztyp sesshafte Jäger und Sammler und Hirten. Die Zeichen für die Signifikanz sind: "****" <0.001 "***" <0.01 "**" <0.05 "." <0.1 " " <1.

Geo. Faktor	Sesshafte Jäger und Sammler			Hirten		
	Koeffizienten	Standard-isierte Koeffizienten	Signifi-kanz	Koeffizienten	Standard-isierte Koeffizienten	Signifi-kanz
(Achsenabschnitt)	3.293	1.403	***	3.071	2.212	***
Temperatur	-0.068	-0.196	***	-0.021	-0.185	**
Niederschlag	0.008	0.068		-0.023	-0.145	.
Saisonalität Niederschlag	-0.009	-0.012		0.035	0.049	
NPP	-0.049	-0.035		-0.035	-0.031	
Küstendistanz	-0.007	-0.026		0.008	0.071	
Höhenmodell+1	-0.145	-0.087	**	-0.090	-0.050	
Höhendifferenz	-0.001	-0.020		-0.001	-0.011	
Anzahl Sprachen	240			67		
Anzahl Gruppen	49			3		

PERSÖNLICHE ERKLÄRUNG

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und die den verwendeten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Zürich, 30. September 2013

Martina Köhli