



**University of
Zurich**^{UZH}

Einfluss von Flüssen und Gebirgen auf die Interpolation von Spracharealen

GEO 511 Master's Thesis

Author

Aaron Kreienbühl
15-735-129

Supervised by

Dr. Gereon Kaiping
Dr. Peter Ranacher

Faculty representative

Prof. Dr. Robert Weibel

25.02.2023

Department of Geography, University of Zurich

1 ABSTRACT

The expansion and diversification of individual speaker groups throughout history have led to the fact that today we speak different languages in different places. These processes depend, among other things, on landscape features such as natural barriers.

This paper examines the influence of rivers and mountains as natural barriers on the spatial location of language areas. South America serves as the study area. For this purpose, the language areas of 295 languages were interpolated with Voronoi tessellation (VT) and the influence of landscape variables on the weighting of VT was investigated. Three models were tested an unweighted VT, a VT considering rivers as weights and a VT considering mountains as weights. Known point geometries of languages from two different sources were used as seed points. These are, on the one hand Glottolog, which is freely available and on the other hand Ethnologue, which requires a license for access.

The generated surfaces were then compared for their spatial similarity to corresponding polygons of a reference dataset using the Jaccard Index(JI). Subsequently, the JI of the individual language areas were compared with each other to make statements about the influence of the corresponding landscape variable in the weighting of the VT.

The VT weighted with mountains showed only minimal improvements compared to the unweighted VT. The small differences suggest that mountains can only be considered as a driving factor for the separation of two language areas to a limited extent. When weighted with rivers, the results were significantly worse than for unweighted VT, which contradicts the assumption that rivers are a natural barrier with respect to the dispersion of speaker groups. Due to the limitation in the methodology, the choice of seed points did not play a decisive role.

2 ZUSAMMENFASSUNG

Die Expansion und Diversifizierung von einzelnen Sprechergruppen haben im Laufe der Geschichte dazu geführt, dass wir heute an verschiedenen Orten verschiedene Sprachen sprechen. Diese Prozesse hängen unter anderem von Landschaftsmerkmalen, wie natürlichen Barrieren ab.

Diese Arbeit untersucht den Einfluss von Flüssen und Gebirgen als natürliche Barrieren auf die räumliche Lage von Spracharealen. Als Untersuchungsgebiet dient Südamerika. Dazu wurden die Sprachareale von 295 Sprachen mit Voronoi-Tessellation(VT) interpoliert und der Einfluss von Landschaftsvariablen auf die Gewichtung der VT untersucht. Es wurden drei Modelle getestet eine ungewichtete VT, eine VT die Flüsse als Gewichtung berücksichtigt und eine VT der Gebirge als Gewichtung berücksichtigt. Als Seedpoints wurden bekannte Punktgeometrien der Sprachen aus zwei unterschiedlichen Quellen verwendet. Dabei handelt es sich, einerseits um Glottolog, welcher frei verfügbar ist und andererseits Ethnologue, welcher für den Zugang eine Lizenz benötigt.

Die generierten Flächen wurden anschliessend mit Hilfe des Jaccardindex(JI) auf ihre räumliche Ähnlichkeit zu entsprechenden Polygonen eines Referenzdatensatzes verglichen. Im Anschluss wurden die JI der einzelnen Sprachareale miteinander verglichen, um Aussagen über den Einfluss der entsprechenden Landschaftsvariabel bei der Gewichtung der VT zu machen.

Die mit Gebirgen gewichtete VT zeigte nur minime Verbesserungen gegenüber der ungewichteten VT. Die geringen Differenzen lassen darauf schliessen, dass Gebirge nur bedingt als treibender Faktor zur Trennung von zwei Spracharealen angesehen werden können. Bei der Gewichtung mit Flüssen waren die Ergebnisse signifikant schlechter als bei der ungewichteten VT, was der Annahme, Flüsse in Bezug auf die Ausbreitung von Sprechergruppen als natürliche Barriere zu sehen, widerspricht. Auf Grund der Limitation in der Methodik spielte die Wahl der Seedpoints keine entscheidende Rolle.

3 DANKSAGUNGEN

Zunächst möchte ich mich bei Prof. Dr. Robert Weibel und Dr. Gereon Kaiping bedanken, die bei der Themenfindung geholfen haben und mich in der Anfangsphase mit Inputs unterstützten.

Ebenso möchte ich mich bei meinem Betreuer Dr. Peter Ranacher bedanken, der mich immer motivierte und sich die stets die Zeit nahm, mir konstruktive Feedbacks zu geben.

Speziell möchte ich mich bei Dr. Daniel Carlo Pangerl bedanke, der meine Arbeit durchlas und lektorierte.

Schliesslich möchte ich mich bei meinen Freunden, Kollegen und der Familie für ihre ständige Unterstützung.

4 INHALTSVERZEICHNIS

1	Abstract	1
2	Zusammenfassung	2
3	Danksagungen	3
5	Abbildungsverzeichnis	7
6	Tabellenverzeichnis	9
7	Einleitung	10
7.1	Motivation und Hintergrund	10
7.2	Theoretischer Hintergrund	12
7.2.1	Wie ist Sprache definiert	12
7.2.2	Sprachen als dynamischer Prozess	12
7.2.3	Sprachdiversität	14
7.2.4	Räumliche Eigenschaften von Sprachen	15
7.2.5	Daten zur Darstellung von Sprachen	15
7.2.6	Beispiele von Anwendungen	16
7.2.7	Interpolationsmethoden	17
7.3	Related Work Research Gap	19
7.4	Forschungsfragen	20
8	Daten	23
8.1	Flüsse	23
8.2	DHM (Digitales Höhenmodell)	24
8.3	Glottolog	24
8.4	Ethnologue:	25
8.5	Untersuchungsgebiet	25
9	Methoden	26
9.1	Technische Daten	26
9.2	Generieren der Flächen	26
9.2.1	Toblers Hiking	26

9.2.2	Mincostmap:	29
9.2.3	Gewichtungen	30
9.2.4	VT(Gebirge)	32
9.3	Evaluierung	32
9.3.1	Jaccard-Index:	32
9.4	Analyse	34
9.4.1	Differenzen der Verteilungen	34
9.4.2	Darstellung	35
9.4.3	Flussdichte	35
9.4.4	Durchschnittliche Höhe	35
9.4.5	Regressionsanalyse	35
9.5	Workflow	37
10	Resultate	38
10.1	VT Vergleich Ethnologue vs. Glottolog	38
10.2	Welche Auswirkungen hat die Gewichtung der VT	45
10.3	Korrelation mi den Landschaftsvariablen	47
11	Diskussion	50
11.1	Voronoi-Tesselierung	50
11.1.1	Sprache ist nicht kreisförmig	50
11.1.2	Seedpoint	51
11.2	Unterschiede bei der Wahl des SP	51
11.2.1	Keine Abhängigkeit des Abstandes der Seedpoints	51
11.2.2	Wo ist Glottolog besser?	52
11.3	Gebirge	53
11.3.1	Thematisch	53
11.3.2	Methodik	54
11.4	Flüsse	55
11.4.1	Thematik	56
11.4.2	Methodik	59
11.5	Evaluierung	61
12	Fazit	62
12.1	Erreichtes	62

12.2	Grenzen	62
12.3	Zukunftsaussichten	63
13	<i>Bibliografie</i>	64

5 ABBILDUNGSVERZEICHNIS

ABBILDUNG 1: ILLUSTRATION EINER UNGEWICHTETEN VT. _____	17
ABBILDUNG 2: SKIZZE ZEIGT DIE RÄUMLICHE AUSWIRKUNG DER GEWICHTUNG DER VT DURCH EINEN FLUSS (RECHTS) GEGENÜBER DER UNGEWICHTETEN VT (LINKS) AUF DIE TRENNLINIE (GRENZE DER MENGEN). 18	
ABBILDUNG 3: GRAFIK ZEIGT GESCHWINDIGKEIT (WALKING SPEED) IN ABHÄNGIGKEIT ZUR STEIGUNG (SLOPE IN DEGREES) NACH TOBLER (TOBLER, 1993). _____	27
ABBILDUNG 4 GRAFIK ZEIGT DEN OUTPUT DER KOSTENFUNKTION BEI EINER NACH GEWICHTUNG (DHM) VT. SKALIERUNG ZEIGT DIE KOSTEN, _____	28
ABBILDUNG 5: DIE FÜR JEDEN SP ERSTELLTEN OUTPUTRASTER DER KOSTENFUNKTION (SIEHE ABBILDUNG 4 WURDEN MITEINANDER VERGlichen UND DER MINIMALE WERT WURDE ANSCHLIESSEND FÜR JEDE ZELLE ÜBERNOMMEN. ZELLEN DIE KEINE ÜBERSCHNEIDUNGEN MIT DEN INPUTKARTEN HABEN WURDE DER WERT 500 ZUGEWIESEN. _____	30
ABBILDUNG 6: RESULTAT DER MINCOSTMAP. BASIEREND AUF DEN WERTEN VON ABBILDUNG 5 WURDE JEWEILS DIE REFERENZNUMMER DES RASTERS BZW. DER SPRACHE EINER ZELLE ZUGEWIESEN, WELCHE AN DIESER STELLE MINIMALE KOSTEN HAT. _____	30
ABBILDUNG 7: GRAFIK ZEIGT DEN OUTPUT DER KOSTENFUNKTION BEI EINER NACH FLÜSSEN (LINKS FLÜSSE (MITTEL) UND RECHTS FLÜSSE (GROSS)) GEWICHTETEN VT. SKALIERUNG ZEIGT DIE KOSTEN, WELCHE BENÖTIGT WERDEN, UM DEN ENTSPRECHENDEN PUNKT VOM SP AUS ZU ERREICHEN. _____	31
ABBILDUNG 8: GRAFIK ZEIGT DEN OUTPUT DER KOSTENFUNKTION BEI EINER VT(GEBIRGE). SKALIERUNG ZEIGT DIE KOSTEN, WELCHE BENÖTIGT WERDEN, UM DEN ENTSPRECHENDEN PUNKT VOM SP AUS ZU ERREICHEN. _____	32
ABBILDUNG 9: VISUELLE DARSTELLUNG DES JI DER SPRACHE 'CENTRAL AYMARA'. WOBEI ZELLEN DIE SOWOHL DEM MIT VT GENERIERTEN SPRACHAREAL ALS AUCH DEM ENTSPRECHENDEN REFERENZSPRACHAREAL ZUGEWIESEN WURDEN, DEN WERT 3 HABEN. ZELLEN, DIE DEM MIT VT GENERIERTEM SPRACHAREAL ZUGEWIESEN WURDE, ABER NICHT DEM REFERENZSPRACHAREA,L HABEN DEN WERT 2. ZELLEN, DIE DER REFERENZSPRACHAREAL, ABER NICHT DEM MIT VT GENERIERTEM SPRACHAREAL ZUGEWIESEN WURDEN, HABEN DEN WERT 1. ZELLEN, DIE KEINEM DER BEIDEN FLÄCHEN ZUGEWIESEN WURDEN, HABEN DEN WERT 0. _____	34
ABBILDUNG 10: BEISPIELE DER MODELLGÜTE IN ABHÄNGIGKEIT DER KORRELATION DER ZU VERGLEICHENDEN VARIABLEN (SONNECK, 2020) _____	36
ABBILDUNG 11: GRAFISCHER ÜBERBLICK AUF DIE IM KAPITEL 9 BESCHRIEBENEN METHODEN. RECHTECKE BESCHREIBEN PROZESSE. KREISE BESCHREIBEN INPUT- RESPEKTIVE OUTPUTDATEN. _____	37
ABBILDUNG 12: VERTEILUNGEN DES JI(VT(UNGEWICHTET));ETHNOLOGUE) (LINKS) UND DES JI(VT(UNGEWICHTET);GLOTTOLOG) (RECHTS) DER EINZELNEN POLYGONE – WOBEI DIE Y-ACHSE DER ANZAHL POLYGONE ENTSPRICHT. DA ES SICH BEIM JI UM KONTINUIERLICHE WERTE ZWISCHEN 0 UND 1 HANDELT, ENTSPRECHEN DIE BALKEN AGGREGIERTEN WERTEN. _____	39

ABBILDUNG 13: VERTEILUNGEN VON ABBILDUNG 10 ALS BOXPLOT. WEITER WERDEN DIE VERTEILUNGEN BZW. DIE WERTE UNTER BZW. ÜBER DEM MEDIAN MITEINANDER VERGLICHEN. _____	40
ABBILDUNG 14: RÄUMLICHE LAGE DER POLYGONE. SKALIERUNG ENTSpricht DEM JI(VT(UNGEWICHTET);ETHNOLOGUE)) DER EINZELNEN POLYGONE. _____	41
ABBILDUNG 15: RÄUMLICHE LAGE DER POLYGONE. SKALIERUNG ENTSpricht DEM JI(VT(UNGEWICHTET);GLOTTOLOG)) DER EINZELNEN POLYGONE. _____	42
ABBILDUNG 16: RÄUMLICHE LAGE DER POLYGONE. SKALIERUNG ENTSpricht DER DIFF(SEEDPOINTS) DER EINZELNEN POLYGONE. _____	43
ABBILDUNG 17: VERTEILUNG DER ABSTÄNDE DER GLOTTOLOGPUNKTE ZU DEN ETHNOLOGUEZENTROIDEN. EIN ABSTAND BEZIEHT SICH IMMER AUF EINE SPRACHE. _____	44
ABBILDUNG 18: GRAFIK VERGLEICHT DIE DIFF(SEEDPOINTS) (ABBILDUNG 14) IN ABHÄNGIGKEIT ZU DEN ABSTÄNDEN DER SEEDPOINTS (ABBILDUNG 15) DER GLEICHEN SPRACHE. _____	44
ABBILDUNG 19: VERTEILUNGEN DES JI(VT(GEBIRGE);ETHNOLOGUE) DER EINZELNEN POLYGONE – WOBEI DIE Y-ACHSE DER ANZAHL POLYGONE ENTSpricht. DA ES SICH BEIM JI UM KONTINUIERLICHE WERTE ZWISCHEN 0 UND 1 HANDELT, ENTSprechen DIE BALKEN AGGREGIERTEN WERTEN. _____	45
ABBILDUNG 20: VERTEILUNGEN DES JI(FLÜSSE(MITTEL);ETHNOLOGUE) (LINKS) UND DES JI(FLÜSSE(GROSS);GLOTTOLOG) (RECHTS) DER EINZELNEN POLYGONE – WOBEI DIE Y-ACHSE DER ANZAHL POLYGONE ENTSpricht. DA ES SICH BEIM JI UM KONTINUIERLICHE WERTE ZWISCHEN 0 UND 1 HANDELT, ENTSprechen DIE BALKEN AGGREGIERTEN WERTEN. _____	45
ABBILDUNG 21: VERTEILUNGEN VON ABBILDUNG 10 (LINKS), ABBILDUNG 17 UND ABBILDUNG 18 (BEIDE) _____	46
ABBILDUNG 22: RÄUMLICHE LAGE DER POLYGONE. SKALIERUNG ENTSpricht DER DIFF(GEBIRGE) DER EINZELNEN POLYGONE. _____	47
ABBILDUNG 23: GRAFIK VERGLEICHT DIE DIFF(GEBIRGE) IN ABHÄNGIGKEIT ZU DER DURCHSCHNITTLICHEN HÖHE DES GLEICHEN SPRACHAREALS. _____	48
ABBILDUNG 24: GRAFIK VERGLEICHT DIE DIFF(FLÜSSE(GROSS)) (LINKS) BZW. DIE DIFF(FLÜSSE(MITTEL)) (RECHTS) IN ABHÄNGIGKEIT ZU DER JEWEILIGEN FLUSSDICHTe DES GLEICHEN SPRACHAREALS. _____	49
ABBILDUNG 25: ETHNOLOGUEPOLYGON DER SPRACHE QAWASQAR (GELB) MIT ENTSprechENDEN SP VON ETHNOLOGUE (ROT) BZW. GLOTTOLOG (BLAU) _____	52
ABBILDUNG 26: KARTE VERGLEICHT DIE RÄUMLICHE LAGE DER FLÜSSE(MITTEL) MIT DEN POLYGONEN DER AUSREISSER DE VERTEILUNG DER JI(VT(FLÜSSE(MITTEL))). _____	57
ABBILDUNG 27: KARTE VERGLEICHT DIE RÄUMLICHE LAGE DER FLÜSSE(GROSS) MIT DEN POLYGONEN DER AUSREISSER DE VERTEILUNG DER JI(VT(FLÜSSE(GROSS))). _____	58

6 TABELLENVERZEICHNIS

TABELLE 1: KATEGORIEN DER FLÜSSE NACH ABFLUSSMENGE VON GLORIC _____ 23

TABELLE 2: TABELLE ZEIGT DURCHSCHNITTLICHE ABFLUSSMENGEN VOM 08.02.2023 VON DREI SCHWEIZER
FLÜSSEN (BAFU, 2023) _____ 24

7 EINLEITUNG

7.1 MOTIVATION UND HINTERGRUND

Sprache ist ein wesentlicher Bestandteil der Kommunikation. Seit Anbeginn der menschlichen Geschichte erlaubt sie uns, Gedanken und Gefühle auszudrücken, Ideen auszutauschen und Beziehungen aufzubauen. Sie ist die Grundlage für den Austausch von Wissen, die Entwicklung von Traditionen und das Übermitteln von Ideen. Sprache kann dazu beitragen, die Kultur und das Verständnis zwischen Gruppen zu stärken (Theory of knowledge, 2023).

Sprachen sind ein wichtiges kulturelles Merkmal jeder Gesellschaft. Ihre Entwicklung steht in engem Zusammenhang mit historischen und politischen Entwicklungen sowie umweltbedingten Einflüssen und spiegeln die Geschichte der Menschheit wider. So bildet eine gemeinsame Sprache meist den Grundbaustein einer menschlichen Gruppe, die sich der derselben Population oder Kultur angehörig fühlt (Bouckaert *et al.*, 2012; Gavin *et al.*, 2013).

Sprechergruppen/Populationen dürfen dabei nicht als statisch gesehen werden. Eine Population, welche dieselbe Sprache spricht, hat eine gewisse Mobilität, wodurch ein kultureller Austausch mit anderen Populationen über die Zeit entstehen kann. So führen verschiedene Prozesse wie Kriege oder das Entdecken von neuen Handelsrouten dazu, dass Sprechergruppen voneinander getrennt werden oder sich mit anderen vermischen. Dies kann darin münden, dass neue Sprachen entstehen oder alte Sprachen verschwinden (Gavin *et al.*, 2013; Túlio *et al.*, 2019). Diese Prozesse haben dazu geführt, dass heute weltweit etwa 7000 Sprachen gesprochen werden (Eberhard, Simons and Fenning, 2022).

Damit es zu einem solchen kulturellen/sprachlichen Austausch kommen kann, ist ein Kontakt zwischen den Sprechergruppen essenziell. Die Wahrscheinlichkeit, dass ein solcher Kontakt entsteht, ist von verschiedenen Faktoren abhängig. Der intuitivste Faktor ist der räumliche Abstand. Je weiter zwei Sprechergruppen voneinander entfernt leben, desto geringer ist die Wahrscheinlichkeit, dass ein Kontakt entsteht (Nichols, 1997; Greenhill, 2014). Daneben gibt es aber noch weitere räumliche Faktoren, die Einfluss auf den Kontakt zwischen zwei Sprechergruppen haben können. Beispielsweise können natürliche Barrieren wie Gebirge und Flüsse den Austausch zwischen Sprechergruppen erschweren (Greenhill, 2014).

Interessant ist, welchen Einfluss natürliche Barrieren wie Gebirge und Flüsse (Greenhill, 2014) tatsächlich auf die Gebiete, in welchen eine Sprechergruppe lebt, haben. Hätten

natürliche Grenzen einen signifikanten Einfluss auf die Sprachgebiete, würden sich die Grenzen zwischen zwei Sprachgebieten demzufolge in Richtung eines Flusses oder einer Gebirgskette verschieben. Ein ansehnliches Beispiel ist der Rhein. Diese natürliche Grenze bildet heute die Trennlinie zwischen dem deutschsprachigen Gebiet auf der Ostseite und dem französischsprachigen Gebiet auf der Westseite (Suckow, 2012). Auch die Alpen können als natürliche Barriere zwischen der italienischen und der deutschen Sprache gesehen werden. Interessant ist es nun, zu analysieren, ob die genannten natürlichen Grenzen generell als Indikator für Sprachgrenzen gesehen werden können und inwiefern sie gegenüber anderen Faktoren ins Gewicht fallen (Sgobba, 2019).

Insbesondere für Studien, welche sich mit der phonogenetischen Ausbreitung einer Ursprache wie dem Indogermanischen oder der Bantu-Expansion beschäftigten (Neureiter *et al.*, 2021; Koile *et al.*, 2022), können Resultate in diesem Bereich Aufschluss darüber geben, wie man natürliche Barrieren als zusätzliche Wegkosten in ein Modell einbauen lassen. Weiter ist es im Bereich der GIScience (Geographic Information Science) interessant, welche bekannten Methoden geeignet sind, um Sprachareale zu generieren.

Zudem können Forschungen in diese Richtung einerseits für die Linguistik, da Analysen in diesem Bereich helfen, die Entstehung und Verbreitung von einzelnen phonologischen Merkmalen (Wortlauten) oder grammatikalischen Sonderheiten, durch zusätzliche Miteinbeziehung von geographischen Faktoren (in diesem Fall natürliche Grenzen), zu verstehen. Da – wie erwähnt – eine Sprache auch mit einer eigenen Kultur einhergeht, können Modelle in diesem Bereich zudem auf einen kulturhistorischen Ansatz ausgeweitet werden und wären auch in dieser wissenschaftlichen Disziplin von Bedeutung (Mantiri, 2015; Robert Henry Robins, 2022). Man könnte noch einen Schritt weiter gehen und Parallelen von der Entstehung von Sprache zur Entstehung der Arten ziehen. Solche Ansätze wurden schon von mehreren Forschenden untersucht (Stepp, Castaneda and Cervone, 2005; Perrigo, Hoorn and Antonelli, 2020) und sind vor allem im Bereich der Evolutionstheorie von Bedeutung, müssen aber mit Vorsicht genossen werden, da die menschliche Geschichte und damit ebenso die Geschichte der Sprache vergleichsweise jung sind.

Auch im soziologischen und historischen Kontext ist die Evolution von Sprachen von Relevanz. In vielen Fällen der Geschichte bis in die Gegenwart wurden/werden die gemeinsame Sprache und von den herrschenden Schichten missbräuchlich verwendet, um die Einverleibung anderer Staaten oder militärische Aktionen im Namen der gemeinsamen Kultur zu rechtfertigen (Reagan, 2003). Als Beispiele dafür können der 2. Weltkrieg (Anschluss

Österreichs; Lebenserweiterung im Osten im Namen der deutschen Kultur) sowie der russische Überfall auf die Ukraine im Jahr 2022 (Panslawismus) gesehen. Durch das Aufzeigen, welche Faktoren überhaupt zur Entstehung bzw. Ausbreitung einer Sprache führen, können solche negativen, mit Konfliktpotenzial behafteten Diskurse effektiv unterbunden und bekämpft werden.

7.2 THEORETISCHER HINTERGRUND

7.2.1 Wie ist Sprache definiert

Sprachen sind, wie biologische Arten, durch komplexe Abstammungsmuster miteinander verwandt. Eine Sprache unterscheidet sich dabei von anderen Sprachen durch ihre Divergenzen in ihren phonologischen Merkmalen (Aussprache der Laute), der Morphologie (Aufbau der Wörter) sowie der Syntax (grammatikalischer Aufbau der Sätze) (Gavin *et al.*, 2013).

Die Abstammungsmuster sind auch räumlich zu erkennen. So liegen Sprachen, die linguistisch gesehen miteinander verwandt sind, in den meisten Fällen auch näher beieinander (Gavin *et al.*, 2013).

Sprachen werden wiederum in verschiedene Dialekte unterteilt. Es handelt sich dabei um lokale Varietäten einer Sprache, die sich in ihrer Phonologie, Morphologie und Syntax leicht unterscheiden. Wann genau ein Dialekt als eigene Sprache kategorisiert wird, ist nicht einheitlich definiert und variiert von Person zu Person (Picout, 2013). Eine mögliche Definition ist das Vorhandensein einer gemeinsamen Hochsprache (beispielsweise Hochdeutsch). Oft wird auch die gegenseitige Verständlichkeit als Kriterium genannt, dass zwei Dialekte derselben Sprache angehören (Gabor, 2023).

7.2.2 Sprachen als dynamischer Prozess

Um zu verstehen, wo Sprachen genau liegen, muss betrachtet werden, welche Prozesse dazu führen, dass an unterschiedlichen Orten verschiedene Sprachen gesprochen werden.

Die Entstehung und die Verbreitung von Sprachen bzw. Kulturen können mit evolutionsbiologischen Modellen verglichen werden. In beiden Fällen finden sich

phylogenetische Bäume, die auf gemeinsame biologische/linguistische Merkmale verschiedener Arten/Sprachen schliessen lassen (Wichmann, Müller und Velupillai, 2010; Neureiter *et al.*, 2021). Weiter wurden schon viele Studien veröffentlicht, welche einen Zusammenhang zwischen der Biodiversität und der Sprachdiversität untersuchen (Stepp, Castaneda and Cervone, 2005; Gavin *et al.*, 2013; Axelsen and Manrubia, 2014; Hua *et al.*, 2019).

Ein ansehnliches Beispiel bieten die indoeuropäischen Sprachen. Sie alle entstammen einer Ursprache, die vor ca. 10 000 Jahren nördlich des Kaukasus gesprochen wurde. Diese Ursprache breitete sich immer weiter aus. Mit der Zeit wurde die Population in kleinere Populationen/Sprechergruppen unterteilt, die zunehmend isoliert voneinander lebten. Infolgedessen wurde der Austausch zwischen den Sprechergruppen immer geringer. Durch den minimierten Kontakt entwickelten sich die Phonologie und die Grammatik der Sprache einer Population unabhängig voneinander, wodurch erst neue Dialekte und später neue Sprachen entstanden. Dieser Prozess wiederholte sich über die Jahrhunderte stetig und kann als Expansion der Sprachen bezeichnet werden (Bouckaert *et al.*, 2012).

Andererseits gibt es auch Differenzen im Vergleich zur Evolutionsbiologie, wo ein genetischer Austausch zwischen zwei divergierten Arten nicht mehr möglich ist. So kann es zwischen verschiedenen Sprechergruppen, die sich schon seit Jahrhunderten aufgeteilt haben und demzufolge eine andere Sprache sprechen, zu einem erneuten Kontakt kommen. Dies kann dazu führen, dass sich einzelne Sprachen wieder vermischen können, was sich unter anderem durch Lehnwörter erkennbar macht. Dieser Prozess wird auch als «Borrowing» bezeichnet (Bouckaert *et al.*, 2012; Neureiter *et al.*, 2021). Eine typische, stark von Sprachmigration beeinflusste Sprache ist Englisch. Ursprünglich war diese Sprache dem Deutschen sehr nahe, da sie von den Angelsachsen während der Völkerwanderung auf die britische Insel gebracht wurde und sich dort sehr schnell ausbreitete. Der Ärmelkanal bildete eine natürliche Grenze, wodurch sich die Bevölkerung und die Sprache vom ursprünglichen Sächsischen zunehmend isolierten. Später, während der Wikingerzeit, fand eine erste schwache Sprachmigration durch die einfallenden Skandinavier statt. Eine zweite stärkere folgte im 11. Jahrhundert, als die Normannen auf der Insel einfielen und die herrschende Schicht komplett verdrängten. So wurde vom englischen Adel bis in die frühe Neuzeit meist Französisch gesprochen. Dies hatte einen prägenden Einfluss auf die englische Sprache. So finden sich heute zahlreiche französische Lehnwörter (Robert Henry Robins, 2022).

Die beschriebenen Prozesse können dabei unter anderem von spezifischen örtlichen Gegebenheiten wie natürlichen Barrieren beeinflusst werden (Stepp, Castaneda and Cervone, 2005).

7.2.3 Sprachdiversität

Ausgehend von den in Kapitel 7.2.2 beschriebenen Prozessen entsteht in einem Gebiet eine gewisse Vielfalt von Sprachen. Wie häufig sich neue Sprachen herausbilden, ist regional unterschiedlich und hängt von verschiedenen Faktoren ab (Stepp, Castaneda and Cervone, 2005; Hua *et al.*, 2019; Tulio *et al.*, 2021).

Einerseits haben die klimatischen Bedingungen einen grossen Einfluss auf die Sprachvielfalt in einer Region. So führen unter anderem eine höhere Durchschnittstemperatur, eine umfangreichere Niederschlagsmenge, eine höhere primäre Nettoproduktivität sowie geringe saisonale klimatische Schwankungen zu einer erhöhten Sprachdiversität. Dies hat zur Folge, dass sich weltweit gesehen in den Tropen, verglichen mit anderen geographischen Breiten, die höchste Dichte an verschiedenen Sprachen vorfinden lässt (Hua *et al.*, 2019; Túlio *et al.*, 2019; Tulio *et al.*, 2021). Dieses Phänomen wird als latitudinaler Gradient bezeichnet (Collard and Foley, 2002; Currie and Mace, 2009; Gavin and Stepp, 2014).

Weiter hat die Landschaft, wie die natürlichen Barrieren, einen Einfluss auf die Sprachdiversität. Sowohl Berge als auch Flüsse können den sozioökonomischen Austausch innerhalb einer Sprechergruppe erschweren, wodurch die Wahrscheinlichkeit steigt, dass sich zwei isolierte Sprechergruppen entwickeln; gleichzeitig wird der Kontakt zweier verschiedener Sprechergruppen erschwert, wodurch der Effekt des «Borrowing» abgeschwächt wird. Beides führt zu einer erhöhten Sprachdiversität (Stepp, Castaneda and Cervone, 2005; Bouckaert *et al.*, 2012; Hua *et al.*, 2019; Tulio *et al.*, 2021). In Bezug auf Flüsse wird meist die Flussdichte mit der Sprachdiversität verglichen. So korreliert die Anzahl der Flüsse mit der Anzahl an Sprachen in einem Gebiet positiv (Hua *et al.*, 2019; Tulio *et al.*, 2021). Bei Gebirgen wurden die Variablen durchschnittliche Höhe, maximaler Höhenunterschied und Rauigkeit beschrieben. Auch in diesem Fall ist die Korrelation mit der Sprachdiversität positiv (Hua *et al.*, 2019). Als weitere Einflussfaktoren der Landschaft werden die Grösse von Inseln bzw. Kontinente genannt (Tulio *et al.*, 2021).

Zudem können verschiedene soziokulturelle Faktoren einen Einfluss auf die Sprachdiversität haben, wie Handelsrouten oder Kriege. Diese sind nicht so leicht quantifizierbar und können

stark von örtlichen Gegebenheiten abhängen (Tulio *et al.*, 2021). Eine weitere Variable, welche mit der Sprachdiversität zusammenhängt, ist die Bevölkerungsdichte. Auch hier ist eine positive Korrelation festzustellen (Cashdan, 2001; Túlio *et al.*, 2019).

7.2.4 Räumliche Eigenschaften von Sprachen

Um den Einfluss von natürlichen Grenzen auf Sprachareale zu untersuchen und die Resultate anschliessend darzustellen, müssen passende Datensätze, welche Informationen zu einzelnen Sprachen liefern, verwendet werden. Dazu müssen die räumlichen Eigenschaften von Sprachen betrachtet werden. Bei Sprachen handelt es sich um räumliche Objekte, deren Grenzen nicht eindeutig sind. So finden sich häufig zweisprachige Gebiete, beispielsweise im Kanton Graubünden (Deutsch und Rätoromanisch), oder zweisprachige Städte, wie Biel (Deutsch und Französisch). Die Übergänge sind häufig fließend. Sprachgrenzen werden nicht künstlich von Menschenhand definiert, sondern entstehen durch die oben beschriebenen Prozesse. Man spricht in diesem Fall auch von Bona-Fide-Grenzen. Diese stehen im Gegensatz zu Fiat-Grenzen, wobei es sich um von Menschen klar definierte, scharfe Grenzen handelt. Ein Beispiel für eine Fiat-Grenze wäre die Landesgrenze zwischen der Schweiz und Deutschland. Diese entstand zwar durch verschiedene natürliche Prozesse, wurde aber zu einem gewissen Zeitpunkt genau festgelegt und ist eindeutig verortbar (Smith and Varzi, 2000; Vogt *et al.*, 2012). Um die Analyse und die anschliessende Darstellung der Resultate einfach zu halten, sollten Datensätze verwendet werden, welche die Sprache als Fiat-Objekt darstellen.

7.2.5 Daten zur Darstellung von Sprachen

Weiter stellt sich die Frage, welche Arten von Daten für die Kartographierung verwendet werden.

Einerseits können Rasterdaten eingesetzt werden. Diese sind gitterartig angelegt. Jeder Zelle (Pixel) in diesem Raster können bestimmte Informationen zugewiesen werden, wie die an diesem Ort gesprochene Sprache oder die Anzahl an SprecherInnen der Sprache. Die Zelle speichert dabei ihre relative Lage im Raster (zum Beispiel in Form von Koordinaten).

Dadurch können einzelne Zellen mit Zellen aus anderen Rastern, welche die gleiche Position im Raster haben, verrechnet werden, sofern die Raster dieselbe Zellgrösse haben (Stoeckle, 2014; Landesamt für Geoinformation und Landentwicklung Niedersachsen, 2021; Kröner,

2023). In Bezug auf Sprachen liesse sich beispielsweise die kontinuierliche Abnahme von SprecherInnen über ein Gebiet (Eine Zelle entspricht der Anzahl an SprecherInnen) darstellen.

Vektordaten stellen räumliche Objekte als Geometrien dar. Es gibt drei Hauptgeometrien: Punkte, Linien und Flächen. Ein Punktobjekt wird dabei durch ihre Koordinaten erfasst. Eine Linie verbindet zwei Punkte an unterschiedlichem Ort. Bei einem Polygon handelt es sich um eine Fläche, welche aus mehreren Punkten und Linien besteht. Die Flächen besitzen somit eine scharfe Grenze sowie einen Schwerpunkt, der als Zentroid definiert ist. Bei einem Vektorobjekt können zusätzlich zur Geometrie (ähnlich wie bei einer Rasterzelle) ein Attribut oder mehrere Attribute hinzugefügt werden (Landesamt für Geoinformation und Landentwicklung Niedersachsen, 2021; Kröner, 2023) .

Sprachen können durch Punkte dargestellt werden. Die Vorteile von Punkten sind die Einfachheit und die geringe Speichergrösse. Nachteilig ist, dass man keine Informationen über die Ausbreitung der Sprache haben, was wichtig ist, wenn man Sprachareale bzw. Sprachgrenzen analysieren möchte. Um eine Fläche mit scharfen Grenzen zu erhalten, muss zuerst eine Interpolationsmethode angewendet werden.

Mit Polygonen lässt sich ein Gebiet aufzeigen, in welchem eine bestimmte Sprache gesprochen wird, wodurch Informationen über deren Verbreitung vorhanden sind und analysiert werden können. Ein Beispiel bildet der Datensatz von Ethnologue (Eberhard, Simons and Fenning, 2022) (siehe Kapitel 8.4). Hier werden 7151 verschiedene Sprachen – die gesamte Welt betreffend – als Polygon bzw. Multipolygon dargestellt. Dabei wird jedes Polygon nur einer Sprache zugeordnet, während einer Sprache mehrere Polygone zugewiesen werden können (Multipolygone).

7.2.6 Beispiele von Anwendungen

Je nach Verwendungszweck können den Karten noch weitere Attribute hinzugefügt werden, wie die Sprecherzahl, oder Sprachen können anhand ihrer phylogenetischen Abstammung klassifiziert und anschliessend entsprechend visualisiert werden (Bouckaert *et al.*, 2012; Bouckaert, Bowerman and Atkinson, 2018), um nur einige Anwendungsbeispiele zu nennen. Weiter können auch Variablen im Zusammenhang mit Sprache (beispielsweise die Sprachdiversität (Stepp, Castaneda and Cervone, 2005; Hua *et al.*, 2019; Tulio *et al.*, 2021) modelliert werden. Bei den genannten Arbeiten wurde jeweils ein Modell kreiert, welches

aufzeigen soll, wie viele verschiedene Sprachen in einem Gebiet gesprochen werden. Hier spielen die Namen der Sprachen keine Rolle, da nur deren Häufigkeitsverteilung untersucht wurde.

Als Zusatz können weitere Elemente einer Karte hinzugefügt werden, die je nach Thematik variieren. Als Basis gilt meist eine Hintergrundkarte des zu betrachtenden Gebiets. Des Weiteren können zusätzliche Landschaftsmerkmale aufgeführt werden. In Bezug auf die Analyse von natürlichen Grenzen können beispielsweise für Flüsse Linien und für Gebirge ein DHM (siehe Methodik) verwendet werden.

7.2.7 Interpolationsmethoden

Wie in Kapitel 7.2.5 erwähnt, bieten Punktdaten keine Informationen über die räumliche Verteilung einer Sprache. Für Analysen über die räumliche Ausbreitung von Sprachen ist dies essenziell. Daher müssen die Sprachareale ausgehend von den Punktdaten geschätzt werden. Die GIScience verwendet zu diesem Zweck verschiedene Interpolationsmethoden. Diese variieren stark in ihrer Komplexität und eignen sich unterschiedlich gut, je nach Anwendungsgebiet. Eine im Vergleich simple Methode stellt die Voronoi-Tesselierung (VT) dar (Meng, Liu and Borders, 2013; Geographic Information Technology Training Alliance, 2016; Bhunia, Shit and Maiti, 2018; GISGeography, 2022).

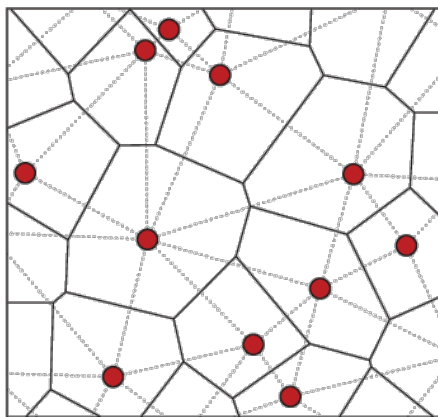


Abbildung 1: Illustration einer ungewichteten VT. Rote Punkte dienen als SP (Leipold et al., 2017).

Die normale VT ist eine räumliche Interpolation, die einem Punkt(A) aus dem zu interpolierendem Datensatz, auch Seedpoint (SP) genannt, alle Punkte(B) zuweist, welche sich näher an Punkt(A) befinden als an dem am nächsten benachbarten Punkt (C) (welcher

ebenfalls als Seedpoint (SP) fungiert) des zu betrachteten Punktes(A). Dem Punkt(c) wiederum werden alle Punkte (D) zugewiesen, welche sich näher an Punkt(C) befinden als an Punkt(A) (siehe Abbildung 2 (links)).

Ein klassisches Voronoi-Polygon entsteht, indem man die Mittelsenkrechten zwischen allen benachbarten Punkten miteinander verbindet. Des Weiteren ist es möglich, Voronoi-Polygone mit Rasterdaten zu erstellen. Hier wird meist ein NN(Nearest Neighbour)-Algorithmus verwendet, wobei die nächsten Zellen so lange der Ursprungszelle zugeordnet werden, bis sie auf eine Zelle treffen, welche bereits einem anderen Ursprungsort zugewiesen werden kann (Geographic Information Technology Training Alliance, 2016; GISGeography, 2022).

Der Vorteil der Rasterberechnung ist, dass nichteuklidische Metriken, Gewichtungsfaktoren etc. berücksichtigt werden können. So kann beispielsweise die Überquerung eines Flusses oder eines Gebirges mit erhöhten Kosten verbunden sein, wodurch die einzelnen Voronoi-Polygone entsprechend angepasst werden. Dieser Ansatz wird auch gewichtete Voronoi-Polygone genannt. Die Methode der Gewichtung kann sich nach Anwendungsgebiet stark unterscheiden (Mu, 2004; Geographic Information Technology Training Alliance, 2016; Liu *et al.*, 2019; Kaiping, 2021).

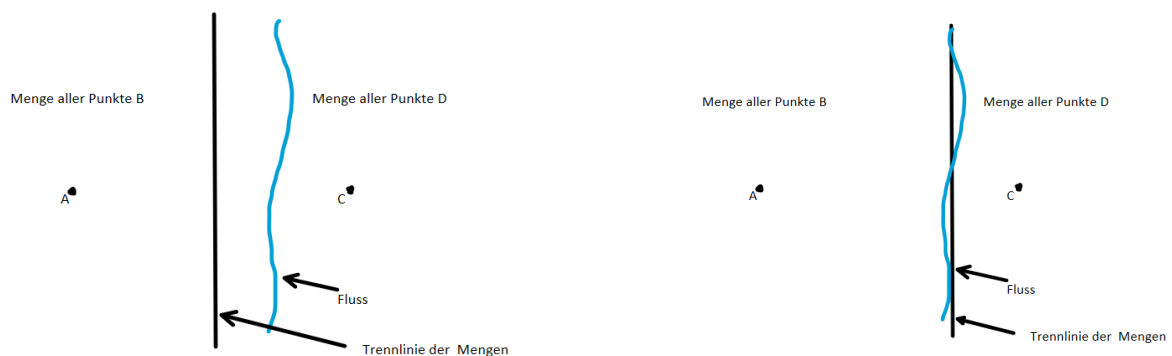


Abbildung 2: Skizze zeigt die räumliche Auswirkung der Gewichtung der VT durch einen Fluss (rechts) gegenüber der ungewichteten VT (links) auf die Trennlinie (Grenze der Mengen).

7.3 RELATED WORK RESEARCH GAP

Isolationstheorie vs. ökologische Risikohypothese

Unter der Berücksichtigung all der im letzten Abschnitt erwähnten Einflüssen untersuchten (Hua *et al.*, 2019) dabei zwei Hypothesen, welche zu einer erhöhten Sprachdiversität führen können. Die Untersuchungen basieren auf den Resultaten von (Axelsen and Manrubia, 2014), die eine Korrelation sowohl der Flussdichte als auch der Rauigkeit mit der Sprachdiversität auf allen Kontinenten feststellen konnten.

Einerseits geht die ökologische Risikohypothese davon aus, dass gleichmässige über das Jahr verteilte klimatische Verhältnisse, wie sie in den Tropen herrschen, zu einer über das Jahr gesehen stabilen Nahrungsmittelproduktion führen. Dies wiederum resultiert darin, dass sich Menschen, welche dort leben, ganzjährig vor Ort ernähren können. Im Gegensatz dazu stehen Kulturen, welche in den gemässigten Klimazonen leben. Hier ist die Nahrungsmittelproduktion stark saisonal abhängig. Dies führt dazu, dass Sprechergruppen, um ihr Überleben zu sichern, eher auf den Austausch mit anderen Kulturen angewiesen sind, da sie nur saisonal von ihren eigenen Erträgen leben können, was wiederum bedingt durch den Austausch mit Handelspartnern einen dämpfenden Effekt auf das «Borrowing» und dadurch auf die Sprachvielfalt hat (Hua *et al.*, 2019).

Dagegen steht die Hypothese der geographischen Isolation. Hier wird untersucht, ob ein kausaler Zusammenhang zwischen Sprachdiversität und Landschaft, welche durch vier Variablen definiert wird (durchschnittliche Höhe, maximaler Höhenunterschied, Landschaftsrauigkeit und Flussdichte), besteht. Diese Faktoren sollen laut der Hypothese dazu führen, dass die sozioökonomische Mobilität von verschiedenen Sprach/Bevölkerungsgruppen durch natürliche Barrieren erschwert wird, was den kulturellen sowie den sprachlichen Austausch zwischen diesen Gruppen erschwert (Hua *et al.*, 2019).

Die empirischen Untersuchungen von (Hua *et al.*, 2019) unterstützen die ökologische Risikohypothese klar. Dies liegt vor allem daran, dass die minimale Anzahl an Individuen der kleinsten Population in einem bestimmten Raum (auf die Grösse bezogen) kleiner wird, bei höherer Sprachdiversität, was darauf schliessen lässt, dass gewisse Faktoren gegeben sind, dass auch kleinste Populationen ohne Kontakt zu anderen Populationen überleben können (Hua *et al.*, 2019).

Die Korrelation von Flussdichte und Sprachdiversität wird in der Analyse von (Hua *et al.*, 2019) auf die ökologische Risikohypothese zurückgeführt. Dies kommt daher, dass Flüsse neben einer natürlichen Barriere auch eine natürliche Ressource sind, welche das ganze Jahr über verwendet werden kann, zum Beispiel Fischfang.

Bisher wurde untersucht, ob zwischen den genannten Faktoren (Flussdichte, maximale unterschiedliche Höhe, durchschnittliche Höhe und Landschaftsrauigkeit) eine Korrelation zur Häufigkeit der Sprachen, auf eine Region betrachtet, besteht. So bestätigen die Resultate die ökologische Risikohypothese im Grossen und Ganzen (Hua *et al.*, 2019). Allerdings lassen sich Isolationseffekte nicht ausschliessen. Um dies zu untersuchen, muss nicht mehr die Sprachdiversität betrachtet werden, sondern man kann sich auf einzelne Sprachgebiete fokussieren und deren Flächen mit den entsprechenden Grenzen auf den räumlichen Einfluss von Flüssen und Gebirge genauer analysieren, um herauszufinden, inwiefern die Isolationshypothese tatsächlich Einfluss auf die Grenzen zwischen den Sprachen nimmt.

Um den Einfluss der Isolationshypothese zu untersuchen, sollten Polygone aus einem Datensatz mit Spracharealen, welche mit einer ungewichtete simplen VT (Voronoi-Tessellierung) generiert wurden und den Zentroid des entsprechenden Polygons als SP (Seedpoints) verwenden, verglichen werden. Derselbe Vergleich kann nun mit denselben Polygonen und Spracharealen, die mit einer gewichteten VT generiert wurden durchgeführt werden. Die Resultate der jeweiligen Vergleiche können im Anschluss auf Unterschiede analysiert werden.

In einem weiteren Schritt kann geprüft werden, inwiefern sich Punkte aus einem frei verfügbaren Datensatz als SP für dieselbe Methode eignen.

7.4 FORSCHUNGSFRAGEN

Die vorhin erwähnte Untersuchung bezüglich der Isolationshypothese kann nun in Bezug auf die zu verwendenden Daten ausformuliert werden. In einem ersten Schritt sollte gezeigt werden, inwiefern die VT geeignet ist, um Sprachareale zu interpolieren. Als Vergleichsfläche wird dabei immer das Polygon/Multipolygon derselben Sprache aus dem Ethnologuedatensatz (siehe Kapitel 8.4) verwendet. Die räumliche Ähnlichkeit zwischen dem

interpolierten Sprachareal und dem damit referenziertem Sprachpolygon wird mit dem Jaccardindex (genaue Beschreibung in Kapitel 9.3.1) ermittelt.

Für die VT werden einmal die Punkte von Glottolog (siehe Kapitel 8.3) und einmal die Zentroidpunkte von Ethnologue (Ethnologuezentroids) als Seedpoints (SP) verwendet, was Aufschluss darüber geben sollte, ob die VT mit Daten, welche frei zur Verfügung stehen, ähnlich gute Resultate erzielt, wie mit Daten, die nur mit Lizenz erhältlich sind, wie dies bei Ethnologue der Fall ist. Um die VT mit unterschiedlichen SP zu vergleichen, werden bei der Evaluierung nur Sprachen verwendet, welche in beiden Datensätzen vorhanden sind.

(Hua *et al.*, 2019) erwähnt in Bezug auf die Isolationshypothese, dass sowohl Flüsse als auch Gebirge als mögliche Einflussfaktoren in Betracht gezogen werden können. In einem weiteren Schritt wäre es interessant, zu sehen, welchen Einfluss die natürlichen Barrieren auf die mit VT generierten Sprachflächen haben. Dazu kann geprüft werden, wie sich eine Gewichtung der VT mit einem Flussbarrierenraster -> **VT(Fluss)** bzw. mit einem digitalen Höhenmodell (DHM) -> **VT(Gebirge)** auf die resultierenden Sprachflächen auswirkt. Diese können im Anschluss mit den aus der ungewichteten VT -> **VT(ungewichtet)** resultierenden Sprachflächen verglichen werden, um so Aufschluss darüber zu erhalten, welche Polygone wie stark durch die jeweiligen Gewichtungen beeinflusst worden sind.

Weiter können Sprachflächen, die durch gewichtete VT stark von der VT(ungewichtet) abweichen, mit Parametern verglichen werden, welche als Indikator für eine flussreiche bzw. gebirgige Region gelten. (Hua *et al.*, 2019) nennen dabei unter anderem die Flussdichte und die durchschnittliche Höhe eines Sprachareals als Indikatoren.

Daraus entstehen folgende Forschungsfragen:

Forschungsfrage 1: Wie gut stimmen Sprachareale, die mit VT ermittelt wurden, mit den Referenzspracharealen aus Ethnologue überein?

Hypothesen:

1 a)

Mit VT generierte Sprachareale, die Glottolog Punktgeometrien als SP verwenden, stimmen über die ganze Verteilung gesehen mit den Referenzspracharealen (Wert des JI) schlechter überein als solche, die Ethnologue Zentroide als SP verwenden, da diese neben der semantischen Abhängigkeit (Sprache) auch eine geometrische Abhängigkeit zu den Referenzspracharealen haben. Die Übereinstimmung ist umso tiefer, je weiter die beiden zu vergleichenden SP auseinander liegen.

1b)

VT Sprachpolygone, die den Einfluss von Gebirgen und Flüssen berücksichtigen, stimmen mit den Referenzspracharealen besser überein als solche, die nur die räumliche Entfernung berücksichtigen.

Forschungsfrage 2: Hängt der Unterschied zwischen VT (ungewichtet) und VT (Gebirge) bzw. von VT (ungewichtet) und VT (Fluss) von entsprechenden Landschaftselementen ab?

Hypothesen:

2 a)

Der Unterschied zwischen VT(ungewichtet) und VT(Fluss) hängt von der Flussdichte ab. VT(Fluss) generiert signifikant bessere Polygone in Gebieten, wo die Flussdichte hoch ist.

2 b)

Der Unterschied zwischen VT(ungewichtet) und VT(Fluss) hängt von der durchschnittlichen Höhe eines Sprachareals ab. VT(Gebirge) generiert signifikant bessere Sprachareale in Gebieten, wo die durchschnittliche Höhe hoch ist.

8 DATEN

8.1 FLÜSSE

Für die Flussdaten wurde das Global River Classification (GloRiC) verwendet. Hierbei handelt es sich um eine Datenbank, die einzelne Flussabschnitte als Linien speichert und ihnen verschiedene Attribute zuweist. Als Basis für die geographische Unterteilung der Flussabschnitte dient HydroRivers (Lehner and Grill, 2013). Die hydro-ökologischen Merkmale wurden mit Hilfe der Hydro-Atlas-Datenbank erstellt (Linke *et al.*, 2019).

Version 1.0 von GloRiC bietet eine hydrologische, klimatische und geomorphologische Unterklassifizierung sowie einen aus verschiedenen Merkmalen kombinierten Flusstyp für jeden Flussabschnitt, was zu insgesamt 127 Flussabschnittstypen führt. Insgesamt beinhaltet der Datensatz 8.5 Millionen Flussabschnitte mit einer Gesamtlänge von 35.9 Millionen km über die ganze Welt verteilt (Ouellet Dallaire *et al.*, 2019).

Eine genaue Dokumentation über die korrekte Verwendung des Datensatzes und die genaue Bedeutung der einzelnen Attribute finden man bei (*GloRiC Global River Classification Technical Documentation Version 1.0*, 2018). Für diese Arbeit war die durchschnittliche Abflussmenge relevant. Diese wurde in fünf Kategorien unterteilt:

Tabelle 1: Kategorien der Flüsse nach Abflussmenge von GloRiC

Kategorie	Abflussmenge [$\text{m}^3 \text{s}^{-1}$]
Very low	0.1–10
Low	10–100
Medium	100–1000
High	1000–10 000
Very high	> 10 000

Tabelle 2 zeigt die Abflussmengen von Schweizer Flüssen, um einen Grössenvergleich darzustellen.

Tabelle 2: Tabelle zeigt durchschnittliche Abflussmengen vom 08.02.2023 von drei Schweizer Flüssen (BAFU, 2023)

Name (Ort)	Abflussmenge [$\text{m}^3 \text{s}^{-1}$]
Thur (Andelfingen)	19
Aare (Brugg)	149
Rhein (Basel)	541

8.2 DHM (DIGITALES HÖHENMODELL)

Beim DHM handelt es sich um eine Darstellung topografischer Oberflächen der Erde ohne Bäume, Gebäude und andere Oberflächenobjekte (USGS, 2023).

Als DHM wurde Global Land One-kilometer Base Elevation (GLOBE) verwendet. Hierbei handelt es sich um einen Rasterdatensatz, bei welchem jeder Zelle ein Wert zugeschrieben wird, welcher die Höhe über Meer repräsentiert. Dieses deckt die gesamte Welt ab. Die genaue Grösse einer Rasterzelle wird mit 30 Bogensekunden angegeben, was nominal 1 km entspricht (Hastings and Dunbar, 1999).

8.3 GLOTTOLOG

Bei Glottolog handelt es sich um eine Datenbank, die verschiedenen Informationen zu über 7000 Sprachen beinhaltet. Diese ist für jede Person zugänglich. Dabei kann über einen Katalog, welcher in die einzelnen Sprachfamilien unterteilt ist, oder über direkte Eingabe nach einer bestimmten Sprache gesucht werden. Jeder Sprache ist eine Punktkoordinate zugewiesen, welche im Gebiet liegt, in dem diese Sprache gesprochen wird. Genaueres zur Datenerhebung wird nicht erwähnt. Die jeweiligen Punkte lassen sich auf einer Weltkarte visuell darstellen. Punktdaten können von der Website inklusive der relevanten Metainformation in Form eines Geojson-files heruntergeladen werden (Hammarström *et al.*, 2021). Die Glottologpunkte wurden als SP für die VT verwendet.

8.4 ETHNOLOGUE:

Ethnologue ist eine weitere Datenbank, die verschiedene Information über die weltweit verbreiteten Sprachen beinhaltet. Es handelt sich, im Gegensatz zu Glottolog, um einen Datensatz, der nur mit entsprechender Lizenz genutzt werden kann. Jeder Sprache im Datensatz werden ein Polygon/Multipolygon und weitere Attribute, zum Beispiel der Name der Sprache, zugewiesen. (Eberhard et al., 2021). Für diese Arbeit war nur der Name der Sprache von Relevanz, um einen semantischen Link zu den Glottologdaten zu erhalten.

Die einer Sprache zugewiesenen Polygone/Multipolygone basieren dabei auf aggregierte Daten, welche von verschiedenen Sprachexperten in den entsprechenden Regionen gesammelt wurden (Eberhard et al., 2021).

Die Ethnologue-Polygone wurden als Referenzsprachareale verwendet, wobei die generierten Sprachareale mit den Sprachpolygonen des Ethnologuedatensatzes verglichen wurden. Des Weiteren wurde die Zentroidpunkte als SP verwendet.

8.5 UNTERSUCHUNGSGBIET

Als Untersuchungsgebiet wurde Südamerika gewählt. Der Kontinent ist für die Analyse gut geeignet, da man hier ein komplexes Flusssystem (Amazonasgebiet) und eine Gebirgsregion (Anden) findet. Weiter liegt die nördliche Hälfte in den Tropen, wo man eine hohe Sprachdiversität vorfindet wie in Kapitel 7.2.3, was zu mehr Sprachgebieten führt, welche untersucht werden können. Überdies sind in den Tropen laut (Hua *et al.*, 2019) die Effekte der ökologischen Risikohypothese besonders ausgeprägt, wodurch sich Vergleiche zur Isolationshypothese vornehmen lassen. Zudem sind viele indigene Kulturen beheimatet, welche unterschiedliche Sprachen sprechen, was weiter für die Sprachvielfalt spricht (Germani and Knapp, 2023). Der Einfluss der Kolonialsprachen (Spanisch und Portugiesisch) wurde nicht berücksichtigt.

9 METHODEN

9.1 TECHNISCHE DATEN

Als CRS (coordinate reference system) wurde für alle beschriebenen Methoden EPSG:4326 (WGS 84) verwendet, da dies eine weltweite Gültigkeit hat und alle verwendeten Daten bereits in diesem Referenz-System waren. Abgesehen vom DHM waren die anderen Datensätze jeweils in Vektoren gegeben und mussten zuerst rasterisiert werden. Als Zellgrösse wurde die Grösse einer Zelle (nominal 1 km) des DHM-Datensatzes gewählt. Geringe räumliche Verzerrungen, die auf Grund der zweidimensionalen Projektion entstehen, wurden dabei bewusst in Kauf genommen.

Der grösste Teil des verwendeten Codes wurde in der Programmiersprache Python 3 geschrieben. Als Entwicklungsumgebung diente Jupyter Notebook (Version 2022.10). Sämtliche Raster wurden im Xarray-Format (Hoyer *et al.*, 2016) bearbeitet. Weitere spezielle verwendete Libraries sind Geopandas (Jordahl *et al.*, 2020) zum Vorprozessieren und Plotten sowie Geocube zur Rasterisierung von Vektordaten.

Die Kostenfunktion basierte auf der movecost-function von (Alberti, 2019). Der Code wurde in R-Studio (Version 2022.07 (R4.2)) ausgeführt.

Sämtliche Prozesse wurden mit dem ScienceApps-Server der Universität Zürich durchgeführt.

Für den schriftlichen Teil wurde das Programm Word verwendet.

9.2 GENERIEREN DER FLÄCHEN

In einem ersten Schritt wurden mit VT die Sprachareale generiert, welche später mit den Polygonen/Multipolygonen von Ethnologue Referenzspracharealen verglichen wurden.

9.2.1 Toblers Hiking

Die Hiking-Funktion wurde von Waldo Tobler entwickelt (Tobler, 1993). Es handelt sich dabei um eine Funktion, welche die Geschwindigkeit eines Wanderers in Abhängigkeit von der Steigung der zurückzulegenden Strecke berechnen soll. Tobler geht davon aus, dass die

Höchstgeschwindigkeit eines Wanderers 6 km/h beträgt und bei einer Steigung von -0.05 (dh/dx) erreicht wird. Des Weiteren nimmt die Geschwindigkeit bei erhöhter Steigung (sowohl auf- als auch abwärts) exponentiell ab Abbildung 3.

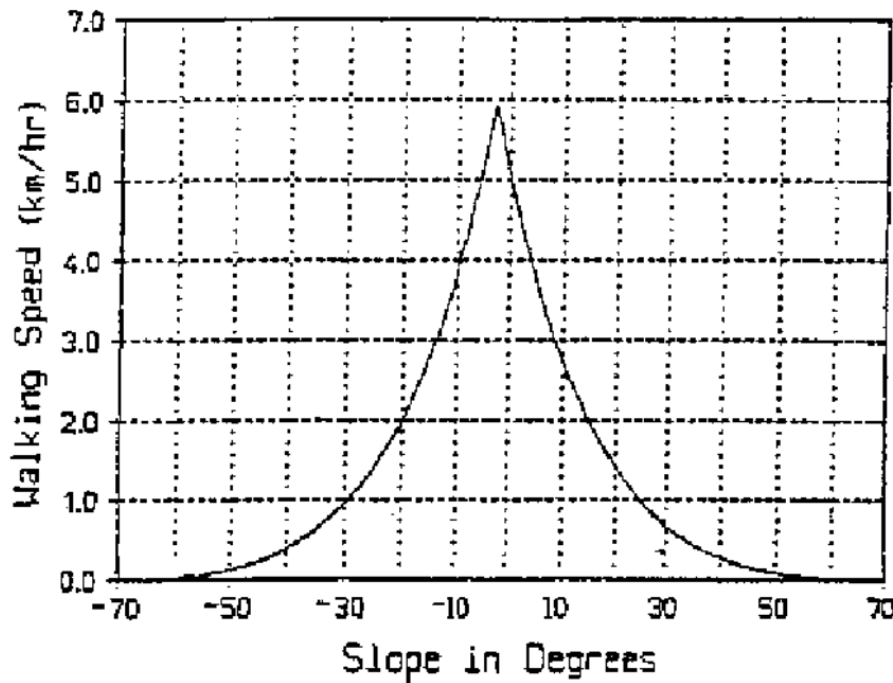


Abbildung 3: Grafik zeigt Geschwindigkeit (Walking Speed) in Abhängigkeit zur Steigung (Slope in Degrees) nach Tobler (Tobler, 1993).

Dieses Modell kann dafür verwendet werden, die Dauer einer Reise zwischen zwei Punkten in Abhängigkeit zur Topologie zu berechnen (Tobler, 1993; Goodchild, 2020).

$$W = 6 \exp(-3.5 * \text{abs}(S + 0.05))$$

W = Geschwindigkeit [km/h]

S = Steigung = dh/dx (müssen in der gleichen Einheit sein) = tan(Θ)

Θ = Winkel der Steigung (Inklination)

(Tobler, 1993)

(Magyari-Saska, Magyari-Sáska and Dombay, 2012) beschrieben eine Methode, die einen NN(Nearest-Neighbour)-Algorithmus basierend auf der Hiking-Funktion verwendet, welcher eine Karte ausgibt, die alle Punkte(B) berücksichtigt, welche in einer gewissen Zeit von einem Punkt(A) zu Fuss erreicht werden können. Als Input dient dabei ein Rasterdatensatz, welcher

die Kosten darstellt (meist ein DHM), welches den Wert einer Zelle definiert. Für jede Zelle wird dabei mit Hilfe eines 8er-NN-Algorithmus die Steigung gegenüber der benachbarten Zelle berechnet. Dadurch lassen sich mittels der oben beschriebenen Formel und der Angabe der Strecke [km] die Kosten in Form von Zeit berechnen, welche benötigt werden, um einen gewissen Punkt zu erreichen.

Dieser Ansatz kann beispielsweise genutzt werden, um die Ausbreitung einer Sprache in Abhängigkeit der Steigung zu simulieren. Für jede Sprache wurde dabei ein fixer Seedpoint (SP) als Input genommen, von welchem aus der Algorithmus startet. Bei der anschliessend bearbeiteten Karte entspricht der Wert einer Zelle der Höhe der Kosten (Zeit), welche aufgewendet werden müssen, um vom SP aus der Zelle zu erreichen.

Die movecost-function (Alberti, 2019) welche auf der Hiking-Funktion basiert, kann auch zur Kostenberechnung der Flüsse verwendet werden (siehe Kapitel 9.2.3). Daher wird im weiteren Verlauf der Arbeit der Begriff Kostenfunktion verwendet.

Ein ähnlicher Ansatz, um die Kosten für die Ausbreitung von Sprache zu berechnen, wurde bereits von (Kaiping, 2021) verwendet.

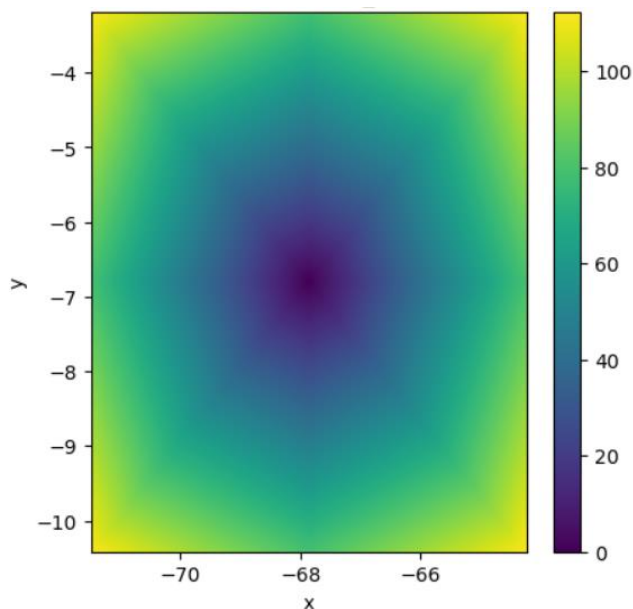


Abbildung 4 Grafik zeigt den Output der Kostenfunktion bei einer nach Gewichtung (DHM) VT. Skalierung zeigt die Kosten, welche benötigt werden, um den entsprechenden Punkt vom SP aus zu erreichen.

9.2.2 Mincostmap:

Weiter wurde eine Methode benötigt, welche mehrere Karten (mit der derselben Geometrie) aggregieren kann, wobei jeder Sprache, welche sowohl im Glottologdatensatz als im Ethnologuedatensatz vorkam, eine Referenznummer (2 bis 297) zugewiesen wurde, um sie später miteinander vergleichen zu können. Ziel war es, aus mehreren Input-Karten des Untersuchungsgebiets eine einzelne Karte zu erstellen. Dabei stellen die Zellen jeder Inputkarten die Kosten dar, welche benötigt wurden, um eine bestimmte Zelle vom SP aus zu erreichen. Die Zelle, auf welcher sich der entsprechende SP befindet, hat somit den Wert 0 (Abbildung 4). Die Inputkarten wurden für alle Gewichtungen erstellt, wobei einmal der Glottologdatensatz und einmal der Ethnologuedatensatz als Grundlage für die SP dienten. Handelte es sich um eine Sprache, die nicht in beiden Datensätzen vorhanden waren, wurden ihr die Referenznummer 1 zugewiesen und die entsprechende Sprache wurde für die Mincostmap, aber nicht für die weitere Analyse verwendet. Zellen im Hintergrund, welche keiner Sprache zugewiesen wurden erhielten den Wert -100.

Ziel war es, eine Karte zu erstellen, bei der für jede einzelne räumlich referenzierte Zelle der Minimumwert über aller Input-Karten an genau dieser Zelle ausfindig gemacht werden konnte (siehe Abbildung 5). Weiter registriert der Algorithmus die Referenznummer der Karte, an der die Zelle minimal ist. In einer neuen Karte wird nun dieser Zelle die Referenznummer der entsprechenden Karte zugewiesen, bei welcher die Kosten minimal sind (siehe Abbildung 6) Als Resultat erhält man ein Raster, bei dem erkennbar ist, bei welcher Inputkarte (gekennzeichnet mit Referenznummer der Sprache) die Kosten einer bestimmten Karte am kleinsten sind.

Wie in Abbildung 5 und Abbildung 6 zu erkennen ist, wurden nicht alle Punkte erreicht, die Flächen auf Grund der hohen Datenmengen begrenzt waren, was aber nur in sehr wenigen Fällen zu Unsicherheiten führte.

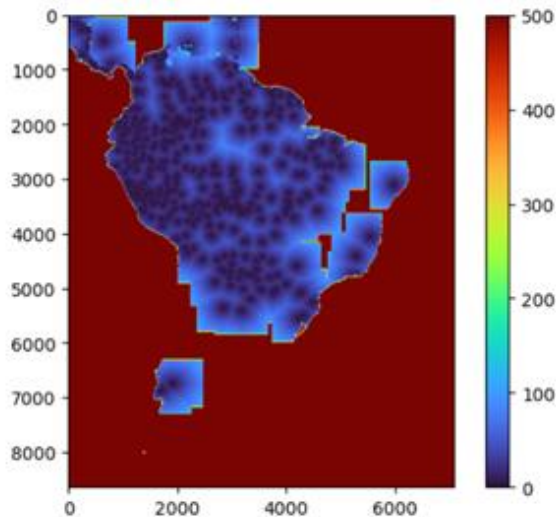


Abbildung 5: Die für jeden SP erstellten Outputraster der Kostenfunktion (siehe Abbildung 4) wurden miteinander verglichen und der minimale Wert wurde anschliessend für jede Zelle übernommen. Zellen die keine Überschneidungen mit den Inputkarten haben wurde der Wert 500 zugewiesen.

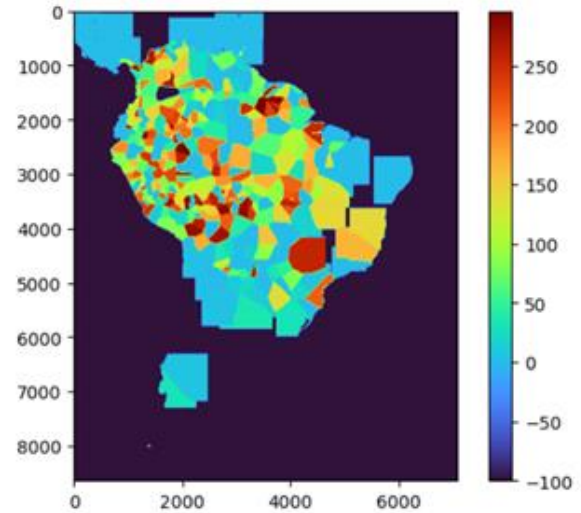


Abbildung 6: Resultat der Mincostmap. Basierend auf den Werten von Abbildung 5 wurde jeweils die Referenznummer des Rasters bzw. der Sprache einer Zelle zugewiesen, welche an dieser Stelle minimale Kosten hat.

9.2.3 Gewichtungen

Je nach Gewichtung wurde für die Kostenfunktion ein anderes Kostenraster als Input genommen. Als SP wurden für jede Analyse nach Gewichtung einmal die Glottologpunkte und einmal die Ethnologuezentroids verwendet.

VT(ungewichtet)

In diesem Schritt wurden keine natürlichen Grenzen berücksichtigt. Die jeweilige Grösse der Spracharealen hängt nur vom Abstand zu den nächstliegenden Punkten ab. Die Methode entspricht der ungewichteten VT (GISGeography, 2022). Auf Grund des 8er-NN-Algorithmus ist die Ausbreitung der VT nicht perfekt kreisförmig, wie in Abbildung 4 gut erkennbar ist.

Als Basis für die Kostenfunktion diente hier ein Raster mit konstanten Werten (in diesem Fall der Wert 1).

VT(Flüsse)

Hier wurden die Flüsse als Kosten berücksichtigt. Zuerst wurden die Linien mit einem Buffer zu Polygonen erweitert, welcher beidseitig der Zellgröße vom DHM entsprach. Damit konnten die Vektordaten rasterisiert werden und es entstanden keine Lücken, welche der NN-Algorithmus hätte passieren können.

Anschließend wurden die entsprechenden Daten rasterisiert. Ergebnis war ein Raster mit Zellen der Kategorie «Fluss» und Zellen der Kategorie «nicht Fluss». Zellen der erstgenannten Kategorie wurde der Wert 100 zugewiesen. Alle anderen Zellen entsprechen dem Wert 1. Dadurch ergibt sich an den Grenzen zu den Flüssen eine Steigung von 100, was mit Betrachtung auf die Kostenfunktion (basierend auf der Hiking-Funktion) ($S = 100$) als unüberwindbares Hindernis gesehen werden kann.

Da in der Literatur keine genaueren Hinweise gegeben wurden, welche Eigenschaften ein Fluss (Breite, Tiefe etc.) haben muss, um als natürliche Barriere im Zusammenhang mit dieser Arbeit zu gelten, wurden zwei verschiedene Annahmen überprüft. Einerseits wurde ein Output generiert, bei welchem nur Flüsse aus den Kategorien «High» und «Very high» (Tabelle 1) verwendet wurden. Der damit generierte Output wird folglich als **VT(Flüsse(gross))** bezeichnet. Zudem wurde ein zweiter Output generiert, bei welchem die Flüsse aus den Kategorien «High» und «Very high» mit den Flüssen aus der Kategorie «Medium» ergänzt wurden (Tabelle 1). Der damit generierte Output wird folglich als **VT(Flüsse (mittel))** bezeichnet. Die beiden Kategorien mit den kleinsten Abflüssen «Low» und «Very low» wurden nicht berücksichtigt.

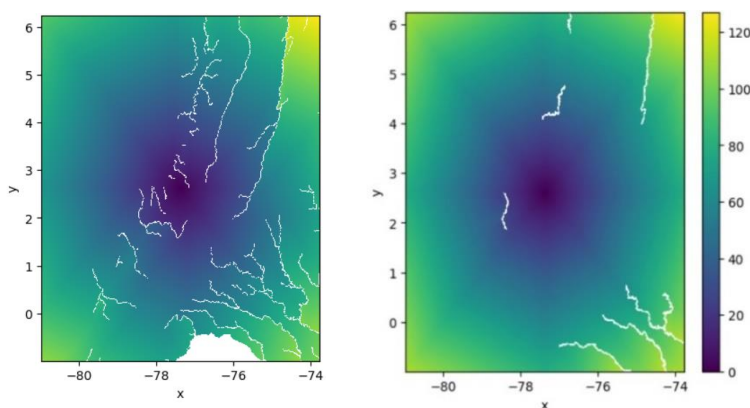


Abbildung 7: Grafik zeigt den Output der Kostenfunktion bei einer nach Flüssen (links Flüsse (mittel) und rechts Flüsse (gross)) gewichteten VT. Skalierung zeigt die Kosten, welche benötigt werden, um den entsprechenden Punkt vom SP aus zu erreichen.

9.2.4 VT(Gebirge)

In diesem Schritt wurden die Kosten, welche durch die Topografie bedingt sind, berücksichtigt. Als Kostenraster wurde das DHM verwendet. Wobei der Einfluss der Gewichtung nur minimal zu erkennen ist im Vergleich zu Abbildung 7.

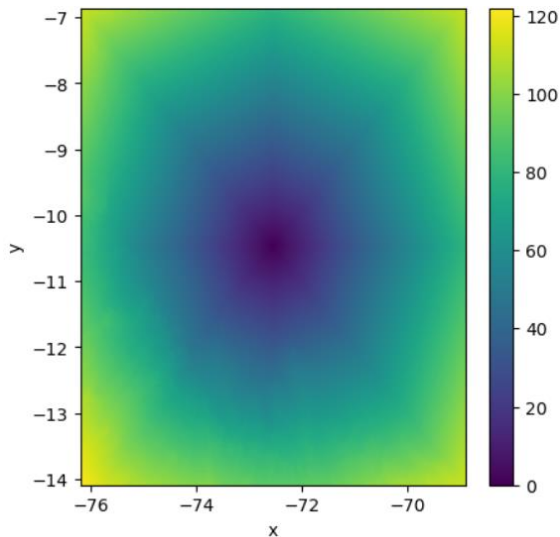


Abbildung 8: Grafik zeigt den Output der Kostenfunktion bei einer VT(Gebirge). Skalierung zeigt die Kosten, welche benötigt werden, um den entsprechenden Punkt vom SP aus zu erreichen.

9.3 EVALUIERUNG

9.3.1 Jaccard-Index:

Der Jaccard-Index (JI) ist eine gute Möglichkeit, um zwei Mengen auf ihre Ähnlichkeit zu überprüfen. Hierbei werden die beiden zu vergleichenden Mengen auf gemeinsame Elemente untersucht. Die Anzahl gefundener gemeinsamer Elemente wird anschliessend durch die Anzahl aller in beiden Datensätzen vorkommenden Elemente geteilt. Schlussendlich erhält man einen Wert zwischen 0 (keine Übereinstimmung) und 1 (perfekte Übereinstimmung). Der Jaccard-Index wird in vielen wissenschaftlichen Disziplinen angewendet und kann in verschiedenen Variationen berechnet werden, wobei in dieser Arbeit der «Basic Jaccard Index» relevant ist (Shi, Ngan and Li, 2014; Da Fontoura Costa, 2021; Glen, 2022).

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

A = erste zu vergleichende Menge

B = zweite zu vergleichende Menge

(Da Fontoura Costa, 2021)

Dieses Konzept kann auch für räumliche Flächen genutzt werden. Dabei ist entscheidend, dass zwei zu vergleichende Flächen einer klaren Referenz (in diesem Fall einer gemeinsamen Sprache) zugewiesen werden können. Zwei Flächen können auf ihre räumliche Übereinstimmung geprüft werden, indem den beiden Mengen jeweils alle Zellen einer Fläche zugewiesen werden. Wichtig dabei ist, dass die räumliche Position (Longitude;Latitude) jeder Zelle als zusätzliche Information gespeichert wird. Im Anschluss werden die beiden Mengen auf Zellen untersucht, welche dieselbe Position besitzen. Dies kann laut der obigen $A \cap B$ betrachtet werden. Anschliessend wird $A \cup B$ gebildet, was in diesen Fall der Gesamtzahl aller Zellen, die den beiden Flächen zugewiesen wurden, entspricht. Der JI einer Fläche ergibt sich aus $A \cap B / A \cup B$ und wurde für jede Sprache, die sowohl im Glottolog- als auch im Ethnologue-Datensatz vorkommt, berechnet.

Abbildung 9 zeigt eine optische Darstellung des JI als Beispiel.

Der JI ergibt sich in folglich aus:

$$JI = \frac{\text{(Zellen mit Wert 3)}}{\text{(Zellen mit Wert 2)+(Zellen mit Wert 1)– (Zellen mit Wert 3)}}$$

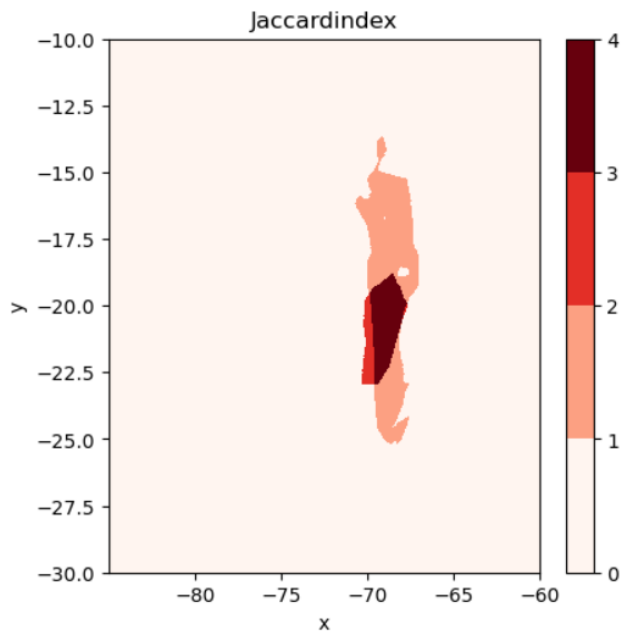


Abbildung 9: Visuelle Darstellung des JI der Sprache 'Central Aymara'. Wobei Zellen die sowohl dem mit VT generierten Sprachareal als auch dem entsprechenden Referenzsprachareal zugewiesen wurden, den Wert 3 haben. Zellen, die dem mit VT generiertem Sprachareal zugewiesen wurde, aber nicht dem Referenzsprachareal haben den Wert 2. Zellen, die der Referenzsprachareal, aber nicht dem mit VT generiertem Sprachareal zugewiesen wurden, haben den Wert 1. Zellen, die keinem der beiden Flächen zugewiesen wurden, haben den Wert 0.

9.4 ANALYSE

9.4.1 Differenzen der Verteilungen

Um zu analysieren, wie sich der JI der einzelnen Sprachareale verändert, wurde jeweils für jede der evaluierten Sprachen die Differenz zwischen dem JI mit Gewichtung und dem JI ohne Gewichtung berechnet. Hier wurden stets die Ethnologuezentroide als SP verwendet.

Daraus entstanden folgende Differenzen:

JI der generierten Sprachareale VT(Gebirge) - JI der generierten Sprachareale VT(ungewichtet) -> **Diff(Gebirge)**

JI der generierten Sprachareale VT(Flüsse(gross)) - JI der generierten Sprachareale VT(ungewichtet) -> **Diff(Flüsse(gross))**

JI der generierten Sprachareale VT(Flüsse(mittel)) - JI der generierten Sprachareale VT(ungewichtet) -> **Diff(Flüsse(mittel))**

Weiter wurde für jede der evaluierten Sprachen die Differenz des JI mit Glottologpunkten als SP und des JI mit Ethnologuepunkten als SP berechnet. Dabei wurde die VT(ungewichtet) verwendet.

JI der generierten Sprachareale VT(ungewichtet) mit Glottologpunkten als SP - JI der generierten Sprachareale VT(ungewichtet) mit Ethnologuezentroiden als SP ->

Diff(Seedpoints)

9.4.2 Darstellung

Um Selbstkorrelationseffekte zu vermeiden, wurden die Evaluierungswerte (JI) der einzelnen Gewichtungen bzw. deren Differenzen den Vektordaten bzw. deren Polygone/Multipolygon von Ethnologue zugewiesen. In einem weiteren Schritt wurden die Verteilungen der Evaluierungswerte gemäss den Fragestellungen visuell dargestellt und analysiert. Dazu wurden Karten, Histogramme sowie Boxplotdiagramme verwendet.

9.4.3 Flussdichte

Als Flussdichte wird die Anzahl von Rasterzellen, welche als Fluss kategorisiert wurden, im Verhältnis zu der Gesamtzahl von Rasterzellen in einem Sprachgebiet bezeichnet.

9.4.4 Durchschnittliche Höhe

Als durchschnittliche Höhe eines Sprachgebiets wird der Durchschnitt der Werte des DHM, welche einer spezifischen Sprache zugewiesen wurden, bezeichnet.

9.4.5 Regressionsanalyse

Um eine die Korrelationen zwischen dem JI der einzelnen Polygone und der Flussdichte bzw. durchschnittlichen Höhe zu ermitteln, wurde eine einfache lineare Regressionsanalyse (RG) durchgeführt. Da die RG nur dafür verwendet wurde, um eine allfällige Korrelation aufzuzeigen, und nicht, um ein Modell für weitere Vorhersagen bezüglich der Variablen zu erstellen, lag der Fokus vor allem auf der optischen Analyse und kann als erweiterter Scatterplot betrachtet werden. Daher waren für die weitere Analyse nur die Modellgüte (R^2) und die Steigung der Geraden relevant.

Die Modellgüte beschreibt dabei den linearen Zusammenhang der untersuchenden Variablen. Der Wert kann in diesem Kontext zwischen 0 (kein linearer Zusammenhang) bis 1 (perfekter linearer Zusammenhang) variieren.

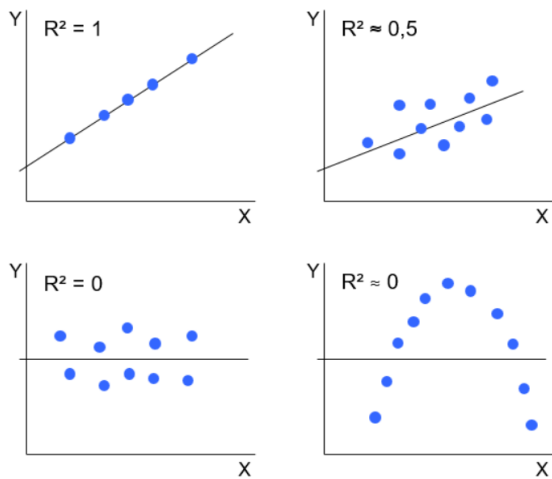


Abbildung 10: Beispiele der Modellgüte in Abhängigkeit der Korrelation der zu vergleichenden Variablen (Sonneck, 2020)

Ein weiterer relevanter Faktor ist die Steigung (Regressionskoeffizient). Sie zeigt an, wie stark der Einfluss der unabhängigen Variabel (x-Achse) auf die abhängige Variable (y-Achse) ist. Die Analyse der Steigung erfolgte dabei rein qualitativ. Eine positive Steigung zeigt eine positive Korrelation, während eine negative Steigung auf eine negative Korrelation hinweist (Loy, 2015; Sonneck, 2020).

9.5 WORKFLOW

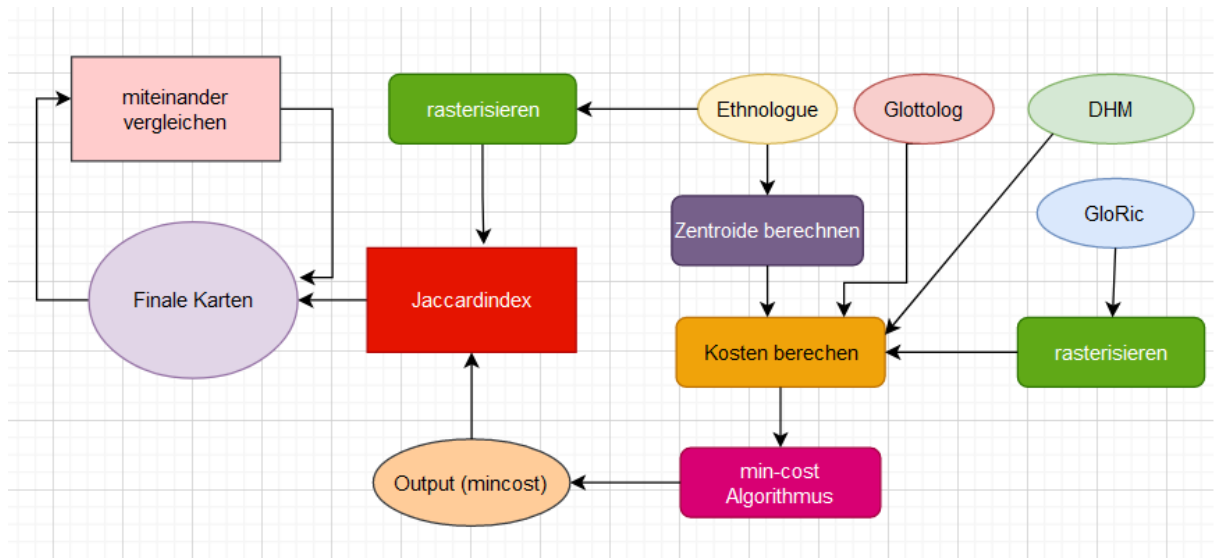


Abbildung 11: Grafischer Überblick auf die im Kapitel 9 beschriebenen Methoden. Rechtecke beschreiben Prozesse. Kreise beschreiben Input- respektive Outputdaten.

10 RESULTATE

Insgesamt wurden 295 Polygone untersucht. Daher gilt für alle Diagramme/Berechnungen $n = 295$.

Der JI(Jaccardindex) zwischen den mit VT(Voronoi-Tesselierung) generierten Spracharealen und den Referenzspracharealen wird zur Vereinfachung immer in der Form: **JI(Gewichtung; Seedpoints)** genannt. Beispielsweise wird der JI, welcher die Referenzsprachareale von Ethnologue mit den Spracharealen, welche durch VT mit Ethnologuezentroiden als Seedpoints (SP) und VT(Gebirge) generiert wurden, als JI(VT(Gebirge);Ethnologue) bezeichnet.

Zur grafischen Darstellung werden immer die Polygone von Ethnologue verwendet.

10.1 VT VERGLEICH ETHNOLOGUE VS. GLOTTOLOG

In Abbildung 12 ist zu erkennen, dass der JI(VT(ungewichtet);Ethnologue) bei über 120 Polygonen einen Wert unter 0.1 annimmt. Es handelt sich dabei um eine rechtsschiefe Verteilung. Zwar finden sich auch mit VT generierte Flächen mit einem hohen JI (0.8), allerdings sind diese im Vergleich zu den kleinen Werten fast bedeutungslos. Bei der Verteilung der Werte von JI(VT(ungewichtet);Glottolog)) liegt dasselbe Muster vor.

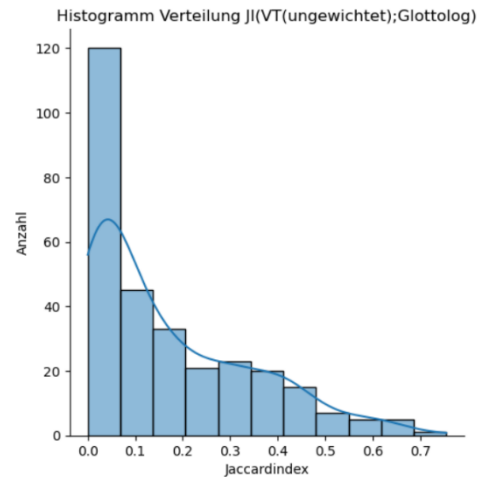
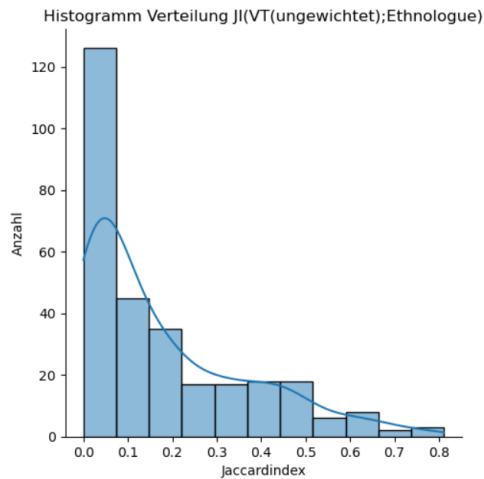


Abbildung 12: Verteilungen des $JI(VT(ungewichtet));Ethnologue$ (links) und des $JI(VT(ungewichtet));Glottolog$ (rechts) der einzelnen Polygone – wobei die y-Achse der Anzahl Polygone entspricht. Da es sich beim JI um kontinuierliche Werte zwischen 0 und 1 handelt, entsprechen die Balken aggregierten Werten.

Abbildung 13 zeigt noch genauer, dass sich die Wahl der SP kaum Auswirkungen auf die Verteilung hat. Vergleicht man die Verteilung der $JI(VT(ungewichtet);Ethnologue)$ mit der Verteilung der $JI(VT(ungewichtet);Glottolog)$, erkennt man keinen signifikanten Unterschied. Bei beiden Verteilungen liegt der Median leicht über 0.1. Allerdings finden sich auch einige Ausreisser, welche höhere JI -Werte anzeigen. Nur etwa ein Viertel der Polygone besitzt einen JI über 0.3. Allerdings existieren auch einige Ausreisser, welche höhere JI -Werte anzeigen.

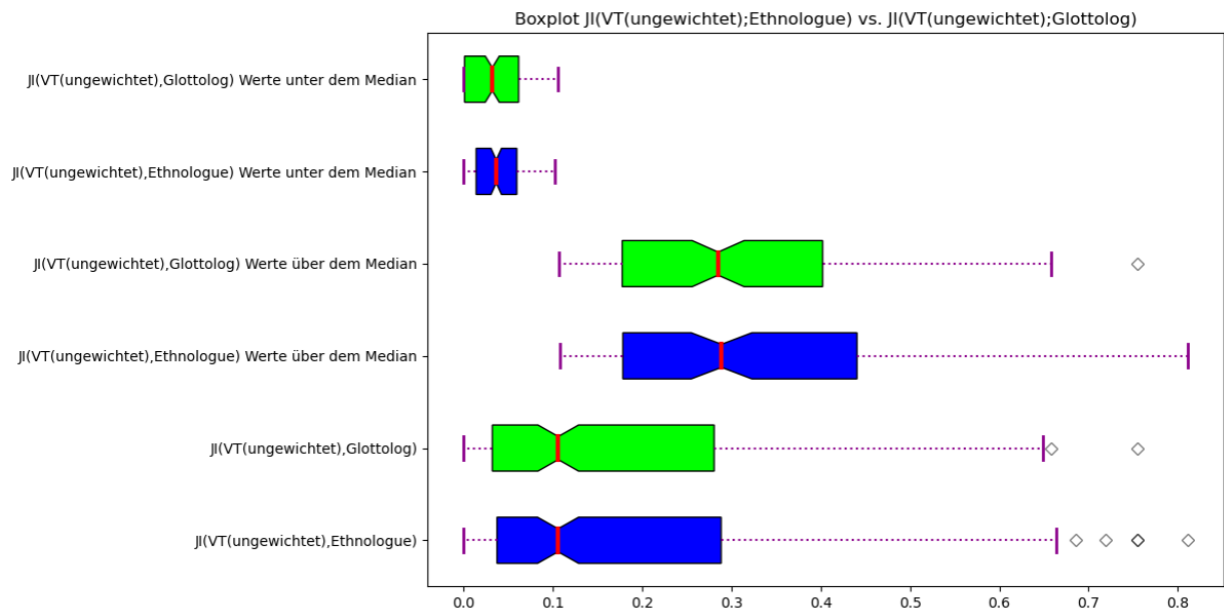


Abbildung 13: Verteilungen von Abbildung 12 als Boxplot. Weiter werden die Verteilungen bzw. die Werte unter bzw. über dem Median miteinander verglichen.

Wie in Abbildung 14 und Abbildung 15 zu erkennen ist, lassen sich keine Muster in Bezug auf Grösse, Form und Lage der Polygone in Bezug auf den JI erkennen. Einen besonders hohen JI findet man in Trinidad und Tobago, wobei es sich um eine Insel handelt.

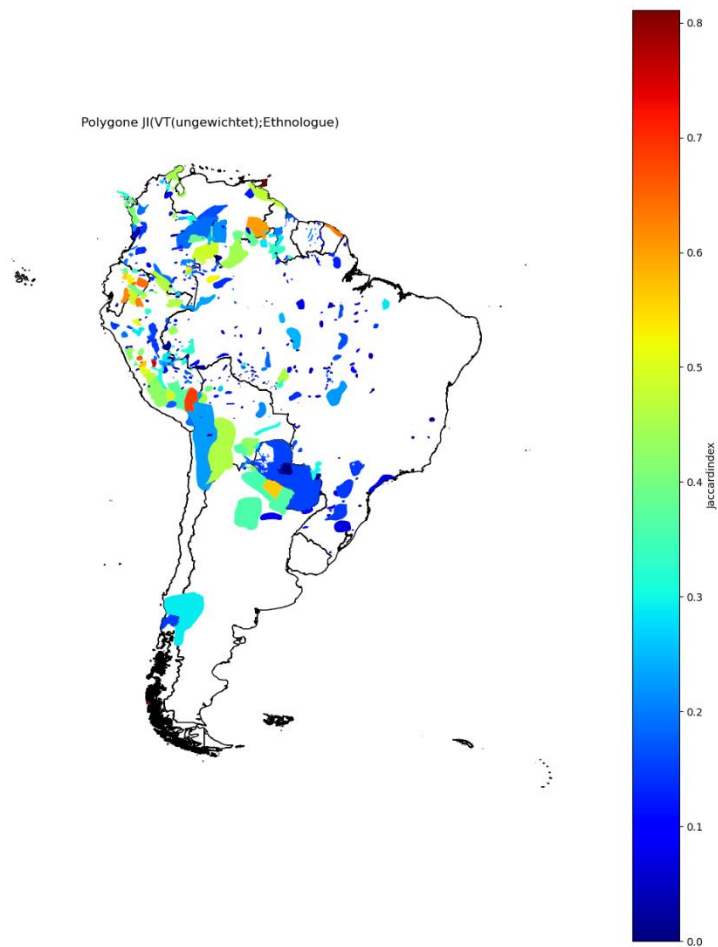


Abbildung 14: Räumliche Lage der Polygone. Skalierung entspricht dem $JI(VT(ungewichtet);Ethnologue)$ der einzelnen Polygone.

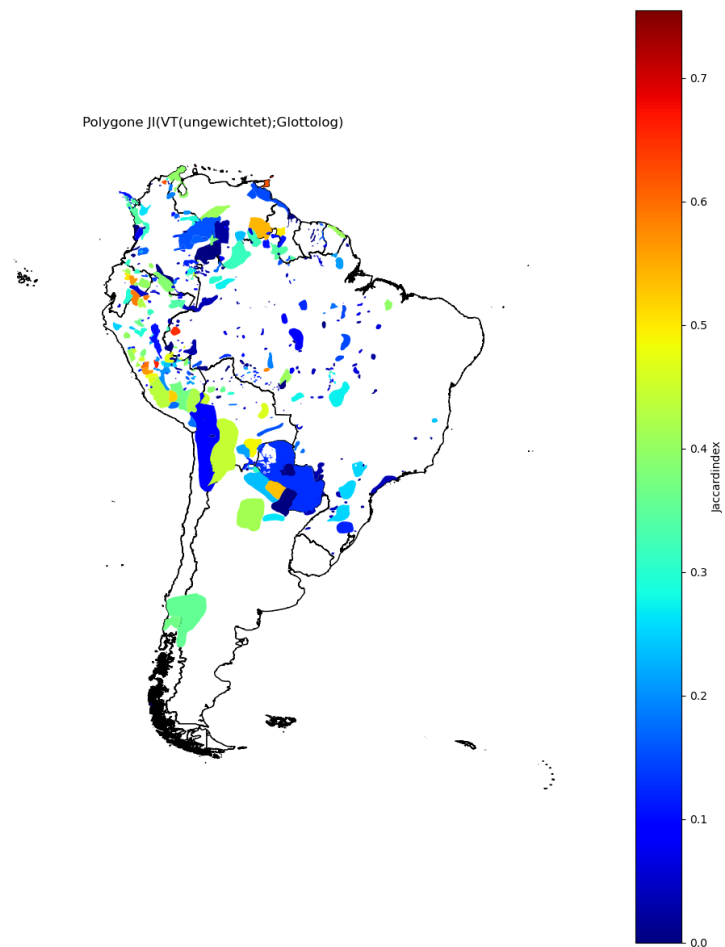


Abbildung 15: Räumliche Lage der Polygone. Skalierung entspricht dem $JI(VT(ungewichtet);Glottolog)$ der einzelnen Polygone.

Obwohl die Verteilungen in Abbildung 13 sich bei der Wahl der SP nicht gross unterscheiden, sieht dies für einzelne Polygone teilweise anders aus. Wie in Abbildung 16 erkennbar ist, variiert die $Diff(Seedpoints)$ bei einzelnen Polygonen im Bereich von -0.2 bis 0.6. Es lassen sich keine räumlichen oder geometrischen Muster für die Polygone, bei welchen die Abweichung der JI besonders gross ist, erkennen.

Teilweise sind einzelne Werte von Polygonen des $JI(VT(ungewichtet);Glottolog)$ höher als die Werte der entsprechenden Polygonen von $JI(VT(ungewichtet);Ethnologue)$. Auffallend ist

hier, dass es sich meist um kleinere Polygone handelt. Bezüglich der räumlichen Verteilung lässt sich auch hier keine klare Aussage treffen.

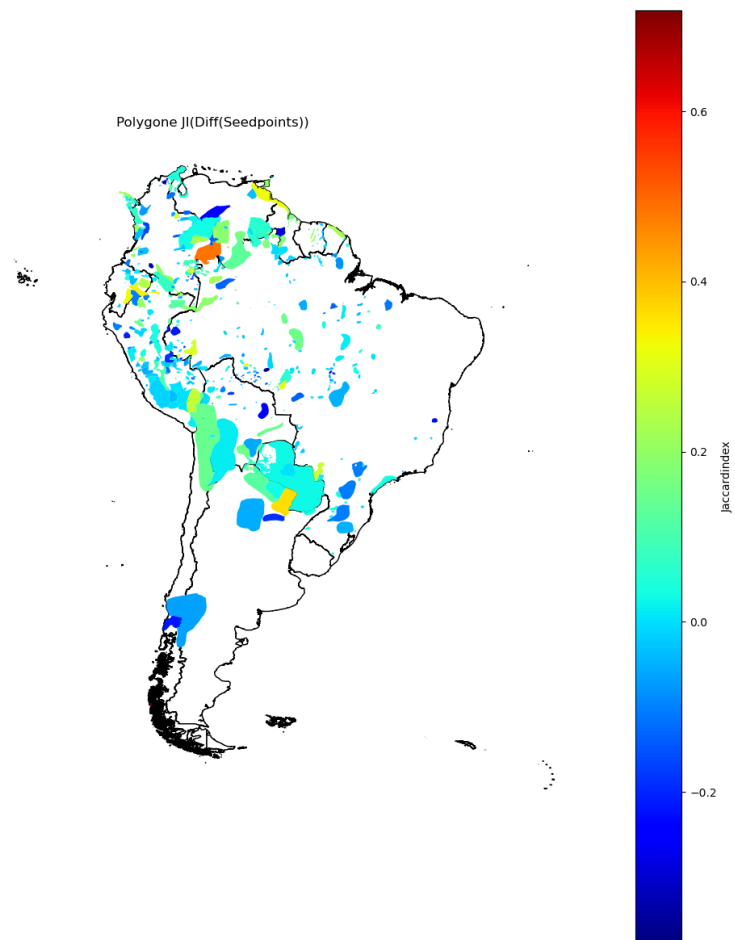


Abbildung 16: Räumliche Lage der Polygone. Skalierung entspricht der Diff(Seedpoints) der einzelnen Polygone.

Wie in Abbildung 17 zu erkennen ist, reichen die Abstände der SP von 0 km bis 40 km. Die Verteilung ist rechtsschief. Es liegen sehr viele SP mit einem Abstand nahe 0 vor. Weiter ist keine Korrelation zwischen dem Abstand der SP und Diff(Seedpoints) zu erkennen. Sowohl Steigung als auch Modellgüte sind nahe 0, wie man in Abbildung 18 deutlich erkennen kann.

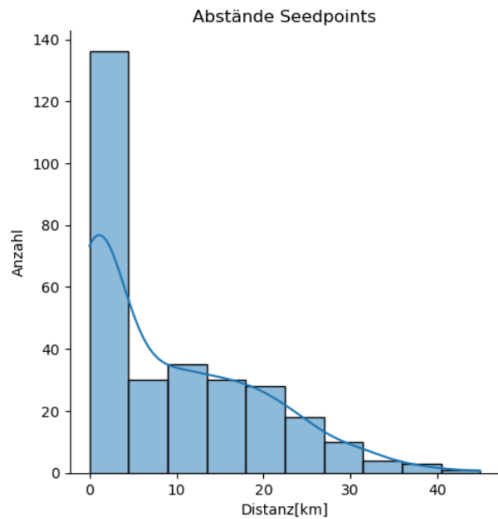


Abbildung 17: Verteilung der Abstände der Glottologpunkte zu den Ethnologuezentroiden. Ein Abstand bezieht sich immer auf eine Sprache.

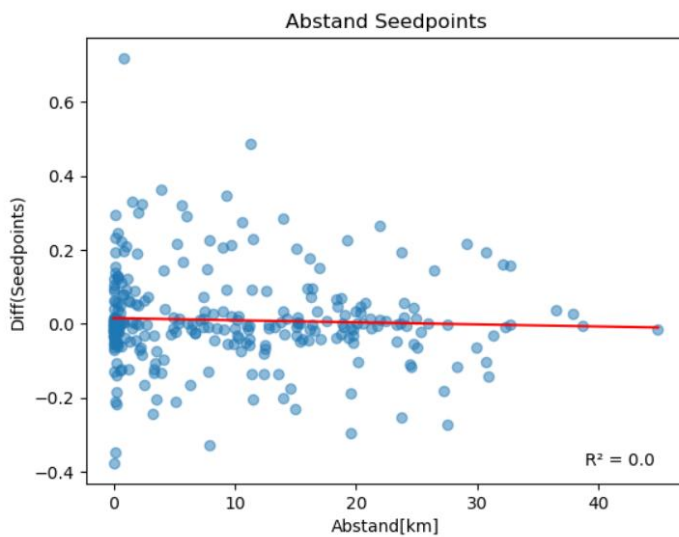


Abbildung 18: Grafik vergleicht die Diff(Seedpoints) (Abbildung 16) in Abhängigkeit zu den Abständen der Seedpoints (Abbildung 17) der gleichen Sprache.

10.2 WELCHE AUSWIRKUNGEN HAT DIE GEWICHTUNG DER VT

Im nächsten Schritt soll aufgezeigt werden, ob sich, wie angenommen, die Werte des JI der einzelnen Polygone verändern, wenn eine gewichtete VT verwendet wird.

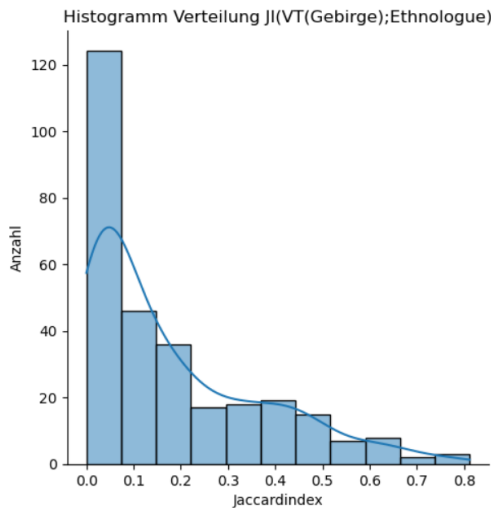


Abbildung 19: Verteilungen des $JI(VT(\text{Gebirge});\text{Ethnologue})$ der einzelnen Polygone – wobei die y-Achse der Anzahl Polygone entspricht. Da es sich beim JI um kontinuierliche Werte zwischen 0 und 1 handelt, entsprechen die Balken aggregierten Werten.

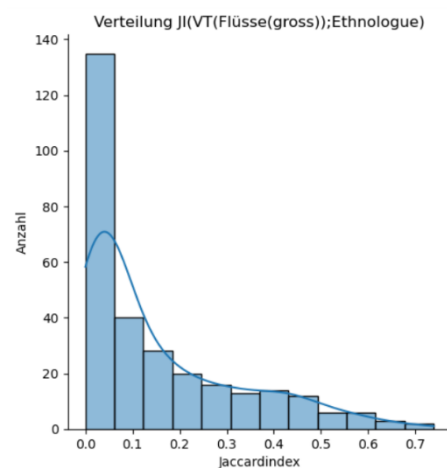
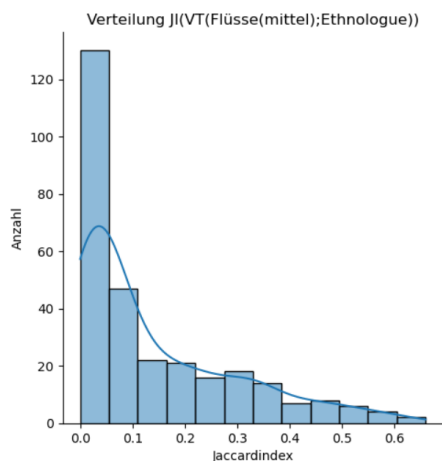


Abbildung 20: Verteilungen des $JI(\text{Flüsse}(\text{mittel});\text{Ethnologue})$ (links) und des $JI(\text{Flüsse}(\text{gross});\text{Ethnologue})$ (rechts) der einzelnen Polygone – wobei die y-Achse der Anzahl Polygone entspricht. Da es sich beim JI um kontinuierliche Werte zwischen 0 und 1 handelt, entsprechen die Balken aggregierten Werten.

Betrachtet man die Histogramme in Abbildung 19 und Abbildung 20, so ist hier eine Rechtsschiefheit bei allen Verteilungen zu erkennen. Die Gewichtung der VT scheint keinen grossen Einfluss auf den JI zu haben. Mit Blick auf Abbildung 21 wird dies noch deutlicher. So unterscheiden sich die Verteilungen des $JI(VT(\text{Gebirge});\text{Ethnologue})$ und des

$JI(VT(ungewichtet);Ethnologue)$ gar nicht. Werden die Flüsse als Gewichtung genommen, liegt ein negativer Effekt vor. Der Median der betreffenden Verteilungen ist erkennbar niedriger als bei der $VT(ungewichtet)$.

Weiter spielt es keine entscheidende Rolle, ob nur grosse oder auch mittelgrosse Flüsse bei der Gewichtung berücksichtigt wurden. Auffallend sind jedoch die vielen Ausreisser bei den Gewichtungen mit Flüssen. Dies kann darauf schliessen lassen, dass es stark von den örtlichen Gegebenheiten abhängig ist, ob ein Fluss Einfluss auf die Sprachgrenze hat.

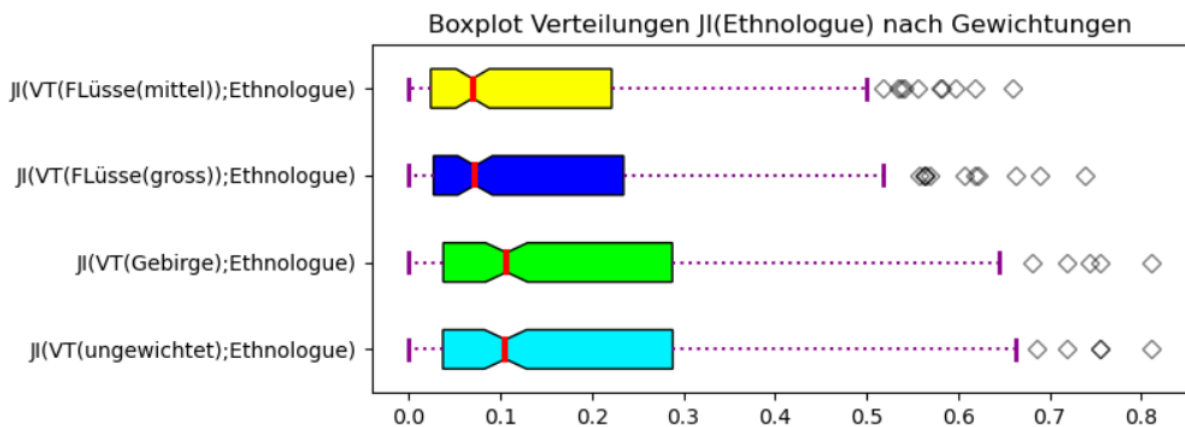


Abbildung 21: Verteilungen von Abbildung 12 (links), Abbildung 19 und Abbildung 20 (beide)

In Abbildung 22 sind vor allem die sehr geringen Abweichungen des $JI(VT(ungewichtet);Ethnologue)$ und des $JI(VT(Gebirge);Ethnologue) \rightarrow (Diff(Gebirge))$ auffallend in einem Intervall von 0.05 bis -0.03. Dies lässt darauf schliessen, dass die Gewichtung wenig Einfluss hat. Dennoch sind räumliche Muster erkennbar. Trotz der im Allgemeinen sehr kleinen Abweichungen sind die grössten Abweichungen meist im Andengebiet zu finden. Ausserdem handelt es sich dabei um sehr kleine Polygone.

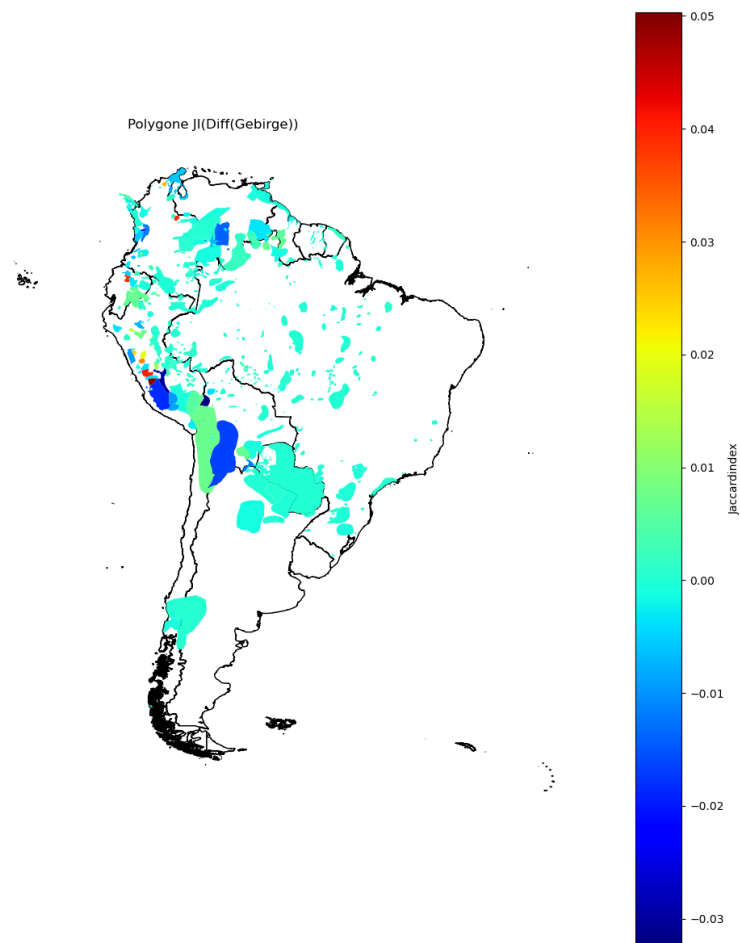


Abbildung 22: Räumliche Lage der Polygone. Skalierung entspricht der Diff(Gebirge) der einzelnen Polygone.

10.3 KORRELATION MI DEN LANDSCHAFTSVARIABLEN

Der folgende Abschnitt soll die Korrelation zwischen der Diff(Gebirge) (Abbildung 22) und der durchschnittlichen Höhe ü. Meer eines Polygons sowie die Korrelation der Diff(Flüsse(gross)) bzw. Diff(Flüsse(mittel)) und der jeweiligen Flussdichte der Polygone analysieren.

In Abbildung 23 ist zu erkennen, dass ein Grossteil der mit VT erhaltenen Sprachgebiete eine durchschnittliche Höhe von unter 1000 Meter hat. Die Werte von Diff(Gebirge) (Abbildung 22) ist in diesem Bereich nahe 0. Bei Werten, bei denen die durchschnittliche Höhe über 1000

Meter über Meer liegt, ist die Streuung grösser und reicht sowohl in den positiven als auch in den negativen Bereich. Die Steigung scheint von der Streuung kaum beeinflusst und zeigt nur eine minim positive Korrelation an. Eine Modellgüte von 0.03 lässt auf keinen linearen Zusammenhang zwischen den Variablen schliessen.

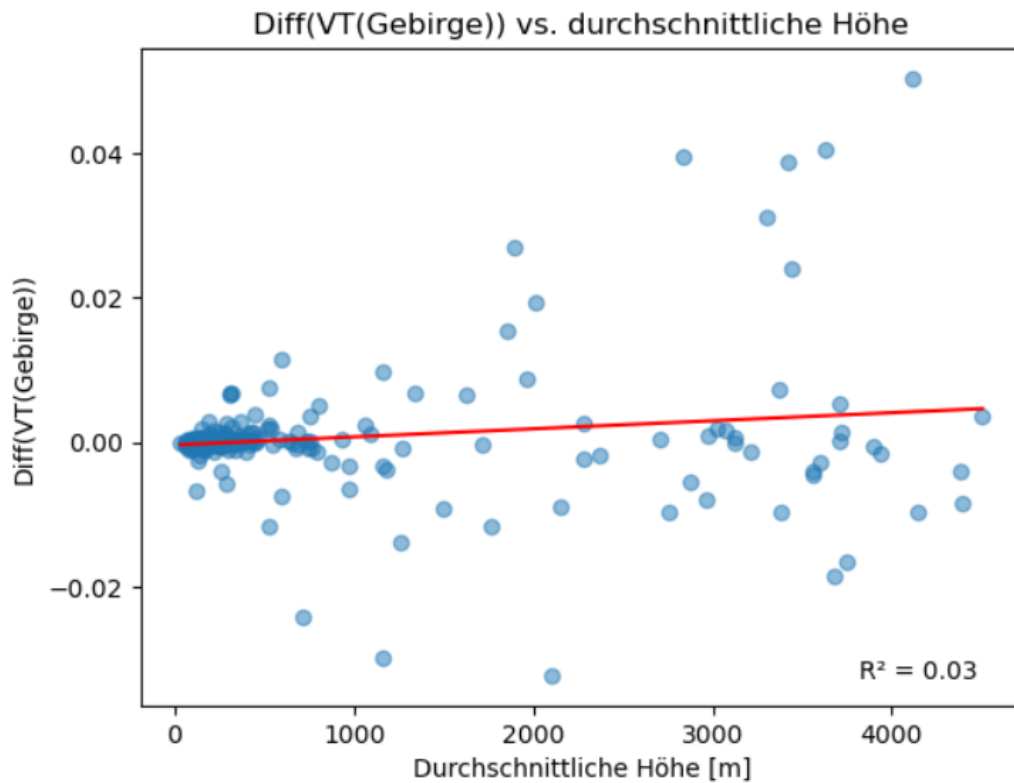


Abbildung 23: Grafik vergleicht die Diff(Gebirge) in Abhängigkeit zu der durchschnittlichen Höhe des gleichen Sprachareals.

In Abbildung 24 (links) ist zu sehen, dass ein grosser Teil der mit VT erhaltenen Sprachgebiete nur eine sehr geringe Flussdichte aufweist, sofern die VT mit grossen Flüssen durchgeführt wurde. Auffallend ist, dass die Schwankungen der Diff(Fluss(gross)) viel grösser sind als die der Diff(DHM). So existieren in Abbildung 24 (links) Werte zwischen -0.2 und 0.8.

Allerdings lässt sich auch hier keine wirkliche Korrelation erkennen. So ist die Steigung nahe 0. Weiter deutet eine Modellgüte von 0.0 auf keinen linearen Zusammenhang der Variablen hin.

Ein ähnliches Bild sieht man bei Abbildung 24 (rechts). Hier sind die Flussdichten im Schnitt pro generierter Fläche etwas höher. Weiter erstreckt sich hier das Intervall der

Diff(Flüsse(mittel)) von -0.4 bis 0.8. Auch in diesem Fall fällt die Steigung gering aus und die Modellgüte von 0.01 lässt nicht auf eine lineare Korrelation schließen.

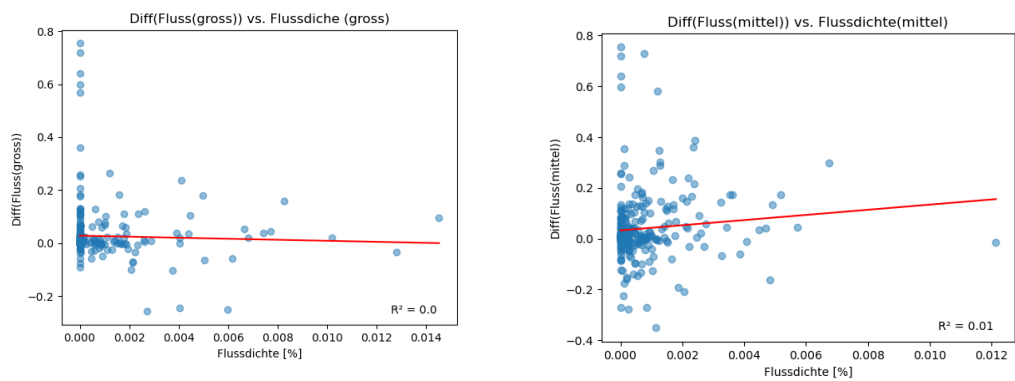


Abbildung 24: Grafik vergleicht die Diff(Flüsse(gross)) (links) bzw. die Diff(Flüsse(mittel)) (rechts) in Abhängigkeit zu der jeweiligen Flussdichte des gleichen Sprachareals.

11 DISKUSSION

11.1 VORONOI-TESSELIERUNG

Basierend auf den tiefen Werten des JI(Jaccardindex) der einzelnen Polygone scheint die VT(Voronoi-Tesselierung) ungeeignet, um Sprachareale von einem Seedpoint (SP) aus zu generieren, welche eine grosse Ähnlichkeit zu Referenzspracharealen aus Ethnologue haben. Dies äussert sich vor allem durch die rechtsschiefen Verteilungen in den Abbildung 12, Abbildung 19 und Abbildung 20, welche für einen Grossteil der Polygone tiefe JI-Werte anzeigen.

11.1.1 Sprache ist nicht kreisförmig

Bei der VT gilt die Annahme, dass sich einzelne Gebiete von einem SP aus kreisförmig ausbreiten, sofern keine natürliche Barriere passiert wird. Dies bedeutet, dass zwei Sprachen bei dieser Modellannahme sich nicht weiter ausbreiten, sobald sie aufeinandertreffen. Wie in Abbildung 14 deutlich erkennbar ist, finden sich immer wieder Polygone, die eine längliche Form haben und somit dieser Annahme widersprechen. Dies lässt darauf schliessen, dass jene Modellannahme für Sprachen ungeeignet ist.

Wie (Tulio *et al.*, 2021) aufzeigt, führen verschiedene Ereignisse dazu, dass sich Sprachen splitten, aussterben oder ineinander aufgehen. (Tulio *et al.*, 2021) spricht dabei von Schocks, welche zu einer Änderung der Bevölkerungsgrösse in einer Region führen. Diese Schocks können negativ -> rapider Bevölkerungsschwund, beispielsweise ausgelöst durch eine Pandemie, oder positiv -> rapider Bevölkerungszuwachs, etwa ausgelöst durch eine neue Erfindung in Agrikultur, sein. So kann eine Seuche einen Landstrich entvölkern und in zwei Sprechergruppen aufsplitten, wodurch zwei neue Sprachen entstehen. Da ein solches Ereignis oft regional sehr beschränkte Auswirkungen haben kann, deutet wenig auf eine kreisförmige Expansion hin, was auch die Resultate von (Tulio *et al.*, 2021) belegen. Andere Literatur, welche sich mit der Expansion von Sprache beschäftigt (Bouckaert *et al.*, 2012; Gavin *et al.*, 2013; Neureiter *et al.*, 2021), lässt ebenfalls nicht auf eine kreisförmige Ausbreitung der Sprache schliessen.

11.1.2 Seedpoint

Zudem muss berücksichtigt werden, dass die Schwerpunkte (Zentroide) der Ethnologuepolygone nicht als Ursprung der sich von dort ausbreitender Sprache gesehen werden dürfen, da es sich hier um eine rein geometrische Einheit ohne jeglichen thematischen Bezug handelt. In Kapitel 11.2 wird noch genauer auf die Problematik der Zentroide als SP eingegangen.

11.2 UNTERSCHIEDE BEI DER WAHL DES SP

Die Verteilung des $JI(VT(\text{ungewichtet}); \text{Ethnologue})$ unterscheidet sich kaum von der Verteilung des $JI(VT(\text{ungewichtet}); \text{Glottolog})$. Einzelne Polygone haben zum Teil verschiedene Werte in Bezug auf den JI , allerdings steht dieser in keiner Abhängigkeit von dem Abstand der SP (Abbildung 18), was der Hypothese 1 a) widerspricht.

11.2.1 Keine Abhängigkeit des Abstandes der Seedpoints

Problematisch ist hier, dass die Grösse der Polygone nicht einbezogen wurde. So kann der Abstand der SP im Vergleich zu grossen Polygonen minimal sein, was aber bei kleinen Polygonen im Vergleich doch ein grosser Abstand ist. Dies ist gut Abbildung 25 ersichtlich. Es handelt sich hier um die Sprache Qawasqar. der Abstand der SP beträgt hier nur etwa 0.8 km, was im Vergleich zum maximalen Wert von 40 km wenig ist. Dennoch sind die SP hier weit auseinander im Vergleich zur Gesamtgrösse des Polygons, was sich auch auf die $\text{Diff}(\text{Seedpoint})$ auswirkt (0.71875). Um den Einfluss auf den Abstand der SP auf das Modell genau beurteilen zu können, ist es daher wichtig, die Grösse des Polygons in das Modell einzubeziehen. Eine Möglichkeit wäre die Regressionsanalyse mit zwei unabhängigen Inputvariablen (Abstand SP und Grösse des Polygons). Sinnvoll wäre für die Grösse eine logarithmische Skalierung, da sich die Verteilung der Grösse auf Grund der vielen kleinen Polygone stark verzerren könnte.

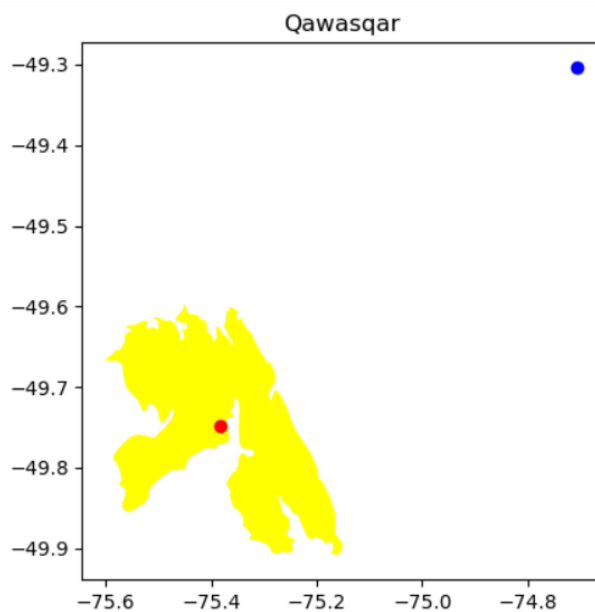


Abbildung 25: Ethnologuepolygon der Sprache Qawasqar (gelb) mit entsprechenden SP von Ethnologue (rot) bzw. Glottolog (blau)

11.2.2 Wo ist Glottolog besser?

Auffallend bei Abbildung 16 ist, dass Polygone, bei welchen der JI besser funktioniert, wenn Glottologpunkte als SP dienen (negative Werte), eher klein sind. Dafür gibt es mehrere mögliche Erklärungen. Eine wäre, dass es sich hierbei um Multipolygone handelt. Dabei besteht die Wahrscheinlichkeit, dass der Zentroid ausserhalb der Polygone liegt, was dazu führt, dass Glottologpunkte als SP geeigneter ist.

Weiter besteht eine Unsicherheit bezüglich der Zentroide von Ethnologue darin, dass es sich um rein geometrische Schwerpunkte handelt. Diese sagen aber in keiner Weise aus, wie die Sprecherzahl innerhalb eines Polygons verteilt ist. Es fehlen jegliche Informationen darüber, wie die Bevölkerung, welche die untersuchende Sprache spricht, verteilt ist. Dies ist vor allem in grossen Spracharealen problematisch, da hier die Wahrscheinlichkeit von Schwankungen bezüglich der Bevölkerung höher ist. Als besonders problematisch können dabei Spanisch und Portugiesisch angesehen werden. Diese erstrecken sich über weite Teile des Kontinents. Der Schwerpunkt wirkt hier willkürlich. Zwar wurden die beiden Sprachen nicht für die Validierung berücksichtigt, trotzdem verzerren sie die generierten durch VT generierten Sprachareale in der Region nahe ihrem Schwerpunkt.

Des Weiteren können sich die Polygone des Ethnologuedatensatzes überschneiden. Damit ist die Summe aller Polygonflächen grösser als die Fläche, welche sie abdecken. Hingegen wird

bei der VT eine Sprache immer nur einem Gebiet zugewiesen – wodurch die Summe der Flächen gleich gross ist, wie das Gebiet, welches sie abdecken. Dies kann zu räumlichen Verzerrungen bezüglich des Ethnologuezentrums führen.

11.3 GEBIRGE

Wie erwartet, finden sich im Andengebiet die Polygone mit der grössten Verbesserung des JI, wenn man die VT mit dem DHM gewichtet (Diff(Gebirge)), was der Hypothese 1 b) entspricht. Diese Resultate unterstützen die Meinung der Literatur, dass Gebirge auf Grund des erschwerten Kontakts zweier Sprechergruppen einen Einfluss auf Sprachgebiete und deren Grenzen haben (Stepp, Castaneda and Cervone, 2005; Gavin and Stepp, 2014). All dies unterstützt zudem die von (Hua *et al.*, 2019) erwähnte Isolationshypothese in Bezug auf die Topografie.

11.3.1 Thematisch

Zwar finden sich auch Polygone im Andengebiet mit Werten bei 0 oder darunter, aber besonders im zentralen bis nördlichen Peru nahe der Küste trifft man ein Muster von Polygonen mit vergleichsweise hohen Werten an. Tatsächlich ist die sprachliche Situation in dieser Region speziell. So haben sich hier im Vergleich zu südlicheren Regionen Sprachen sehr unterschiedlich voneinander entwickelt, was zu hohen Sprachunterschieden bei kleinen räumlichen Abständen führte (genealogische Vielfalt). Im Vergleich dazu war der Kontakt der einzelnen Sprechergruppen in Südperu viel häufiger, was sich in ähnlichen Sprachen widerspiegelt. Dies kann man auch anhand der grösseren Polygone im Süden Perus erkennen (Urban, 2021). Der erhöhte Kontakt der Sprechergruppen im Süden Perus ist vor allem auf die beiden Kulturen Tiahuanaco und Huari zurückzuführen. Beide Kulturen hatten einen starken religiösen Einfluss auf die benachbarten Sprechergruppen, sodass ihre Zentren weiträumig als Pilgerstätte bekannt waren und somit von Personen aus anderen Sprechergruppen besucht wurden. Des Weiteren hatten die Kulturen eine expansive Wirtschafts- bzw. Militärpolitik. Diese Gründe führten dazu, dass der Kontakt trotz der gebirgigen Region mit benachbarten Sprechergruppen sehr hoch war – im Vergleich zu nördlicheren Gebieten innerhalb von Peru (Isbell, 2008; Teichert, 2021). Die Gewichtung des sozioökonomischen Austausches überwiegt

in diesem Fall die Kosten der natürlichen Barriere, wodurch sich eine Sprache über ein Gebirge hinweg ausbreiten konnte.

11.3.2 Methodik

Diff(Gebirge) (max = 0.05) der einzelnen Polygone sind verglichen mit den Werten der Diff(Seedpoint) (max = 0.6) und der Diff(Flüsse) (max = 0.8) sehr gering. Dies lässt darauf schliessen, dass die Gewichtung mit dem DHM eher ungeeignet scheint, da Unterschiede zur VT ohne Gewichtung kaum ins Gewicht fallen.

Die Gewichtung mit dem DHM basiert komplett auf der Hiking-Funktion von Tobler (Tobler, 1993). Diese berechnet die Geschwindigkeit einer wandernden Person in Abhängigkeit zur Steigung. Ob diese Annahme für den Kontakt zweier Sprechergruppen taugt, ist zu bezweifeln. Zum einen geht das Modell von einer einzigen Person aus. Damit ein Kontakt entsteht, der dazu führt, dass sich die Sprachen gegenseitig beeinflussen bzw. vermischen, muss die Mobilität einer Gruppe betrachtet werden.

Weiter sagt die Hiking-Funktion nur aus, dass die Geschwindigkeit bei grösserer Steigung abnimmt. Dies sagt allerdings nichts über die Umstände aus, die eine Sprechergruppe dazu bringen würde, ein Tal zu verlassen und Kontakt mit einer Sprechergruppe aus einem anderen Tal aufzunehmen. Die Gründe können vielseitig sein und liegen nicht nur an den zusätzlichen Wegkosten, welche Gebirge verursachen. So sind Gruppen, welche in einem Tal leben, welches sich auf Grund der ökologischen Bedingungen, (zum Beispiel einer langen «mean annual growing season») weniger auf Kontakt angewiesen, um zu überleben (Hua *et al.*, 2019).

Ein alternativer Ansatz zur Berechnung der Kosten zur Überquerung von Gebirgen wäre die «water catchment analysis», welche zur Modellierung von Wassereinzugsgebieten dient (Esri, 2022). Ein Wassereinzugsgebiet ist ein Gebiet, in dem das gesamte Wasser, einschliesslich Regen, Schneeschmelze und Grundwasser, aufgefangen sowie zu einem einzigen Punkt abgeleitet wird. Die Grenzen eines Wassereinzugsgebiets werden durch ein Netz von Bergkämmen (höchste Punkte) gebildet, die eine Wasserscheide darstellen. Das gesamte Wasser, das innerhalb der Grenzen des Wassereinzugsgebiets fällt, wird zum selben Abfluss geleitet, zum Beispiel zu einem Fluss oder See. Wassereinzugsgebiete bilden einen Lebensraum, bei dem verschiedene Pflanzen und Tiere koexistieren (Hill Country Alliance, 2022).

Wassereinzugsgebiete werden mit Hilfe eines DHM modelliert und eignen sich zur Darstellung von Bergkamm-Linien. Da Wasser die Grundlage für alles Leben bildet, könnten Grenzen von Wassereinzugsgebiete als scharfe Grenzen (ähnliche Methode wie bei den Flüssen, wo eine rasterisierte Linie passiert werden muss) für ein Gebiet angesehen werden, in welchem eine Sprechergruppe ohne Kontakt zu anderen Sprechergruppen überleben könnte und damit auch eine unabhängige Sprache entwickelt. Mit den Informationen über den Abfluss könnten beispielsweise Aufschlüsse über die «mean annual growing season» oder andere Faktoren, die laut (Hua *et al.*, 2019) eine Rolle bei der Expansion oder Diversität von Sprachen spielen berücksichtigt werden. So könnte der rein räumliche Ansatz der Hiking-Funktion durch eine zusätzliche sozioökonomische Komponente erweitert werden, wodurch aber auch der reine Fokus auf die Gebirge als natürlichen Grenzen verloren gehen würde.

Durchschnittliche Höhe

Neben der durchschnittlichen Höhe beschreiben (Hua *et al.*, 2019) weiter die Rauheit des Geländes sowie die maximale Höhendifferenz (höchster Punkt – tiefster Punkt) als Indikatoren für das Gelände. Zwar ist die durchschnittliche Höhe besser als die Höhendifferenz geeignet, da hier beispielsweise eine einzelne Randzelle einer Sprachfläche mit stark abweichendem Wert, das Resultat verzerren kann, allerdings sagt die durchschnittliche Höhe nichts über das Gelände aus. So kann es bezüglich der Hochebenen zu Verzerrungen kommen. Somit wäre die Rauheit des Geländes der sicherste Indikator gewesen.

11.4 FLÜSSE

Im Hinblick auf die Resultate kann die Hypothese 2a), dass Flüsse auf Grund ihrer erhöhten Kosten bei einer Überquerung einen signifikanten Einfluss auf den Verlauf der Sprachpolygone haben, nicht bestätigt werden. Auch lässt sich keine Verbesserung des Modells durch Gewichtung mit Flüssen speziell für Flüsse mit einer hohen Flussdichte feststellen.

11.4.1 Thematik

Die Resultate unterstützt damit weitestgehend die Aussage von (Hua *et al.*, 2019), dass Flüsse bei der Analyse der Ausbreitung von Sprache bzw. der Sprachdiversität eher als natürliche Ressource gesehen werden sollten, die es einer kleineren Sprechergruppe ermöglicht, ohne Kontakt zu anderen Sprechergruppen zu überleben (ökologische Risikohypothese), als eine natürliche Barriere, welche den Kontakt zwischen zwei Sprechergruppen erschwert (Isolationshypothese).

Die Annahme, dass Flüsse nicht zwingend als Hindernis gesehen werden sollten, lässt sich auch durch den Ansatz von (Kaiping, 2021) unterstützen. (Kaiping, 2021) geht in seinem Modell zur Analyse der Ausbreitung prähistorischer Kulturen von einer Geschwindigkeit von 3 km/h bei der Überquerung eines Gewässers aus. Bewegt man sich flussabwärts, wird die Fließgeschwindigkeit zu den 3 km/h addiert (Kaiping, 2021). Ein Fluss kann demzufolge als Transportmittel gesehen werden, welches den sozioökonomischen Austausch zwischen zwei Sprechergruppen fördern würde, was aber von (Hua *et al.*, 2019) in Bezug auf die Flussdichte nicht berücksichtigt wurde. Das lässt darauf schließen, dass die Modellannahme, Flüsse seien ein unüberwindbares Hindernis, grundsätzlich falsch ist. Die Annahmen gelten sowohl für die Datensätze Flüsse (gross) als auch Flüsse (mittel).

Analyse der Ausreisser

Obwohl die Verteilung von $JI(VT(\text{Fluss}(\text{gross}); \text{Ethnologue}))$ bzw. die Verteilung $JI(VT(\text{Fluss}(\text{gross}); \text{Ethnologue}))$ tiefere Medianwerte aufweist als die Verteilung bei $JI(\text{Fluss}(\text{gross}); \text{Ethnologue})$, finden sich bei den Flüssen Polygone mit hohem JI .

Ausreisser Polygone JI(Flüsse(mittel))



Abbildung 26: Karte vergleicht die räumliche Lage der Flüsse(mittel) mit den Polygonen der Ausreisser de Verteilung der JI(VT(Flüsse(mittel))).

Ausreisser Polygone JI(Flüsse(gross))



Abbildung 27: Karte vergleicht die räumliche Lage der Flüsse(gross) mit den Polygonen der Ausreisser de Verteilung der JI(VT(Flüsse(gross))).

Wie in den Abbildung 26 und Abbildung 27 gut zu erkennen ist, haben Polygone mit hohem JI(VT(Fluss);Ethnologue) keine Flüsse als Grenze. Ausnahme bildet in Abbildung 27 das Polygon im Norden Argentinens, dessen Grenze an der Grenze zu Paraguay verläuft, welche wiederum vom Rio Parana gebildet wird.

Die Geschichte der argentinisch-paraguayischen Grenze geht auf das späte 18. Jahrhundert zurück und diese war im Lauf der Jahre Schauplatz zahlreicher Konflikte. Ende des 19. Jahrhunderts versuchte die argentinische Regierung, das Gebiet der Provinz Misiones, das auf der paraguayischen Seite der Grenze liegt, zu annektieren. Dies führte zum Dreibundkrieg

zwischen 1865 und 1870, in dem Paraguay von den vereinten Kräften Argentiniens, Brasiliens und Uruguays besiegt wurde (Döhn, 1991; Reber, 1999).

Im Jahr 1932 brach der Chaco-Krieg aufgrund eines Streits über das Gebiet des Chaco aus. Dieser Konflikt wurde 1935 mit einem Abkommen beigelegt, in dem die heutige Grenze zwischen den beiden Ländern festgelegt wurde (Warren, 1961). Als Folge aller Konflikte resultiert hieraus die heute bestehende Grenze zwischen Argentinien und Paraguay. Das lässt darauf schliessen, dass die Sprachgrenze am Rio Parana aus politischen Gründen verläuft. Es ist gut möglich, dass die Sprache auf Grund von Vertreibung der Bevölkerung nur noch einseitig des Flusses gesprochen wird (Warren, 1961; Reber, 1999).

Ein ähnliches Beispiel findet man in Europa im Gebiet Elsass und Lothringen. Diese Gegend war auch häufig Schauplatz vieler Konflikte zwischen Deutschland und Frankreich (respektive deren Vorgängerstaaten), was dazu führte, dass abwechselnd Französisch und Deutsch gesprochen wurde – bis nach dem 2. Weltkrieg endgültig der Rhein als Trennlinie zwischen den Staaten festgelegt wurde, was dazu führte, dass heute in der Region Französisch gesprochen wird (Suckow, 2012).

Folglich trifft auch für die Ausreisser-Polygone die Isolationshypothese (Hua *et al.*, 2019) in Bezug auf Flüsse nicht zu.

11.4.2 Methodik

Gewichtung

Bei der Gewichtung wurde jeder Fluss als unüberwindliches Hindernis und somit als Grenze, welche Sprache nicht passieren kann, betrachtet. Diese Annahme ist nicht korrekt, weil davon ausgegangen wird, dass die ökologische Risikohypothese von (Hua *et al.*, 2019) als grundsätzlich falsch angesehen wird. Vielmehr sollten die zusätzlichen Auswirkungen der Isolationshypothesen genauer betrachtet werden. Wird jedoch – wie in diesem Fall – jeder Fluss als Grenze betrachtet, führt dies zu einem Ausschluss der ökologischen Risikohypothese im Modell. Alternativ hätten beispielsweise nur Flüsse mit besonders wenig Nährstoffen als natürliche Grenze berücksichtigt werden dürfen. Dies war auf Grund der vorhandenen Informationen des Datensatzes nur schwer zu realisieren.

Korrelation mit der Flussdichte

Der Vergleich mit der Flussdichte und der Diff(Flüsse) muss besonders kritisch gesehen werden – vor allem in Anbetracht, dass, wie in Abbildung 24 zu sehen ist, die Polygone mit der höchsten Diff(Flüsse) eine Flussdichte von 0 haben. Dies lässt sich dadurch erklären, dass nicht die Flussdichte der Fläche des mit VT generiertem Polygon dargestellt wird, sondern die Flussdichte des entsprechenden Polygons von Ethnologue. Die Idee dahinter war, Selbstkorrelationseffekte zu verhindern (das mit VT Erstellte ist, wie die Flussdichte, stark von den Flüssen abhängig, die Polygone von Ethnologue sind hingegen unabhängig vom Algorithmus). Die dadurch entstandene Verzerrung senkt die Aussagekraft auf ein Minimum.

Flussdichte

Die Flussdichte muss auch kritisch betrachtet werden. Einerseits gibt GloRic (Ouellet Dallaire *et al.*, 2019; *GloRic Global River Classification Technical Documentation Version 1.0*, 2018) keine genauen Breiten von Flüssen an. Hier wurde nur der Abfluss berücksichtigt. So wurde einfachheitshalber zur Berechnung der Flussdichte die Annahme getroffen, dass alle Flüsse gleich gross sind, was nicht der Realität entspricht. Ausserdem war die Flussdichte von der Auflösung abhängig. So mussten auf Grund der 8er NN – Algorithmus jeder Fluss mindestens zwei Zellen breit sein, damit eine als Fluss kategorisierte Zelle nicht passiert werden konnte. Zwei Zellen entsprechen etwa 2 km, wobei die Länge des grössten Flusses Südamerikas, des Amazonas, mit 1 – 1.6 km angegeben wird (Rieck and Uhlenbrock, 2014).

Weiter wurden für die Analyse der VT(Fluss(gross)) und der VT(Fluss(mittel)) unterschiedliche Flussdichten verwendet, da bei der VT(Fluss(mittel)) auch mehrere Flüsse berücksichtigt wurden. Das zeigt sich ebenso in Abbildung 24 wo die Polygone VT(Fluss(mittel)) grössere Flussdichten aufweisen als VT(Fluss(gross)).

Dieser Fakt führt zu weiteren Verzerrungen, was in diesem Fall auf Grund der minimalen Aussagekraft jedoch kaum eine Rolle spielt.

11.5 EVALUIERUNG

Wegen den allgemeinen tiefen JI-Werten in den Abbildung 12, Abbildung 19 und Abbildung 20 muss die Evaluierungsmethode kritisch betrachtet werden.

Der JI wurde ursprünglich dazu verwendet, gemeinsame Merkmale von zwei verschiedenen Pflanzenarten zu erforschen. Ein räumlicher Kontext war nicht Teil der Berechnung. Es wurden simpel zwei Mengen auf gemeinsame Elemente untersucht (Jaccard, 1901; Da Fontoura Costa, 2021). Der JI wurde zwar schon angewendet, um zwei Raster danach zu untersuchen, ob deren Zellen an derselben georeferenzierten Position demselben Attribut zugewiesen wurden (Kopuchian *et al.*, 2020), aber die Methode an sich wurde nicht analysiert und diskutiert. So ist es schwierig, zu sagen, was genau ein guter JI-Wert für räumliche Analysen ist. Allgemein scheint sich die Literatur nicht auf einen «guten» JI zu fixieren, was darauf schliessen lässt, dass dieser stark von der Thematik abhängt (Jaccard, 1901; Kopuchian *et al.*, 2020; Da Fontoura Costa, 2021). Der JI scheint eher für einen Überblick geeignet als für eine klare statistische Aussage. Trotzdem lassen sich Resultate mit dem JI vergleichen. So sieht man bei Trinidad und Tobago einen der höchsten JI-Werte, was auf Grund der Insellage Sinn ergibt.

Weiter sagt der JI nur aus, wie gross das Verhältnis von $A \cap B$ zu $A \cup B$ ist. Er gibt keinen Aufschluss darauf, wie sich $(A \text{ not } B)$ bzw. $(B \text{ not } A)$ zur Gesamtmenge verhalten. Dazu wären weitere Koeffizienten nötig, wie die Präzision $(A \cap B / ((A \cap B) + (A \text{ not } B)))$ (Data Mines, 2023), um beispielsweise zu analysieren, ob die mit VT erstellten Flächen in den meisten Fällen innerhalb der entsprechenden Polygone von Ethnologue liegen, dies aber nicht umgekehrt der Fall ist. Das ist nur ein Beispiel von Evaluierungen, welche zum Vergleich von zwei Mengen zusätzlich durchgeführt werden können (Data Mines, 2023).

12 FAZIT

12.1 ERREICHTES

Als Erstes sollte diese Arbeit untersuchen, wie gut sich die Voronoi-Tesselation (VT) zur Modellierung von Spracharealen eignet. Dabei stellte sich diese als eher ungeeignet heraus, weil bei der VT von einer kreisförmigen Ausbreitung ausgegangen wird, was sich in Bezug auf die Expansion einer Sprache nicht sagen lässt, da in der Realität zu viele verschiedene Faktoren eine Rolle spielen, womit eine gleichförmige Ausbreitung von Sprachen unwahrscheinlich ist. Die Wahl des Seedpoints spielt demzufolge auch keine Rolle.

Weiter wurde untersucht, ob die Gewichtung der VT mit natürlichen Barrieren (Gebirge und Flüsse) bessere Resultate liefert als einfache VT. Die Resultate zeigten eine minimale Verbesserung bei der Gewichtung mit Gebirgen. Bei der Gewichtung mit Flüssen versagte das Modell hingegen komplett.

Mit den bisherigen Resultaten wurde im Anschluss analysiert, wie stark eine natürliche Barriere als Barriere zwischen Sprechergruppen gesehen werden kann. Dabei zeigte sich, dass Gebirge einen Einfluss auf die Spaltung haben können, es aber viele sozioökonomische Faktoren gibt, welche ebenfalls Einfluss haben. Hingegen sind Flüsse keine Sprachbarrieren. Im Gegenteil – sie können sogar für den Austausch zweier Populationen/Sprechergruppen fördernd sein.

12.2 GRENZEN

Nebst den in der Diskussion behandelten Limitationen, gab es weitere Punkte, wo die Methodik an ihre Grenzen stieß.

So wurde keine Rücksicht auf die phylogenetische Verwandtschaft von Sprachen genommen. Jede Sprache wurde als unabhängig von anderen Sprachen betrachtet. So konnten Cluster von Sprachen, die miteinander verwandt sind, nicht untersucht werden. Auch Effekte des «Borrowing» konnten auf Grund der fehlenden linguistischen Informationen, wie zum Beispiel vorhandenen Lehnwörter, nicht berücksichtigt werden.

Auf Grund der langen Berechnungszeiten mussten die Outputs der Kostenfunktion –Kapitel 9.2– auf einen bestimmten Ausschnitt begrenzt werden, wodurch es zu «leeren» Räumen, die keiner Sprache zugeordnet werden konnten, gekommen ist .

Weiter war die Wahl der Sprachen durch die vorhandenen Datensätze begrenzt. So konnten nur Sprachen, welche sowohl in Ethnologue als auch in Glottolog vorkommen, evaluiert und analysiert werden, wodurch viel vorhandenes Material verloren ging. Zudem beinhaltet der Datensatz von Glottolog für Südamerika insgesamt 668 verschiedene Sprachen mit entsprechenden Geometrien, derjenige für Ethnologue nur 453. Da für die Mincostmap auch Sprachen relevant waren, die nur in einem der beiden Datensätzen waren und bei der Evaluierung nicht berücksichtigt wurden, hatte dies zur Folge, dass die Sprachareale, für welche die Glottologpunkte als SP verwendet wurden, im Schnitt etwas kleiner waren als diejenigen, welche die Ethnologuezentroid als SP einsetzten.

12.3 ZUKUNFTSAUSSICHTEN

Interessant wäre es nun, anhand der gefundenen Resultate das Modell zu erweitern, um bessere Evaluierungsergebnisse für die einzelnen Polygone zu erhalten, wobei auch eine Alternative zum JI in Betracht gezogen werden kann. Die grössten Anpassungen könnten in der Neuinterpretation der Flüsse sowie der zusätzlichen Berücksichtigung von sozioökonomischen Parametern, zum Beispiel der Bevölkerungsgrösse, umgesetzt werden. Überdies könnten die Erkenntnisse für Modelle im Bereich der phylogenetischen Expansion von Sprache genutzt werden.

-

13 BIBLIOGRAFIE

- Alberti, G. (2019) ‘movecost: An R package for calculating accumulated slope-dependent anisotropic cost-surfaces and least-cost paths’, *SoftwareX*. Elsevier B.V. doi: 10.1016/j.softx.2019.100331.
- Axelsen, J. B. and Manrubia, S. (2014) ‘River density and landscape roughness are universal determinants of linguistic diversity’, *Proceedings of the Royal Society B: Biological Sciences*. Royal Society, 281(1784). doi: 10.1098/rspb.2013.3029.
- BAFU, B. für U. (2023) *Aktuelle Daten der Nationalen Daueruntersuchung der Fließgewässer (NADUF)*. Available at: <https://www.bafu.admin.ch/bafu/de/home/themen/wasser/zustand/daten/aktuelle-messwerte-der-gewaesser/aktuelle-daten-der-nationalen-daueruntersuchung-der-fluessgewaes.html>.
- Bhunia, G. S., Shit, P. K. and Maiti, R. (2018) ‘Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC)’, *Journal of the Saudi Society of Agricultural Sciences*, 17(2), pp. 114–126. doi: <https://doi.org/10.1016/j.jssas.2016.02.001>.
- Bouckaert, R. *et al.* (2012) ‘Mapping the origins and expansion of the Indo-European language family.’, *Science (New York, N.Y.)*. United States, 337(6097), pp. 957–960. doi: 10.1126/science.1219669.
- Bouckaert, R. R., Bowern, C. and Atkinson, Q. D. (2018) ‘The origin and expansion of Pama-Nyungan languages across Australia’, *Nature Ecology and Evolution*. Springer US, 2(4), pp. 741–749. doi: 10.1038/s41559-018-0489-3.
- Cashdan, E. (2001) ‘Ethnic Diversity and Its Environmental Determinants: Effects of Climate, Pathogens, and Habitat Diversity’, *American Anthropologist*, 103(4), pp. 968–991. doi: 10.1525/aa.2001.103.4.968.
- Collard, I. F. and Foley, R. A. (2002) ‘Latitudinal patterns and environmental determinants of recent human cultural diversity: do humans follow biogeographical rules?’, *Evolutionary Ecology Research*, 4, pp. 371–383.
- Currie, T. E. and Mace, R. (2009) ‘Political complexity predicts the spread of ethnolinguistic groups’, *Proceedings of the National Academy of Sciences*, 106(18), pp. 7339–7344. doi: 10.1073/pnas.0804698106.

Data Mines (2023) *Was ist eine Konfusionsmatrix? Leicht erklärt* | *DataMines*. Available at: <https://datamines.de/konfusionsmatrix/> (Accessed: 14 February 2023).

Döhn, L. (1991) *Domnick, Heinz Joachim: Der Krieg der Tripel-Allianz in der deutschen Historiographie und Publizistik, MEDIENwissenschaft: Rezensionen* | *Reviews*. doi: 10.17192/EP1991.2.5417.

Eberhard, D. M., Simons, G. and Fenning, C. (2022) *Ethnologue: Languages of the World. Twenty-fifth edition, Dallas, Texas: SIL International*. Available at: https://www.ethnologue.com/?ip_login_no_cache=B%2A%0E%0A%AEP%15%E0&cache= (Accessed: 24 January 2023).

Esri (2022) *Funktionsweise von 'Abflussgebiet'—ArcGIS Pro* | *Dokumentation*. Available at: <https://pro.arcgis.com/de/pro-app/latest/tool-reference/spatial-analyst/how-watershed-works.htm> (Accessed: 16 February 2023).

Da Fontoura Costa, L. (2021) *Further Generalizations of the Jaccard Index*.

Gabor, P. (2023) *Wann gilt ein Dialekt als eigenständige Sprache? - SWR Wissen*. Available at: <https://www.swr.de/wissen/1000-antworten/wann-gilt-ein-dialekt-als-eigenstaendige-sprache-100.html> (Accessed: 24 February 2023).

Gavin, M. C. *et al.* (2013) 'Toward a mechanistic understanding of linguistic diversity', *BioScience*, 63(7), pp. 524–535. doi: 10.1525/bio.2013.63.7.6.

Gavin, M. and Stepp, J. R. (2014) 'Rapoport's Rule Revisited: Geographical Distributions of Human Languages', *PLOS ONE*. Public Library of Science, 9(9), pp. 1–8. doi: 10.1371/journal.pone.0107623.

Geographic Information Technology Training Alliance (2016) *Thiessen-Polygone*. Available at: http://www.gitta.info/Accessibilit/de/html/UncProxAnaly_learningObject4.html (Accessed: 24 January 2023).

Germani, G. and Knapp, G. W. (2023) 'South America - Linguistic patterns | Britannica', *Britannica*. Available at: <https://www.britannica.com/place/South-America/Linguistic-patterns> (Accessed: 23 February 2023).

GISGeography (2022) *What is a Voronoi Diagram? - GIS Geography*. Available at: <https://gisgeography.com/voronoi-diagram-thiessen-polygons/> (Accessed: 30 December 2022).

Glen, S. (2022) *Jaccard Index / Similarity Coefficient - Statistics How To*. Available at:

<https://www.statisticshowto.com/jaccard-index/> (Accessed: 23 November 2022).

GloRiC Global River Classification Technical Documentation Version 1.0 (2018). Available at: <http://www.hydrosheds.org>. (Accessed: 10 November 2022).

Goodchild, M. F. (2020) ‘Beyond Tobler’s Hiking Function’, *Geographical Analysis*, 52(4), pp. 558–569. doi: 10.1111/gean.12253.

Greenhill, S. (2014) ‘Demographic Correlates of Language Diversity’, in Bower, Claire & Evans, B. (ed.) *The Routledge handbook of historical linguistics*. London, pp. 555–578.

Hammarström, H. *et al.* (2021) *Glottolog 4.5 -*, Max Planck Institute for Evolutionary Anthropology. doi: 10.5281/zenodo.5772642.

Hastings, D. A. and Dunbar, P. K. (1999) ‘Global Land One-kilometer Base Elevation digital elevation model (Globe), Documentation’, *Key to Geophysical Records Documentation*, (34), p. 147.

Hill Country Alliance (2022) *Water Catchment Areas (Watersheds)*. Available at: <https://hillcountryalliance.org/our-work/water-resources/water-catchment-areas-watersheds/> (Accessed: 16 February 2023).

Hoyer, S. *et al.* (2016) ‘xarray: v0.8.0’. doi: 10.5281/zenodo.59499.

Hua, X. *et al.* (2019) ‘The ecological drivers of variation in global language diversity’, *Nature Communications*. Nature Publishing Group, 10(1), pp. 1–10. doi: 10.1038/s41467-019-09842-2.

Isbell, W. H. (2008) ‘Wari and Tiwanaku: International Identities in the Central Andean Middle Horizon’, in Silverman, H. and Isbell, W. H. (eds) *The Handbook of South American Archaeology*. New York, NY: Springer New York, pp. 731–759. doi: 10.1007/978-0-387-74907-5_37.

Jaccard, P. (1901) ‘Distribution de la Flore Alpine dans le Bassin des Dranses et dans quelques régions voisines.’, *Bulletin de la Societe Vaudoise des Sciences Naturelles*, 37, pp. 241–272. doi: 10.5169/seals-266440.

Jordahl, K. *et al.* (2020) ‘geopandas: v0.8.1’. Zenodo. doi: 10.5281/zenodo.3946761.

Kaiping, G. A. (2021) *A network for simulating pre-colonial migration in the Americas*. doi: 10.25436/E21598.

Koile, E. *et al.* (2022) ‘Phylogeographic analysis of the Bantu language expansion supports a

rainforest route’, *Proceedings of the National Academy of Sciences*, 119(32), p. e2112853119. doi: 10.1073/pnas.2112853119.

Kopuchian, C. *et al.* (2020) ‘A test of the riverine barrier hypothesis in the largest subtropical river basin in the Neotropics’, *Molecular Ecology*, 29(12), pp. 2137–2149. doi: 10.1111/mec.15384.

Kröner, K. (2023) *Rasterdaten im Vergleich mit Vektordaten | GIS | Repetico, Repetico*. Available at: <https://www.repetico.de/card-16809387> (Accessed: 10 February 2023).

Landesamt für Geoinformation und Landentwicklung Niedersachsen (2021) *Geodaten Technisches Basiswissen*. Available at: https://www.geodaten.niedersachsen.de/download/57102/Geodaten_-_Technisches_Basiswissen.pdf.

Lehner, B. and Grill, G. (2013) ‘Global river hydrography and network routing: baseline data and new approaches to study the world’s large river systems’, *Hydrological Processes*, 27(15), pp. 2171–2186. doi: <https://doi.org/10.1002/hyp.9740>.

Leipold, H. *et al.* (2017) ‘Erratum: ‘Statistical topology of perturbed two-dimensional lattices (2016 J. Stat. Mech. P043103)’’, *Journal of Statistical Mechanics: Theory and Experiment*, 2017, p. 79901. doi: 10.1088/1742-5468/aa78af.

Linke, S. *et al.* (2019) ‘Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution’, *Scientific Data*, 6(1), pp. 1–15. doi: 10.1038/s41597-019-0300-6.

Liu, D. *et al.* (2019) ‘Human-induced eutrophication dominates the bio-optical compositions of suspended particles in shallow lakes: Implications for remote sensing’, *Science of the Total Environment*, 667, pp. 112–123. doi: 10.1016/j.scitotenv.2019.02.366.

Loy, T. (2015) *Wie werden die Koeffizienten in der linearen Regression interpretiert? - Statistik und Beratung - Daniela Keller, Statistik + Beratung*. Available at: [https://statistik-und-beratung.de/2015/06/wie-werden-die-koeffizienten-in-der-linearen-regression-interpretiert/#:~:text=Einfache lineare Regression,und b %3D Steigung der Geraden](https://statistik-und-beratung.de/2015/06/wie-werden-die-koeffizienten-in-der-linearen-regression-interpretiert/#:~:text=Einfache lineare Regression,und b %3D Steigung der Geraden.). (Accessed: 23 February 2023).

Magyari-Saska, Z., Magyari-Sáska, Z. and Dombay, Ş. (2012) *Determining Minimum Hiking Time Using Dem*.

Mantiri, O. (2015) ‘Factors Affecting Language Change’, *SSRN Electronic Journal*. Elsevier BV. doi: 10.2139/ssrn.2566128.

- Meng, Q., Liu, Z. and Borders, B. E. (2013) ‘Assessment of regression kriging for spatial interpolation – comparisons of seven GIS interpolation methods’, *Cartography and Geographic Information Science*. Taylor & Francis, 40(1), pp. 28–39. doi: 10.1080/15230406.2013.762138.
- Mu, L. (2004) ‘Polygon Characterization With the Multiplicatively Weighted Voronoi Diagram’, *The Professional Geographer*. Routledge, 56(2), pp. 223–239. doi: 10.1111/j.0033-0124.2004.05602007.x.
- Neureiter, N. *et al.* (2021) ‘Can Bayesian phylogeography reconstruct migrations and expansions in linguistic evolution?’, *Royal Society Open Science*, 8(1). doi: 10.1098/rsos.201079.
- Nichols, J. (1997) ‘Modeling Ancient Population Structures and Movement in Linguistics’, *Annual Review of Anthropology*, 26(1), pp. 359–384. doi: 10.1146/annurev.anthro.26.1.359.
- Ouellet Dallaire, C. *et al.* (2019) ‘A multidisciplinary framework to derive global river reach classifications at high spatial resolution’, *Environmental Research Letters*. IOP Publishing, 14(2). doi: 10.1088/1748-9326/aad8e9.
- Perrigo, A., Hoorn, C. and Antonelli, A. (2020) ‘Why mountains matter for biodiversity’, *Journal of Biogeography*. Blackwell Publishing Ltd, pp. 315–325. doi: 10.1111/jbi.13731.
- Picout, S. (2013) *Sprache oder Dialekt - Definitionen und Abgrenzungsversuche (kartoniertes Buch) | Buchhandlung Bestenbostel*. 1. Auflage 2012. GRIN Verlag. Available at: <https://bestenbostel.com/shop/item/9783656128236/sprache-oder-dialekt-definitionen-und-abgrenzungsversuche-von-sabine-picout-kartoniertes-buch#> (Accessed: 22 February 2023).
- Reagan, P. T. (2003) ‘Peace, Conflict, and War: The Role of Language and Languages’, pp. 19–34.
- Reber, V. B. (1999) ‘A case of total war: Paraguay, 1864–1870’, *Journal of Iberian and Latin American Research*. Routledge, 5(1), pp. 15–40. doi: 10.1080/13260219.1999.10429960.
- Rieck, S. and Uhlenbrock, K. (2014) *Infoblatt Amazonas*. Available at: https://www.klett.de/sixcms/detail.php?template=terrasse_artikel__layout__pdf&art_id=1017939 (Accessed: 17 February 2023).
- Robert Henry Robins (2022) ‘language - Language and social differentiation and assimilation | Britannica’, *Britannica*. Available at: <https://www.britannica.com/topic/language/Language-and-social-differentiation-and-assimilation> (Accessed: 30 December 2022).

- Sgobba, T. (2019) 'B-737 MAX and the crash of the regulatory system', *Journal of Space Safety Engineering*. Elsevier Ltd, 6(4), pp. 299–303. doi: 10.1016/j.jsse.2019.09.006.
- Shi, R., Ngan, K. N. and Li, S. (2014) 'Jaccard index compensation for object segmentation evaluation', in *2014 IEEE International Conference on Image Processing, ICIP 2014*. Institute of Electrical and Electronics Engineers Inc., pp. 4457–4461. doi: 10.1109/ICIP.2014.7025904.
- Smith, B. and Varzi, A. C. (2000) 'Fiat and Bona Fide Boundaries', *Philosophy and Phenomenological Research*, 60(2), p. 401. doi: 10.2307/2653492.
- Sonneck, P. (2020) *Das Bestimmtheitsmaß der linearen Regression | IfaD*. Available at: <https://www.ifad.de/das-bestimmtheitsmass-der-linearen-regression/> (Accessed: 23 February 2023).
- Stepp, J. R., Castaneda, H. and Cervone, S. (2005) 'Mountains and biocultural diversity', *Mountain Research and Development*, 25(3), pp. 223–227. doi: 10.1659/0276-4741(2005)025[0223:MABD]2.0.CO;2.
- Stoeckle, P. (2014) 'Linguistic Geography and GIS', (August).
- Suckow, D. (2012) *Deutscher Rhein, französischer Rhein | bpb.de*. Available at: <https://www.bpb.de/themen/europaeische-geschichte/geschichte-im-fluss/135684/deutscher-rhein-franzoesischer-rhein/> (Accessed: 30 December 2022).
- Teichert, N. (2021) *Über - Info-Peru*. Available at: <https://info-peru.de/ueber-info-peru/> (Accessed: 15 February 2023).
- Theory of knowledge (2023) *Knowledge and language - TOK*. Available at: <https://tok2022.com/knowledge-and-language/> (Accessed: 22 February 2023).
- Tobler, W. (1993) *Three Presentations on Geographical, National Center for Geographic Information and Analysis*.
- Tulio, M. *et al.* (2021) 'Demographic shifts , inter-group contact , and environmental conditions drive language extinction and diversification'.
- Túlio, M. *et al.* (2019) 'Drivers of geographical patterns of North American language diversity', *Royal Society*, 286(1899). doi: 10.1098/rspb.2019.0242.
- Urban, M. (2021) 'Linguistic and cultural divisions in pre-Hispanic Northern Peru', *Language Sciences*. Elsevier Ltd, 85, p. 101354. doi: 10.1016/j.langsci.2020.101354.

USGS (2023) *What is a digital elevation model (DEM)?* | *U.S. Geological Survey*. Available at: <https://www.usgs.gov/faqs/what-digital-elevation-model-dem> (Accessed: 9 February 2023).

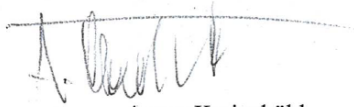
Vogt, L. *et al.* (2012) 'Fiat or Bona Fide Boundary-A Matter of Granular Perspective', *PLoS ONE*, 7(12). doi: 10.1371/journal.pone.0048603.

Warren, H. G. (1961) 'The Conduct of the Chaco War', *Hispanic American Historical Review*, 41(3), pp. 453–454. doi: 10.1215/00182168-41.3.453.

Wichmann, S., Müller, A. and Velupillai, V. (2010) 'Homelands of the world's language families', *Diachronica*, 27(2), pp. 247–276. doi: 10.1075/dia.27.2.05wic.

Persönliche Erklärung

Ich erkläre hiermit, dass die vorliegende Arbeit das Ergebnis meiner eigenen, unabhängigen Arbeit ist.
Alle externen Quellen sind in der Arbeit ausdrücklich angegeben.



Aaron Kreienbühl

Zürich, 25.02.2023