An application of spatial population genetics methods to detect clines in Swiss German morphosyntax

Master Project

Spring 2018

Noemi Romano
Department of Environmental Sciences and Engineering
École Polythechnique Fédérale de Lausanne

Supervisors

Dr. Stéphane Joost
Laboratory of Geographic Information Systems (LASIG)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Dr. Peter Ranacher
Geographic Information Science, Department of Geography

Dr. Curdin Derungs
URPP Language and Space

University of Zurich[UZH]

Abstract

Delineating dialectal boundaries is one of the main challenges in dialectometry, and two main approaches have been proposed to define the spatial organization of dialects: *areas*, in which dialects are delineated by discrete boundaries, and *continua* in which they gradually vary from one region to the next. The transfer of linguistic features between dialects is one of the primary processes occurring when dialects are in contact. This entails a perpetual inter-dialectal influence, leading to the formation of continua. Swiss German dialects are geographically in contact, and it is then more appropriate to address the modeling of dialects continua rather than dialectal areas. The paradigm of dialect continuum finds its parallel in biology with the phenomenon of *geographical cline*, namely a gradual evolution of a biological character along with a geographic gradient. Population genetics methods are the mathematical framework quantifying these geographical clines; more specifically, Bayesian clustering algorithms such as STRUCTURE and TESS have been widely used to infer population structure and detect geographical clines from genetic data. These methods estimate the proportions of each individual's genome (*admixture proportions*) that belong to one or more unknown populations, which are jointly inferred in a Bayesian process. Despite the importance of geography in dialectometry, spatial population genetics methods have not been applied up until now over linguistic data. Hence, in this study, the spatial Bayesian clustering algorithm TESS has been applied by using 88 morphosyntactic phenomena sampled in 383 Swiss German municipalities. The particularity of this algorithms is to incorporate a spatial prior distribution on the individual admixture proportions. The results show that geography has a role in explaining the Swiss German morphosyntactic variation, although the spatial trend of morphosyntax has generally been considered less salient than other linguistic levels. Furthermore, a diversity analysis has been performed and shows that some areas are highly diverse, in contrast to other zones that present clear membership to a specific population. Moreover, geographical clines have been detected. Finally, it has been shown that morphosyntactic traits can be considered as good historical units since migration patterns have been detected.

Key-words **dialect continuum, spatial population genetics, geographical clines, morphosyntax, Bayesian clustering**

Résumé

La délinéation des frontières dialectales est un des plus importants défis en dialectométrie et deux approches ont été proposées afin de définir l'organisation spatiale des dialectes : des aires, dans lesquels les dialectes sont délimités par des frontières discrètes, et des continua, dans lesquels ils varient graduellement d'une région à l'autre. Les dialectes en contact s'influencent constamment ce qui résulte en une forme de continuum dialectal. Cette étude s'intéressant aux dialectes suisses allemands, il s'avère plus pertinent de modéliser leur structure spatiale avec un continuum dialectal plutôt ques des aires, au vu de leur proximité géographique. Le paradigme de continuum dialectal trouve son parallèle en biologie avec le phénomène de cline géographique, qui correspond à l'évolution graduelle d'un caractère biologique le long d'un gradient géographique. Les méthodes de génétique des populations sont le cadre mathématique qui permet de quantifier ces clines géographiques. Plus spécifiquement, les méthodes de classification Bayésiennes Structure et Tess ont largement été appliquées dans le but d'inférer la structure des populations et de détecter les clines à partir de données génétiques. Ces méthodes permettent d'estimer les proportions du génome d'un individu (coefficients de métissage) qui proviennent des plusieurs populations ancêtres et qui sont inférées conjointement dans un processus Bayésien. Malgré l'importance de la géographie en dialectométrie, des méthodes de génétique des populations spatiale n'ont pas encore été appliquées jusqu'à présent sur des données linguistiques. Ainsi, dans cette étude, la méthode Bayésienne de classification Tess a été appliquée sur 88 phénomènes de morphosyntaxe échantillonnés dans 383 municipalités suisse allemandes. La particularité de cet algorithme est d'intégrer de l'information spatiale à priori lors de l'estimation des coefficients de métissage. Les résultats montrent que la composante spatiale a un rôle important dans l'explication de la variation de la morphosyntaxe suisse allemande, malgré sa répartition spatiale soit généralement considérées moins saillante que d'autres niveaux linguistiques. A partir des coefficients de métissage, des analyses de diversité ont été effectuées, et des zones spécifiques montrent une diversité dialectale accrue, alors que d'autres présentent une claire appartenance à une seule famille dialectale. De plus, des clines géographiques ont été détectées dans le nord du pays. In fine, les caractéristiques de morphosynaxe suisses allemandes peuvent être considérée comme des unités héritables, puisque des traces migratoires historiques ont été discernées.

Mots-clés **continuum dialectal, génétique des populations, clines, morphosyntaxe, classification Bayésienne**

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisors Dr. Stéphane Joost, Dr. Peter Ranacher and Dr. Curdin Derungs for their continuous support during my master thesis. It was an incredible experience from which I learned a lot.

Deuxièmement j'aimerais remercier toutes les personnes qui ont fait partie de mon long parcours en SIE, à Sat et au Maupas.

E per ultimo ma non meno importante, vorrei ringraziare i miei genitori e la mia famiglia che mi hanno sostenuta tutti questi anni. Questo è per voi e grazie a voi.

Contents

List of Tables

List of Figures

Acronyms

CAR Conditional Auto Regressive.

CI Credible intervals.

DIC Deviance Information Criterion.

FA Factor Analysis.

H Shannon diversity index.

HAC Hierarchical Agglomerative Clustering.

HWE Hardy-Weinberg equilibrium.

LD Levenshtein Distance.

LE Linkage equilibrium.

MCMC Markov chain Monte Carlo.

MDS Multidimensional Scaling.

MH Metropolis-Hastings.

PCA Principal Component Analysis.

PCM Parametric Comparison Method.

SADS Syntaktischer Atlas der deutschen Schweiz.

SDS Sprachatlas der deutschen Schweiz.

WIV Weighted Identity Value.

Chapter 1

Introduction

1.1 Context

Around 7'000 languages exist in the present-day world, and 90% of these languages are spoken by less than 100'000 people [27]. Over time, many of these languages will disappear due to the extinction of a cultural group or the shift to another language within a speech community. Therefore, languages by evolving or disappearing are closely linked to the evolutionary process of human cultural communities, and their change could give insights on historical human migrations. One example is the English language which contains a large number of French origin words, due to the Norman invasion of 1066 in British island, where French was imposed as the language of culture and administration [98]. Another striking example is the Basque population, which is highly isolated from a linguistic and genetic point of view. The Basque population was not subjected to foreign migrations such as the other European communities, which confirms their linguistic and genetic isolation [11]. Many other examples exist, and the biological and linguistic evolution should be encompassed in an interdisciplinary context since similar processes occur in both disciplines.

It is crucial to distinguish languages from dialects since a language is defined as a collection of dialects specific to a geographic location. Therefore, while languages are investigated at a macro-geographic scale, the study of dialects allows us to look into micro-geographic areas where speakers can understand each other without significant efforts (*mutual intelligibility*). The constant interference between mutual-intelligible dialects causes them to change over time and space constantly. Dialectology is the scientific branch investigating their temporal and spatial evolution and aims to deduce dialects similarities along with migration patterns in specific geographic locations.

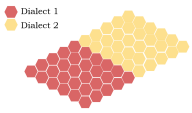
In the last century, thanks to technology progress and the availability of complex linguistic atlases, numerous computational techniques have emerged to investigate the spatial variation of dialects. The branch of dialectology dealing with quantitative and computational tech-

niques is known as dialectometry [67]. It has been introduced by Séguy in 1973 [84], who computed linguistic distances for specific linguistic levels between different locations. Séguy was supported right after by Goebl [39] and the primary aim of their further research was to study dialectal varieties in an aggregate framework by summing the linguistic differences or similarities between distinct locations.

However, standard dialectometric techniques do not give insights into the temporal and historical evolution of dialectal variation. Dialectometric analysis are mostly based on synchronic approaches [106], namely the investigation of a language or dialect at a specific point in time, usually when the data have been collected. Evolutionary biology aims to examine the temporal evolution of species, and the application of these methods in dialectometry could be an interesting interdisciplinary approach to perform diachronic analysis, namely the study of dialectal variation in space and time. Hence, this study aims to apply methods from evolutionary biology to investigate the spatial structure of Swiss German dialects and understand if the linguistic information can be used to detect Swiss German historical migrations.

1.2 Problem statement

One of the main challenge in dialectometry is how dialectal spatial organization can be modeled: *areas*, in which dialects are delineated by discrete boundaries [47] and *continua* in which they gradually vary from one region to the next [12] (Figure 1.1). In the context of dialects in contact, the transfer of linguistic features is an important process implying a constant inter-dialectal influence, leading to the formation of continua [101]. Swiss German dialects could, therefore, be encompassed in the context of continua, due to their continuous spatial organization.



(a) Dialectal areas

(b) Dialect continuum

Figure 1.1: Spatial organization of dialects

Contact between dialects and the inter-influence between them can be paralleled in biology with the process of *gene flow*, occurring when two distinct populations come into contact and gene information is transferred from one population to another [54]. Gene flow between two distinct populations gives rise to an *admixed* gene pool, which has relative contributions from each of the parental populations. The contributions of parental populations to an individual are called *admixture proportions* in biology. Moreover, dialect continuum finds its parallel in biology as well with the term of *geographical cline* [1, 19], which is a gradual evolution of a biological character along a geographic gradient [45].

In evolutionary biology, population genetics methods are the mathematical framework quantifying the genetic admixture of individuals and their geographical clines; more specifically,

spatial Bayesian clustering algorithms such as TESS [32] has been widely used to infer population structure from genetic data. This method estimates the admixture proportions of each individual's genome that belong to one or more unknown populations with the use of spatial priors in a Bayesian process. This study aims to apply TESS to Swiss German linguistic data to simultaneously infer hypothetical ancestral dialectal families and municipality admixture proportions belonging to each of these ancestral families. Despite the importance of geography in dialectometry, TESS has not been performed to dialect data up until now.

1.3 Goal and objectives

The goal of this project is to detect if Swiss German morphosyntactic ancestral families show a spatial structure. To address this goal, several objectives will be pursued in this study:

- Represent and quantify the spatial variation of Swiss German dialectal families
- Compute the dialectal diversity for each municipality and identify the most diverse or homogeneous areas
- Identify geographical clines of Swiss German in order to understand whether Swiss German can be encompassed in a dialect continuum context
- Identify if the inferred ancestral dialectal families show similar patterns of historical migrations

1.4 Research questions

What is the total number of ancestral populations at the origin of Swiss German dialects?

The maximum number of ancestral dialectal families at the origin of Swiss German dialects can be done using a statistical criterion obtained during the simulations.

Is geography an important factor in explaining Swiss German morphosyntactic variation?

This concerns the selection of the model which explains the data statistically the best. Hence, the question if the integration of geography in the model improves model performances is addressed.

How intra-population admixture proportions vary across space?

This question addresses the investigation of morphosyntactic variation within every ancestral dialectal family. This can be done by evaluating whether municipality admixture proportions gradually vary from one point to another or whether discrete changes characterize their spatial variation.

Are inter-population transitions characterized by discrete or gradual changes?

Here, the investigation of morphosyntactic variation is focused on the transitions zones between the different ancestral populations inferred. In other words, attention is paid to zones where ancestral populations come into contact.

1.5 Methodology

In order to apply methods from evolutionary biology to linguistic data, the parallels between biological species and dialects need to be assessed. Furthermore, the Bayesian clustering method performed by TESS underlies the biological assumption of linkage equilibrium, namely the independence between loci (morphosyntactic variables); hence, the Cramer V association measure is applied to SADS morphosyntactic variables and the variables presenting a Cramer's V less than 30% are kept. The Bayesian clustering algorithm TESS is performed by using the independent morphosyntactic phenomena. Three models are performed: the first one estimates the admixture proportions for each municipality without taking into account the spatial distribution of the samples; the second model incorporates a spatial prior distribution on the municipality admixture proportions and the global effect of geography is estimated. The third model evaluates not only the global effect of geography but also the local one, through the integration of spatial autocorrelation in the modeling. Hence, the importance of geography in explaining the morphosyntactic variation is evaluated, and the selection of which model explains best the data is assessed with the *Deviance information criterion* (DIC). In order to check the robustness of the results, the Bayesian clustering algorithm STRUCTURE is also performed. With TESS admixture proportions results, a dialectal diversity analysis is also assessed, by performing the Shannon diversity index. Furthermore, in order to identify geographical clines between the inferred ancestral dialectal families, a color interpolation technique is performed.

The methodological work-flow of this study is presented in figure 1.2.

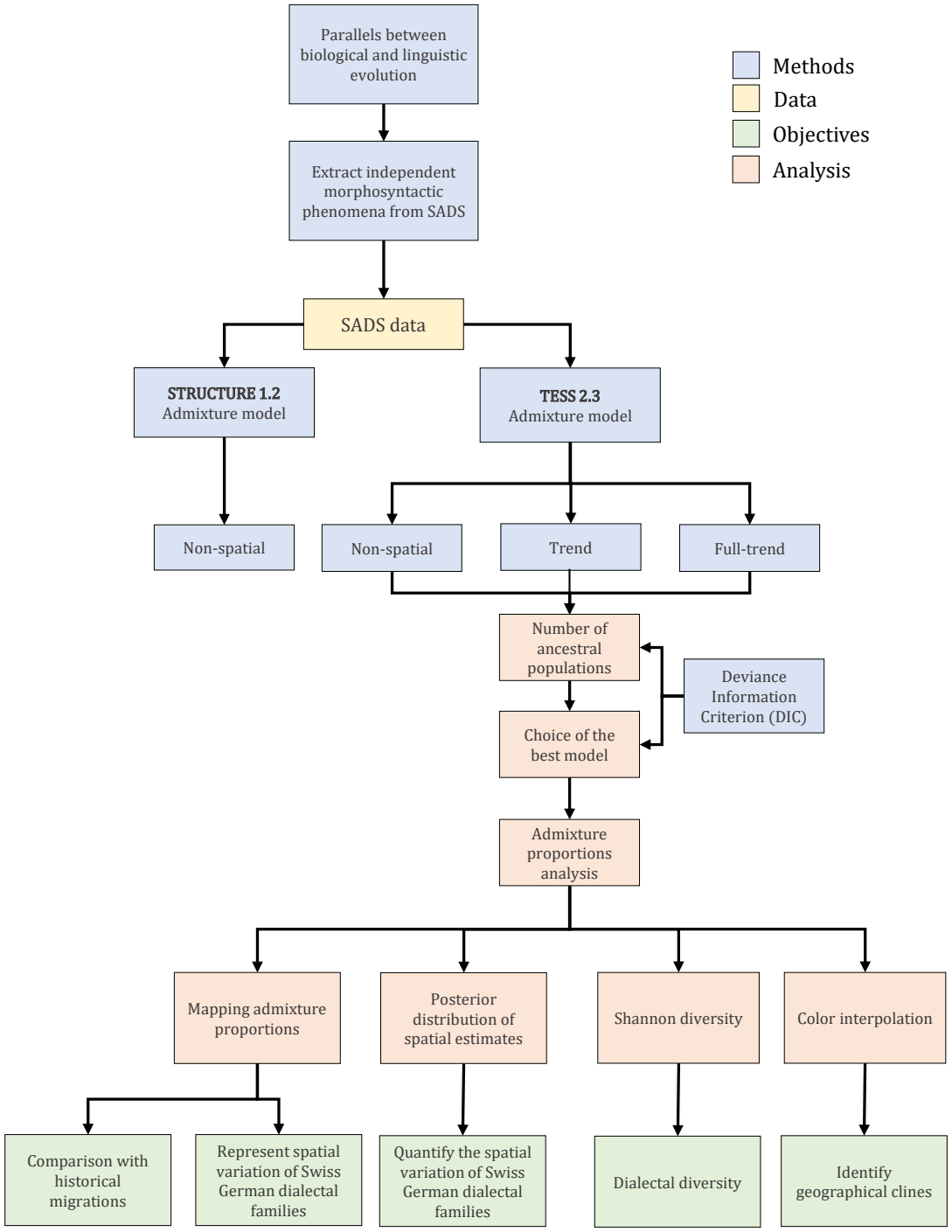


Figure 1.2: Work-flow of this study

Chapter 2

State of the art

Dialectology is the scientific branch of sociolinguistics that aims to investigate accents and linguistic dialects variation specific to a geographic area. In the last century, thanks to technology progress and the availability of complex linguistic atlases, numerous computational techniques have emerged to investigate the spatial variation of dialects. The branch of dialectology dealing with quantitative and computational techniques is known as dialectometry [67]. It has been introduced by Séguy in 1973 [84], who analyzed the linguistic distances for specific linguistic levels between different localities in Gascogne, an ancient province in southwest France. Séguy was supported right after by Goebl [39], who provided a rigorous work-flow to assess dialectometric analysis [68]. The principal criticism arising from their research against traditional dialectology is about the analysis of single linguistic features, which might neglect the general patterns of dialect variation [106]. So, the primary aim of dialectometry is to bring out the overall tendency of dialects spatial variation with the use of linguistic distances across geographic areas.

2.1 Linguistic distances

Dialectometry is mainly based on the investigation of linguistic distances across space. From the beginning of the last century, several techniques have been developed to measure linguistic distances for different linguistic data.

Séguy and Goebl [39, 84] are the pioneers of dialectometry and first measured linguistic distances or similarities at a nominal or categorical level. While Séguy measured dialectal distances, Goebl looked into the investigation on dialectal similarities. Séguy used the Hamming distance [44] which measures the differences based on the comparison of co-occurrence of linguistic features between each pair of locations; it has been further used in dialectometry for Dutch dialects [91]. Goebl [36] introduced the Weighted Identity Value (WIV) measure, which

aims to weight linguistic similarities depending on the rareness of specific linguistic features. Weighted Identity Value (WIV) has been applied to French dialects [38]. Both Hamming distance and WIV measures are well suited for *syntactic*[1] data.

Concerning string distances, Levenshtein distance (LD) [57] has been introduced in dialectometry by Kessler in 1995 [55] and it is used to quantify the distance between strings. It is based on the minimum number of single-character operations (insertions, deletions or substitutions) required to change one string to another [22]. LD is used to measure *phonetic*[2] and *lexical*[3] distances between dialect variants. It has been widely used in dialectometry for phonology analysis, for instance using phonetic data in Dutch [48], Norwegian [40] and Catalan [102] dialects, among others.

These distance measures are used to detect dialectal differences and can be the basis of further cluster analysis that takes distance or similarity matrices as input data.

2.2 Distance-based clustering

As already mentioned, the main aim of dialectologists is to delineate dialectal boundaries and detect the spatial variation of dialectal varieties. To do so, several scholars proposed to perform clustering methods based on linguistic distances.

2.2.1 Hierarchical Agglomerative Clustering

Hierarchical Agglomerative Clustering (HAC) has been widely used with linguistic and dialectical data [37, 39, 74, 81, 95]. This method aims to minimize clusters intra-variation and maximize clusters inter-variation from distance matrices. However, in the context of dialects in contact, areas are not considered a realistic framework. Indeed, the spatial organization of dialects in contact is bound to produce influence between neighbors, leading to the formation of continua. Hence, the general framework of *fuzzy clustering* techniques are more suited for the Swiss German dialectal landscape, and other techniques have been developed to deal with the dialect continuum framework.

2.2.2 Dimensionality reduction

Dimensionality reduction is based on the transformation of a higher-dimensional space to a lower one. Multidimensional scaling (MDS), Principal Component Analysis (PCA) and

[1] *Syntactic*: concerning the syntax, namely how sentences are built.
[2] *Phonetic*: concerning the phonology, namely the pronunciation of words
[3] *Lexical*: concerning lexicon

Factor Analysis (FA) are examples of these statistical techniques assessing this dimensionality reduction. For a more detailed explanation, see appendix A. Several scholars performed dimensionality reduction in dialectometry (e.g., [26,48,65,75,86]). The main aim in applying these algorithms is to reduce the number of variables and gradually color each unit depending on their coordinates values in each dimension in space. Hence, units with similar coordinate values have similar colors, and vice-versa. Proll [75] concluded that this method might be useful to detect the cores of the groups (clusters) and their gradual boundaries. However, only the observations with high loadings are informative from a statistical point of view [106]. These methods are beneficial for explanatory analysis but being implicit models, the assignment of the individuals to a particular cluster is not direct. Hence, model-based clustering methods are an interesting tool to assign directly individuals to clusters and several scholars applied these techniques in both linguistics and dialectology.

2.3 Model-based clustering

Model-based clustering techniques determine the clusters with a probabilistic model inferred from the data. Compared to distance-based clustering, these models take raw data as input matrices, and their inference principle does not rely on the distances between individuals. Bayesian clustering techniques form part of this clustering family and will be explored in this study.

2.3.1 Bayesian clustering

In the last decades, Bayesian clustering techniques have gained popularity in several fields such as genetics, linguistics and text analysis [6,94]. The most widely acknowledged Bayesian clustering algorithm is performed by STRUCTURE [73]. STRUCTURE has been primarily developed for population genetics analysis and has been applied in linguistics as well [8,23,78,94]. This method is considered a powerful tool to infer population structure from genetic data [32,73]. First of all, STRUCTURE infers one to K unknown populations and simultaneously computes individual *admixture proportions*[4] from each of these populations in a Bayesian framework [73]. Hence, every individual can belong to one or several populations, resulting in a probabilistic assignment of each individual to populations (Figure 2.1).

This method falls, therefore, in the broader framework of fuzzy clustering. It helps to detect the cores of the clusters (where individuals belong to only one population) and their fuzzy boundaries (where individuals belong to several populations). Thus, this addresses the modeling of dialect continuum.

[4] The level of membership of each individual to each sub-population.

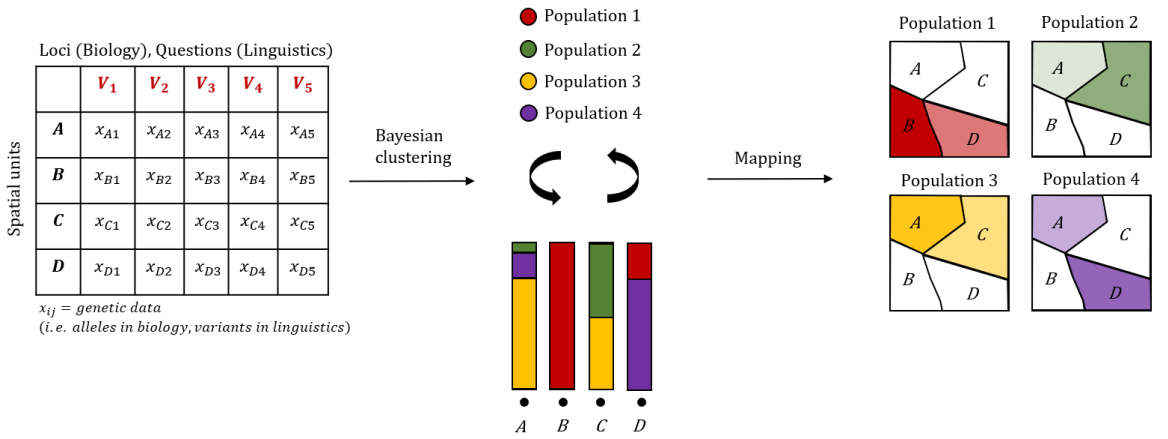


Figure 2.1: Application of Bayesian clustering algorithms in dialectometry

Other Bayesian clustering algorithms similar to STRUCTURE have been introduced since then, each one incorporating new parameters in the model, such as the integration of spatial information [13,15,25,43], *linkage disequilibrium*[5] [16,28,49], *inbreeding*[6] [32,34] or migration [107]. None of the spatial models has been applied to dialect data up until now. This thesis addresses this research gap by integrating the geographic information through the application of the admixture model of TESS [25].

Being TESS primarily developed for population genetics analysis, biological assumptions are taken into account in the inference. So, analogies between evolutionary biology and linguistics need to be assessed.

2.4 Parallels between biology and linguistics

Parallels in biological and linguistic evolutions have emerged since Darwin [20]. Similar evolutionary processes can be encountered in both disciplines. To apply evolutionary methods to linguistic data, the first assumption concerns the units of interest: languages are considered as biological species, linguistic levels as discrete biological characters and linguistic variants as genetic/phenotypic markers [1, 71].

To model dialects continua, several parallels have been drawn. First of all, *geographical clines*, namely a gradual evolution of a biological character across its geographic range [45], is a straightforward biological analogy of dialect continuum. In biology, the formation of geographical clines is consequence of two simultaneous evolutionary forces: *selection* and *gene*

[5] *Linkage disequilibrium*: independence between loci
[6] *Inbreeding*: it occurs when individuals closely related come into contact

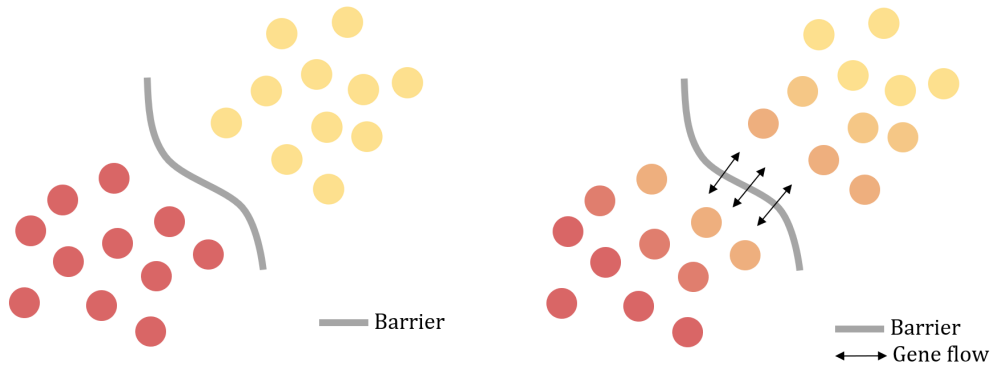
flow[7] [60]. Selection occurs when species adapt to a specific environment and their genetic information changes across generations (Figure 2.2a). Gene flow occurs when two distinct populations come into contact and gene information is transfered from one population to another [54]. Gene flow opposes the divergent force of selection in the zones where populations come into contact and therefore blurs the inter-population transitions. Hence, this leads to the formation of clines (Figure 2.2b).

Gene flow can be found in linguistics where it is called *borrowing* [1, 71]; Borrowing occurs when neighboring languages or dialects are in contact and influence each other. Similar, also the process of selection can be found in linguistics in the form of *social selection* [1, 71], caused for instance by the presence of administrative boundaries or natural barriers.

According to these parallels, two assumptions have been strengthened. Since the evolutionary processes in linguistics and biology are comparable, the use of linguistic data as genetic evidence is not arbitrary. Secondly, continua are a more realistic framework for dialects in contact, due to inter-dialectal influence.

Many other parallels exist between the two fields (Table 2.1), but in-depth analysis are not assessed in this study. However, most of these parallels are discussed in Croft [19], Atkinson and Gray [1] and Pagel [71].

The analogies between biology and linguistics are widely acknowledged in the literature. However, two broader families of language evolution have been proposed at the end of the 19th century: the cladistic theory, which represents language evolution through phylogenetic trees [82] and the wave theory which assumes language evolution as a diffusional process [83].



(a) Selection **(b)** Gene flow

Figure 2.2: Cline formation caused by opposite evolutionary processes: selection and gene flow.

[7] also known as migration.

Biological evolution	Linguistic evolution
Discrete characters	Lexicon, syntax, and phonology
Homologies	Cognates
Mutation	Innovation
Natural selection	Social selection
Cladogenesis	Lineage splits
Horizontal gene transfer	Borrowing
Hybridization	Language Creoles
Correlated genotypes/phenotypes	Correlated cultural terms
Geographic clines	Dialects/dialect chains
Fossils	Ancient texts
Extinction	Language death
Isophenes	Isoglosse

Table 2.1: Parallels between biological and linguistic evolution [1, 71]

2.5 Trees and waves

In linguistics, evolutionary biology has been mostly used for language phylogeny analysis[8], which addresses the detection of genealogical relations between different languages. Trees usually represent genealogical relations. The tree-model considers only the vertical inheritance and assumes that all the languages have a common ancestor. However, this model has been considered inadequate to infer the phylogeny of dialects in contact [30, 31, 41, 64, 69] due to the negligence of the horizontal component of language evolution[9], mostly due to language contact rather than inheritance from a common ancestor. In fact, the underlying tree-model assumption that a parent language splits into several lineages and that each of them evolves independently with the others is unlikely to occur in a context of languages in perpetual contact [104].

In historical linguistics, another theoretical model has been considered more appropriate to represent processes occurring when languages are in contact: the *wave model*. This theory has been introduced by Schmidt [83] in 1873 and affirms that a language trait spreads gradually from a central point to its neighborhood with its specific delimited range. This theory has been the basis for most analysis of dialect continuum since its underlying hypothesis is that language traits are transfered in a diffusional process: they gradually diffuse to their neighborhood as a consequence of contact. This theoretical model has been long after introduced in biology by Menozzi [61] and is known as the demic diffusion model [1]. However, the wave model considers this diffusional process only for single language traits. Tess algorithm is in line with the wave

[8] For a more detailed literature review concerning language phylogeny, I suggest to read Dunn work [22].
[9] i.e. gene flow, borrowing.

model and is well suited to represent the horizontal component of evolution [78]. However, the wave model assumes that this diffusional process occurs for single language traits. Hence, with the use of TESS, this study aims to analyze this diffusional process in an aggregate framework.

The transfer of language traits does not occur in the same way for all linguistic levels. While lexicon and phonology are transfered faster, syntax and *morphology*[10] are subject to slower processes [97]. In the context of this study, the *morphosyntactic*[11] level of Swiss-German will be explored.

2.6 Morphosyntax and geography

Morphosyntax has been of minor interest in general linguistics and dialectometry [2,35] because its variation is considered much subtler and less salient, entailing then the need of richer datasets [56]. However, in the last decades, an important interest has been paid to the collection of morphosyntactic data, by addressing new research questions and by bridging the gap between dialectology studies and typology [2,35,56]. Moreover, a clear spatial tendency of morphosyntax can hardly be detected, which might be another reason why morphosyntax has been investigated less than other linguistic levels [96]. In fact, several studies argued that syntactic variation is less likely to be correlated with geographic distances [95]. However, with the availability of new morphosyntactic datasets, several scholars investigated the spatial structure of syntactic variation [4,35,42,85,96] and some of them agreed that specific syntactic characteristics have recognizable spatial patterns [4,35,85].

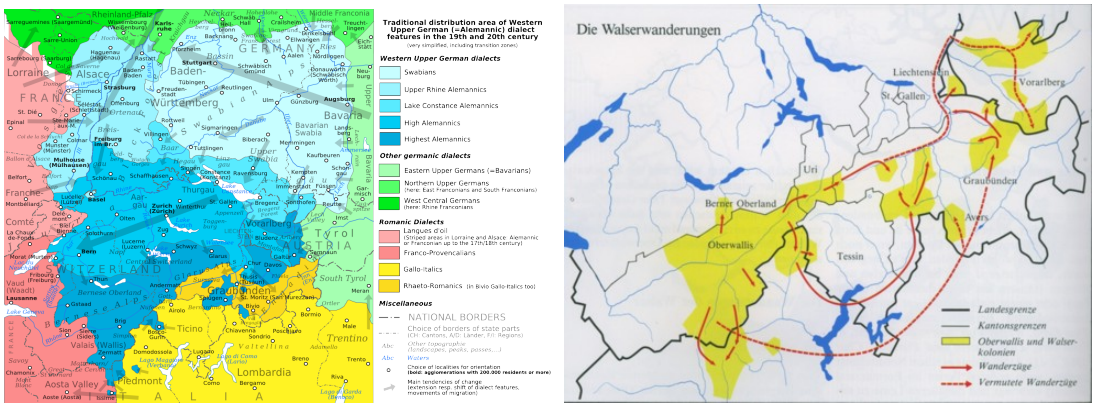
Evolutionary biology methods have been applied to syntactic data, in the context of the phylogeny of languages and population genetics. Longobardi and Guardiano [59] applied Parametric Comparison Method (PCM) to identify the genealogical relatedness with the utilization of morphosyntactic data. Furthermore, Reesink *et al* [78] applied STRUCTURE over abstract structural characteristics of Sahul languages and affirmed that this method is suitable for typological features, which can be both inherited (*vertical transmission*) and borrowed (*horizontal transmission*). Both studies concluded that morphosyntax can be considered as a good signal of historical linguistic relatedness.

[10] *Morphology*: the study of words and their functions

[11] *Morphosyntactic*: related to morphosyntax, involving syntax and morphology linguistic levels. Therefore, how sentences are built and their functions

2.7 Swiss German dialects

Swiss German is one of the liveliest dialectal families in central Europe [81]. It is historically divided into three major dialect groups: Low, High and Highest Alemannic. They are spoken in distinct regions: Low Alemannic in Basel city and in regions around Lake Constance, High Alemannic in the Swiss Plateau and Highest Alemannic, which are spoken in alpine regions of Switzerland (Figure 2.3a). One of the most ancient migrations concerns the Walser population (Figure 2.3b), based in the canton Wallis, which migrated within Switzerland and in the neighboring countries Liechtenstein, Austria, and Italy. Within Switzerland, the Walser migration occurred mostly in the south of canton Uri and Grisons (Figure 2.3b).



(a) Dialectal division map, from Wikimedia [14] **(b)** Walser migration map, from Fibicher [29]

Figure 2.3: Historical maps of Swiss German dialects

In recent times an interest has been paid to the investigation of Swiss german dialects linguistic variation. Since the beginning of the 20^{th} century, great interest has been demonstrated on Swiss digital linguistic data collection, e.g. the *Sprachatlas der deutschen Schweiz* (SDS) [100] that covers phonetic, morphological and lexical variation and the *Syntaktischer Atlas der deutschen Schweiz* (SADS) [10] that captures the morphosyntactic level.

In order to explore the spatial variation of morphosyntax, several dialectometric analysis have been proposed, such as the standard point symbol maps, Kernel Density Estimation to interpolate intensity values of all variants and 3-D visualizations to explore the spatial trends [88,89]. In agreement with Sibler's works, Bart *et al.* [5] applied various spatial statistics techniques such as Moran's I index, semivariograms and trend surface analysis to evaluate the spatial distribution of specific linguistic variants.

Jeszensky and Weibel proposed several GIS methods to assess dialectal boundaries delineation and the spatial variation of specific morphosyntactic variants [52, 53]. They mapped the intensity of specific dominant variants and performed trend surfaces analysis to explore and quantify their spatial variation. These studies analyze individual morphosyntactic phenomena

and, therefore, do not explore the morphosyntactic variation in an aggregate way.

To my knowledge, only Scherrer and Stoeckle [81] analyzed the Swiss German spatial variation in an aggregate way. They performed HAC for four different linguistic levels and assessed the relation between the cluster assignment and geography. They concluded that syntax generally yields a different spatial distribution, in contrast to other linguistic levels.

This study aims to evaluate the morphosyntactic continuum of Swiss German in an aggregate manner, by using the SADS data and applying the spatial Bayesian clustering algorithm Tess.

Chapter 3

Materials and Methods

3.1 Data

3.1.1 Syntaktischer Atlas der deutschen Schweiz SADS

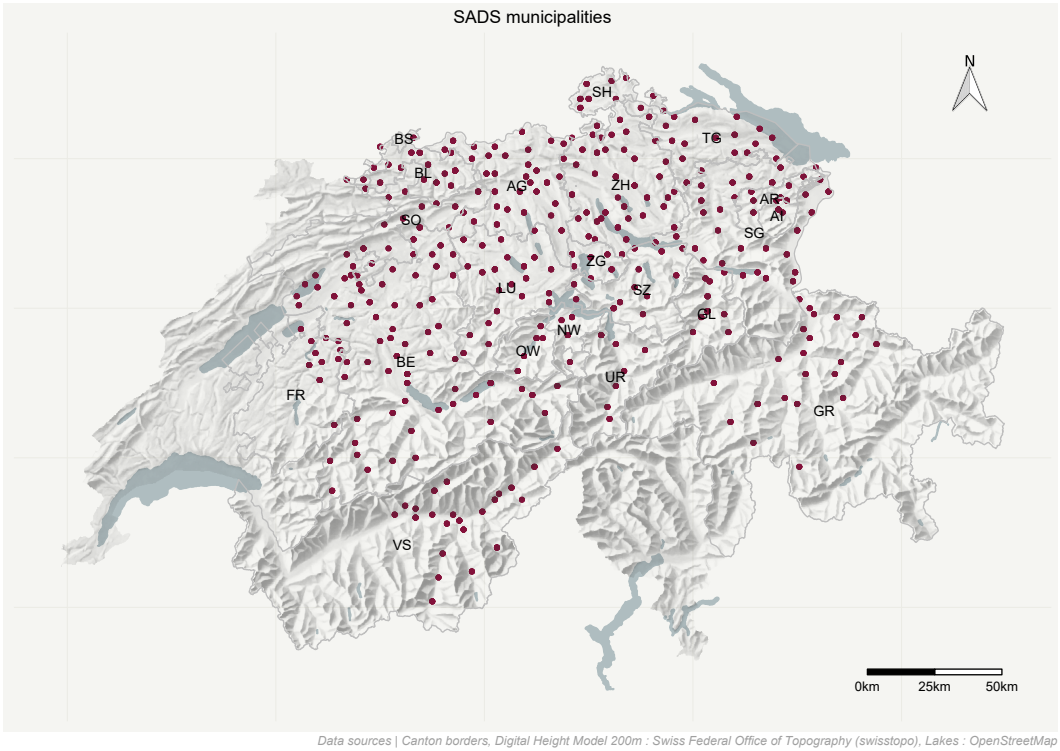


Figure 3.1: Municipalities of the SADS

In this study, the *Syntaktischer Atlas der deutschen Schweiz* (SADS) [10] has been used as input data for the analysis. As mentioned in section 2.7, the SADS covers the morphosyntactic variation of Swiss German. 383 Swiss German speaking municipalities have been surveyed

between 2000 and 2002 (Figure 3.1). The particularity of this linguistic dataset is that multiple respondents per municipality were surveyed. In fact, in the context of linguistics, only one respondent per spatial unit is usually sampled [52], leading to less robust conclusions about the linguistic variation. Around 3000 respondents have been surveyed, with 3 to 26 individuals per municipality (median value equal to 6-7). We excluded respondents with an unknown geographic origin, resulting in 2970 individuals. Furthermore, 88 independent phenomena have been chosen after an association measure analysis that is discussed in section 3.3.2.

3.1.2 Spatial data

Spatial data have been used for the maps visualizations.

Digital height model The digital height model with 200 m grid has been integrated to the maps in order to give a general overview of Swiss topography. It is freely provided by the Swiss Federal Office of Topography (SWISSTOPO) [93].

Lakes The vector file of Swiss lakes has been used as well in order to include water bodies on the maps. The vector file is freely available on OpenStreetMap [70].

Canton borders The vector file of Swiss canton borders have been used to include the administrative boundaries. It is freely provided by the Swiss Federal Office of Topography (SWISSTOPO) [92].

3.2 Principles of Bayesian inference

Before looking at TESS in more detail, it is essential to explore the fundamental ideas of Bayesian inference. A short overview of Bayesian statistics in genetics is found in Beaumont and Rannala [6].

Bayesian inference is based on Bayes' theorem, which aims to estimate the probability of an event depending on available prior information related to this event. Hence, in line with this theorem, Bayesian inference aims to incorporate background information while estimating the appropriate model and its parameters that influence the observed data.

Let's denote the observed data as D and the parameters as θ. The main principle of Bayesian inference is given by equation 3.1.

$$Pr(\theta|D) \propto Pr(\theta)Pr(D|\theta) \tag{3.1}$$

Bayesian inference aims to estimate the joint probability distribution of D and θ $Pr(\theta|D)$, by assuming that D are observed random variables and θ unobserved random variables.

The joint probability, also known as posterior distribution, is the product of the prior information on θ $Pr(\theta)$ and the likelihood $Pr(D|\theta)$, i.e. it is the probabilistic model describing the distribution of D given θ (Figure 3.2).

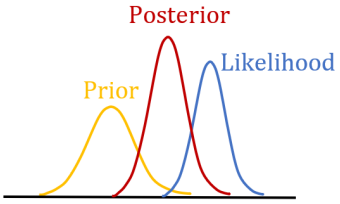


Figure 3.2: Product of prior distribution and likelihood gives rise to posterior distribution

3.3 TESS

TESS is a Bayesian clustering algorithm that is inspired by STRUCTURE, a widely used algorithm in population genetics that aims to infer population structure from multi-locus genotype data. In contrast to STRUCTURE, TESS allows the incorporation of spatial priors that account for local and global spatial effects in the modeling. This model takes into account the importance of geography, which is a relevant parameter in ecology since the divisions of different species are due to their ecological environments, which are specific to a geographic place. Furthermore, when spatial autocorrelation is not taken into account in ecological studies, the statistical inference can be biased [21,58].

In the same way, it is acknowledged that dialectal varieties are more similar between neighboring spatial units than distant ones. This is in line with Tobler's first law of geography [99] and the *Fundamental Dialectology Principle* [66], affirming both that geographically proximate varieties tend to be more similar than distant ones. TESS' admixture model is in line with these principles, and it will be used to test whether including geography in the modeling can yield a better inference of dialectal division.

The version 2.3 of TESS has been used and the simulations have been ran on 4-cores 16GB RAM laptop.

3.3.1 Data formatting

In order to apply this method to the SADS, the parallel between genetic data and the SADS data needs to be assessed.

Multi-locus genotype data generally contain *loci*[1] in which *alleles*[2] have been sampled for several individuals of the same species. The first parallel that can be done is between the loci and the different questions asked in the SADS[3]. Secondly, the different answers that individ-

[1] *Locus*: location of the gene
[2] *Allele*: variant of a gene
[3] i.e. the morphosyntactic phenomena that have been sampled

uals chose for the SADS can be considered the alleles. In the same way, as alleles are variants of the gene, we can consider the different answers as variants of a specific morphosyntactic phenomenon (see input data of Figure 2.1). Hence, from now on, loci and alleles are denoted as morphosyntactic phenomena and variants respectively.

3.3.2 Biological assumptions

Like STRUCTURE, TESS is an individual and model-based algorithm that aims to group individuals in sub-populations which are in *Hardy-Weinberg* (HWE) and *linkage* (LE) equilibrium. HWE states that the population inferred remains stable (allele frequencies of this population do not vary from generation to generation) without the occurrence of evolutionary forces, such as selection and gene flow. HWE populations would, therefore, correspond to ancestral populations that date further back in time and remained stable until the data were collected. Morphosyntax is considered a relatively stable characteristic of language or dialect when compared to other linguistic levels. This means that morphosyntactic transfer processes are slower than the lexical or phonological ones. Hence, using morphosyntactic data, we expect to detect historical signals that go further back in time when compared to other linguistic levels.

LE is the independent association of alleles at different loci in a given population. To comply with this assumption, the *Cramer's V* association measure [17] has been computed for all morphosyntactic features to keep only the variables that are poorly associated. This measure is a widely used technique in genetics to measure LE between multi-allelic loci. It indicates how strong two categorical variables are associated and ranges from 0 to 1. It is given by the following relation:

$$\phi_c = \sqrt{\frac{\chi^2}{N(k-1)}} \tag{3.2}$$

where ϕ_c corresponds to the *Cramer's V* coefficient, χ^2 is obtained with Pearsons's chi-squared test, N is the number of observations (individuals), and k is the number of variables. Only the variables having less than the threshold value of *30%* of association have been kept. This threshold value indicates a low strength of association, hence, a high LE.

3.3.3 Statistical model

TESS proposes two models: the *no-admixture* model and the *admixture* model. Both models allow to infer population structure, but are based on two different assumptions: while the no-admixture model assumes that each individual belongs to only one population, the admixture one assumes that an individual can come from several populations K ($K = 1, 2, ...K_{max}$).

The admixture model is appropriate when the ancestral populations are unknown, and it is expected that they were in contact, giving rise to admixed individuals. Hence, this method has been considered a more realistic framework for Swiss German dialects, where one would rather expect mixed dialects because of the historical migration and recent contact.

3.3.3.1 Variables

Let's assume X the linguistic data that have been sampled at different morphosyntactic phenomena and $\tilde{X}$ their respective geographic coordinates, Z the unknown ancestral populations, P the variants frequencies [4] within the unknown ancestral populations at each morphosyntactic phenomenon and Q the individual admixture proportions. Each of these vectors is composed as follow:

$$X : x_{im}$$

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{bmatrix}$$

where x_{im} is the answer that a given individual i $(i = 1, 2, ..., N)$ gave at a particular morphosyntactic question m $(m = 1, 2, ..., M)$.

$$Z : z_{im}$$

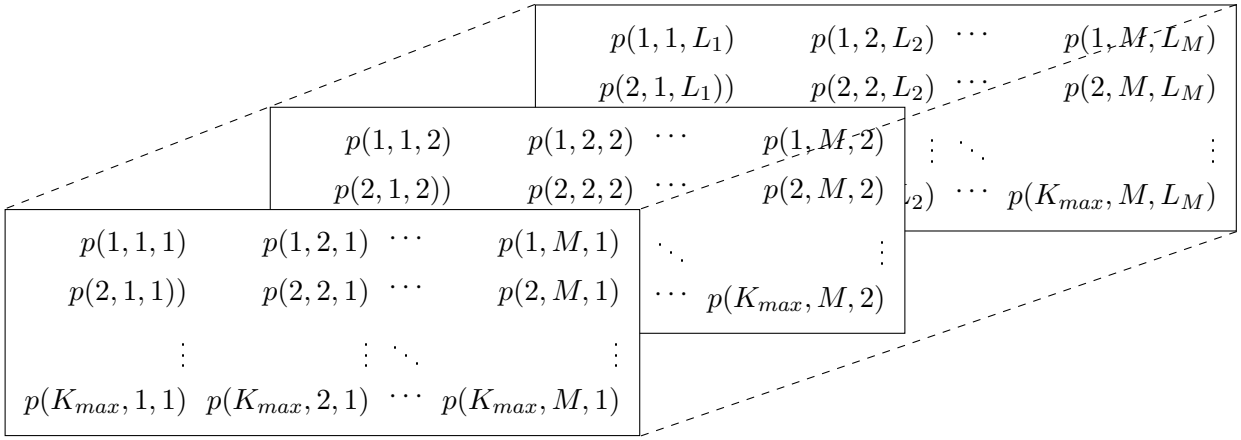
$$\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1M} \\ z_{21} & z_{22} & \cdots & z_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ z_{N1} & z_{N2} & \cdots & z_{NM} \end{bmatrix}$$

where z_{im} is the population from which a particular morphosyntactic phenomenon m arises from for every individual i. In simple terms, we want to predict the origin of the information that every individual carries for each morphosyntactic phenomenon.

The linguistic variant frequencies at each morphosyntactic phenomenon within the ancestral populations K are given by:

$$P : p_{kml}$$

[4] See section 3.3.1 for parallel done in this study between linguistic variant and allele.



where p_{kml} is the frequency of a particular linguistic variant l ($l = 1, 2, ..., L_m$) for a morphosyntactic phenomenon m in population k ($k = 1, 2, ..., K_{max}$) and

$$Q : q_{ik}$$

$$\begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1K} \\ q_{21} & q_{22} & \cdots & q_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N1} & q_{N2} & \cdots & q_{NK} \end{bmatrix}$$

where q_{ik} is the admixture proportion q of individual i from population k. In biological terms, q_{ik} refers to the proportion of an individual genome belonging to a particular population k, so the level of membership of this individual to a given cluster.

3.3.3.2 Likelihood

Knowing the ancestral populations and the frequencies of the linguistic variants for each morphosyntactic phenomenon, the following formula gives the likelihood function:

$$Pr(D|\theta) = Pr(X|Z, P, Q) = \prod_{i=1}^{N} \prod_{m=1}^{M} p_{z_{im}mx_{im}} \tag{3.3}$$

From the double product follows that morphosyntactic variant frequencies are sampled independently, hence LE is assumed.

3.3.3.3 Priors

The advantage of using Bayesian techniques lies the possibility to incorporate prior information into the model. Hereafter, the prior distributions of the parameters Z, P and Q are described.

As proposed by Balding and Nichols [3], the prior distribution of the linguistic variant frequencies at each morphosyntactic phenomenon m in each population K contained in the vector P follows a Dirichlet distribution

$$Pr(P) \sim Dir(\lambda_1, \lambda_2 ... \lambda_{L_m}) \tag{3.4}$$

where λ_l is a coefficient to which the frequencies of a given morphosyntactic variant l are proportional. We assume that $\lambda_1 = \lambda_2 = ... = \lambda_{L_m} = 1$, which implies a uniform distribution on the morphosyntactic variant frequencies [73].

The prior distribution of z_{im} given P and Q is

$$Pr(z_{im} = k|P, Q) = q_{ik} \tag{3.5}$$

and finally, the prior distribution of Q is a Dirichlet distribution

$$Pr(Q|\alpha) \sim Dir(\alpha_{i1}, \alpha_{i2} ... \alpha_{iK_{max}}) \tag{3.6}$$

where α_{iK} is a *hyperparameter*[5] to which the average admixture proportions $E(q_{ik})$ are proportional. While in STRUCTURE the α coefficients are described as a uniform distribution as in equation 3.4, the admixture model of TESS introduces a log-normal model with the following α hyperparameters

$$log(\alpha_{ik}) = f(\tilde{x}_{ik})^T \beta_k + y_{ik} \tag{3.7}$$

where $log(\alpha_{iK})$ is an unobserved response variable, $\tilde{x_i}$ is a vector containing the geographic coordinates of each individual i, β_K contains the spatial trend coefficients of each population K and y_{iK} refers to a zero-mean spatially autocorrelated random variable. This log-linear regression model is inspired by the Universal Kriging model [18, 79] which predicts global and local spatial effects; in fact, the first term of the right hand side of equation 3.7 can be decomposed in two parts: $f(x_{ik})^T \beta_k$ corresponds to the mean response specified by a trend

[5] Hyperparameter: in the context of Bayesian statistics, hyperparameters are parameters of the prior distributions that need to be distinguished from the parameters of the model.

surface, and y_{ik} corresponds to the local response derived from a conditional auto-regressive (CAR) Gaussian model [7,103]. The CAR model is in line with the Fundamental dialectology principle and Tobler's First law of geography. In fact, the expected value of the residual term y_{ik} is given by the weighted sum of the values in its neighborhood as follows:

$$E(y_{ik}|y_{jk}) = \rho \sum_{i \sim j} w_{ij} y_j \tag{3.8}$$

where j is an individual in the neighborhood of i, ρ is the spatial autocorrelation, w_{ij} is the influence of individual j on individual i. Hence, this relation implies that neighboring dialectal varieties are more likely to be in contact and influence each other than distant or isolated dialectal varieties.

To define neighborhood, François [32] proposed a Voronoi tessellation, which defines neighbor spatial units as those with common edges. The weights are given by the exponential covariance matrix

$$w_{ij} = exp(\frac{-d_{ij}}{\gamma}) \tag{3.9}$$

where d_{ij} is the distance separating the spatial units i and j and γ is the average great-circle distance between the individual geographic coordinates. Based on equation 3.8, the covariance matrix is as follows:

$$\Lambda = \sigma^2 (Id - \rho W)^{-1} \tag{3.10}$$

where σ^2 is the variance of the CAR, Id is the identity matrix, and W is the matrix of the weights w_{ij}

$$W = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1j} & \cdots & w_{1N} \\ w_{21} & 0 & \cdots & w_{2j} & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots & & w_{3N} \\ w_{i1} & w_{i2} & \cdots & 0 & \cdots & w_{iN} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{Nj} & \cdots & 0 \end{bmatrix}.$$

Tess allows changing the spatial priors, such as the degree of the trend surface (no trend, linear, quadratic, cubic) and the degree of spatial autocorrelation ρ that can be left constant or estimated from the data. In this study, three different models are proposed by changing the

spatial prior information on the admixture proportions. Firstly, a model with a uniform prior (such as in STRUCTURE), by considering no spatial trend and letting ρ equal to 0. Secondly, only the global effect is incorporated, by letting the trend degree equal to 1 (linear trend) and ρ equal to 0. Moreover, the global and local effects are taken into account by considering the trend degree equal to 1 and ρ estimated by the data. From now on, these models will be named *non-spatial*, *trend* and *full-trend model* respectively. An overview of the priors on the individual admixture proportions used in this study is given in table 3.1.

Model	**Prior**
Non-spatial	$log(\alpha_{ik}) = 0$
Trend	$log(\alpha_{ik}) = f(x_{ik})^T\beta$
Full-trend	$log(\alpha_{ik}) = f(x_{ik})^T\beta + y_{ik}$

Table 3.1: Priors on the individual admixture proportions

A summary of the parameters estimated and their relative prior distributions notation is given in table 3.2

Parameters	**Description**	**Prior notation**
Z	Ancestral populations	$Pr(Z\|Q)$
P	Variant frequencies for each morphosyntactic phenomenon in each population	$Pr(P)$
Q	Admixture proportions	$Pr(Q\|\alpha)$
α	Hyper-parameter proportional to admixture proportions	$Pr(\alpha\|y, \beta, \tilde{X})$
y	Spatial autocorrelation	$Pr(y\|\rho, \sigma^2)$
σ^2	Variance of covariance matrix	$Pr(\sigma^2)$
ρ	Strength of spatial autocorrelation	$Pr(\rho)$
β	Trend coefficients	$Pr(\beta)$

Table 3.2: Admixture model parameters

3.3.4 Inference of parameters

In line with equation 3.1 and the prior distributions described in table 3.2, the posterior distribution is therefore given by

$$\begin{aligned} Pr(Z, P, Q, \alpha, y, \rho, \sigma^2, \beta|X, \tilde{X}) \propto Pr(X|Z, P, Q)Pr(Z|Q)Pr(Q|\alpha)\times \\ Pr(\alpha|y, \beta, \tilde{X})Pr(P)Pr(y|\rho, \sigma^2)Pr(\rho)Pr(\beta) \end{aligned} \tag{3.11}$$

The posterior distribution cannot be calculated explicitly due to its multidimensionality; however, Markov chain Monte Carlo MCMC is the statistical framework to sample from such

complex distributions. MCMC aims to approximate a target distribution[6], by generating random numbers (Monte Carlo) in a probabilistic space. A Markov chain implies that each iteration depends only on its previous iteration. The main aim of the algorithm is to construct a Markov chain with stationary distribution. In this study, several runs have been performed to check how many iterations were needed to converge to a stationary distribution, and 30000 iterations were enough to converge. Since a Markov chain implies that samples depend on the starting values, the first 10000 iterations have been discarded. Detailed explanations about the MCMC processes and the methods used by TESS to update the parameters are given in appendix B.

3.3.5 Model choice and number of ancestral populations

TESS provides a statistical criterion to choose which model best fits the data and how many ancestral populations are detected. The statistical criterion is called Deviance information Criterion (DIC) [90] and can be decomposed into two model measures: the goodness of fit and the number of free parameters[7] used in the model. The lower value of the DIC, the better is the model.

Let X the data and θ the parameters used in the model. DIC is given as follows:

$$DIC = E_{\theta|X}[D(\theta)] + p_D \tag{3.12}$$

where $E_{\theta|X}[D(\theta)]$ is the posterior expectation of the deviance, namely a statistical measure of goodness of fit, and p_D is the number of free parameters in the model. $D(\theta)$ is given by

$$D(\theta) = -2log(Pr(X|\theta)) + C \tag{3.13}$$

where $Pr(X|\theta)$ is the likelihood function (see equation 3.1) and C is a constant.

p_D is a measure of model complexity and is given by

$$p_D = E_{\theta|X}[D(\theta)] - D(\bar{\theta}) \tag{3.14}$$

where $\bar{\theta}$ is the expectation of θ. The penalization of the goodness of fit through the number of effective parameters is because the more parameters we use, the more the model will fit best, but it does not imply that the model is a good model. Hence, this issue is tackled by taking into account the model complexity while calculating the DIC.

[6] The posterior distribution that we want to sample from

[7] effective parameters

TESS measures the DIC for every MCMC generation and it used to choose the number of ancestral populations K_{max}, and, fixed K_{max}, the best of the candidate models [25]. Within TESS, the DIC is calculated for several K_{max} values[8] and then plotted against each K_{max}. The K_{max} where the DIC values reach a plateau is considered the most suited number of ancestral populations. In other words, when the DIC values reach a plateau it means that from this point the model does not improve significantly. In this study, K_{max} was ranged from 2 to 8 for 15 runs. Furthermore, the 5 lowest DIC values have been averaged and plotted against each K_{max}.

Once K_{max} has been fixed, we can choose which model best fits our data. Hence, non-spatial, trend and full-trend models are compared by their average DICs, and the model showing the lowest DIC is chosen for further analysis described in the next sections. The errors in DIC comparisons are of order $\sqrt{N}$ where N is the number of sampled individuals [25,80]. In this study, $\sqrt{N} = \sqrt{2970} = 54.5$.. Hence, in order to make robust conclusion, the choice of the model does not only depend on the DICs values themselves but also on the differences between each model DIC.

3.4 Analysis

Every analysis is performed on the estimates of admixture proportions q_{ik} obtained with TESS. First, admixture proportions q_{ik} are mapped for every ancestral population K to have an overview of their spatial variation. Second, the spatial trend estimates β are analyzed for every population in order to evaluate the importance of geography in explaining their morphosyntactic variation. Third, from q_{ik} results, the Shannon diversity index is calculated for each spatial with the aim of evaluating their dialectal diversity and figure out whether we find hybrid or homogeneous areas. Moreover, a novel method is introduced to visualize the dialect continuum from admixture proportions q_{ik}.

3.4.1 Admixture proportions q_{ik}

First of all, the 5 runs with the lowest DICs have been chosen, and the q_{ik} estimates have been averaged with the software CLUMPP [51]. CLUMPP software has been used in order to match the populations inferred between the different runs since in each run they are differently labeled. With the use of this software, q_{ik} are then averaged for each individual. This is done to have a more correct Bayesian estimate of the individual admixture proportions [24].

The usual way of visualizing admixture proportions is with a stacked barplot for every individual such as those presented in the Bayesian clustering schema (Figure 2.1). In order to

[8] K_{max} is chosen by the user

get a first insight about the spatial distribution of the admixture proportions, the different individuals' admixture proportions are visualized with stacked barplot ordered by latitude and longitude.

All the individuals sampled in the SADS have been submitted to TESS. However, knowing only their municipality origin, the individuals' geographic information within each municipality overlap. Hence, to have one value per spatial unit (municipality), a Voronoi tessellation has been created from the municipalities geographic coordinates and the q_{ik} has been aggregated at the Voronoi cell level by taking the mean of q_{ik}, which from now on is denoted as $\bar{q}_{ik}$. The Voronoi tessellation has been chosen because it has been the basis of the spatial prior on the admixture proportions in TESS (see section 3.3.3.3). Furthermore, with the Voronoi tessellation we yield a continuous spatial structure[9].

The aggregation has been done after the simulations in order to keep all the information. In fact, if we did the aggregation at the municipality level before the integration in TESS, we should have taken some arbitrary choices, such as choosing the dominant variant in each municipality since TESS takes as input data only categorical variables.

After the aggregation at the Voronoi cell level, the mapping of $\bar{q}_{ik}$ has been accomplished with the `ggplot` [105] library in `R` [77]. The Voronoi tessellation and the aggregation of individual admixture proportions have been performed with the *Join attributes by location* plug-in in QGIS software [76].

3.4.2 Spatial trends

The spatial trends β decomposed in latitudinal (β_{lat}) and longitudinal trend (β_{long}), will be analyzed for each ancestral populations inferred. The uncertainty of the spatial trend estimates is assessed by calculating the 95% credible intervals[10] ($\text{CI}_{95\%}$) of the β estimates.

Since we cannot average the different distributions over the different runs, the spatial model (full-trend or trend model) presenting the run with the lowest DIC is chosen and analyzed. If the value 0 is in the range of the spatial trends $\text{CI}_{95\%}$, there is no significant spatial trend; if the range of $\text{CI}_{95\%}$ do not include the value 0, there is then a significant spatial trend. $\text{CI}_{95\%}$ have been calculated with the `quantile` library in `R`.

[9] For instance, if we were using the municipalities borders as spatial units we would have had a discontinuous spatial structure since not all Swiss German municipalities have been submitted to the SADS questions.

[10] Credible intervals are similar to the notion of confidence intervals, but are specific to Bayesian approaches.

3.4.3 Dialectal diversity

Dialectal diversity analysis is performed with the well known Shannon diversity index (H) [87]. This index has been first introduced to quantify the uncertainty in strings of text and later applied in ecology to calculate species diversity in a community. In line with the use of H in ecology, the dialectal diversity for each spatial unit i is performed as follows:

$$H_i = -\sum_{k=1}^{K_{max}} \bar{q}_{ik} ln(\bar{q}_{ik}) \tag{3.15}$$

Hence, in the same way such as in ecology, H gives how much a spatial unit is diverse and helps to detect hybrid or homogeneous zones.

3.4.4 Color interpolation

The H index gives a general overview of spatial unit diversity but does not provide any information about which populations contribute to their diversity. Generally, research using Tess make use of simple Kriging interpolation to show all the ancestral populations inferred in one map; however, zones where individuals have equal contributions of two populations take the white color such as they do not belong to any of these populations. To tackle this problem, Tess Ad-Mixer software [63] provides a framework where transitions areas, namely where populations come into contact, are colored with RGB color interpolation. However, it can be used only on raster files obtained with Kriging interpolation and not with vector files such as the case of this study. Hence, a novel method is proposed to allow this color interpolation in a vectorial framework.

First of all, for the sake of simplicity, only the two dominant ancestral populations contributing to each spatial unit are extracted, and the municipality admixture proportions are recalculated such that they sum up to 1. Depending on the mixing proportions (the amount of contribution from dominant population 1 and dominant population 2), categorical classes representing the type of mixture are designated (Figure 3.3).

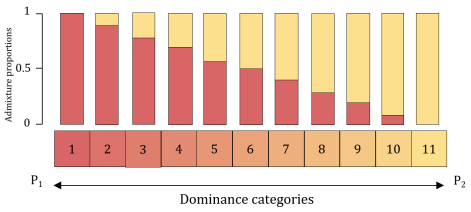


Figure 3.3: Categories to which different mixing proportions are assigned.

Then, N unique color palettes has to be created, and N is given by

$$N = \binom{K_{max}}{c} = \frac{K_{max}!}{c!(K_{max} - c)!} \tag{3.16}$$

where K_{max} the total number of ancestral populations and c the number of ancestral populations contributing to each individual ($c = 2$ in this study). Hence, N color palettes are created containing a color interpolation of each unique pairwise combination. The color palettes are created with the function `colorRampPalette` of R. In this study, 11 dominance categories have been chosen, and each spatial unit has been assigned to one of $11xN$ colors.

3.4.5 Robustness of the model

STRUCTURE has been also performed to assess the robustness of TESS model. The main difference between the two models is the prior information on the admixture proportions. While TESS includes a spatial prior, STRUCTURE assumes a uniform distribution. STRUCTURE model is therefore very similar to the non-spatial model of TESS. STRUCTURE results are in appendix E and will serve as a basis for robustness assessment.

Chapter 4

Results

4.1 Deviance information criterion

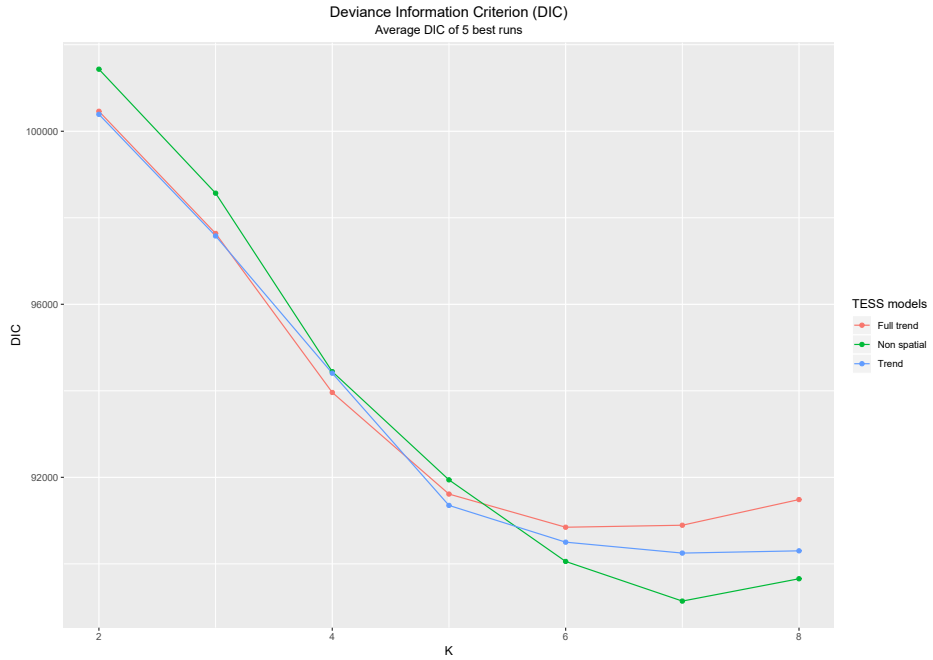


Figure 4.1: Deviance Information Criterion (DIC)

Figure 4.1 presents the average DICs over the 5 best runs for each model. Generally, the spatial models outperform the non-spatial model until K_{max} equal to 6, except for 4 ancestral populations where the trend model yields average DICs values similar to the non-spatial one. With more than 5 ancestral populations, the non-spatial model outperforms the spatial models. DICs values do not seem to diminish significantly after K_{max} equal to 5 for spatial models. The non-spatial model seems to best fits the data with more than 5 ancestral popula-

tions; furthermore, non-spatial average DIC values reach a plateau at 6 ancestral populations. Since spatial and non-spatial models yield different conclusions as regards to the number of ancestral populations ($K_{max} = 5$ for spatial models, $K_{max} = 6$ for non-spatial models), both configurations are analyzed hereafter. For $K_{max} = 5$, the trend model has been chosen and all the populations inferred are presented. For the non-spatial model, only the additional populations detected are presented and the rest of the populations maps can be found in appendix D.

The full-trend model outperforms the other models only for $K_{max} = 4$, which is not considered a suitable number of ancestral populations since the plateau is not reached yet. Moreover, the difference between DIC_{trend} and $\text{DIC}_{full-trend}$ is 3396, and trend model can be chosen against the full-trend model certainly. Therefore, the full-trend model is not investigated in detail. However, the results of $K_{max} = 5$ for the lowest DIC run, including β trends and the spatial autocorrelation ρ results are presented in the appendix C.

4.2 Admixture proportions q_{ik}

The following maps represent the municipality admixture proportions for each ancestral population detected with the trend and non-spatial models.

4.2.1 $K_{max} = 5$, Trend model

4.2.1.1 Barplots

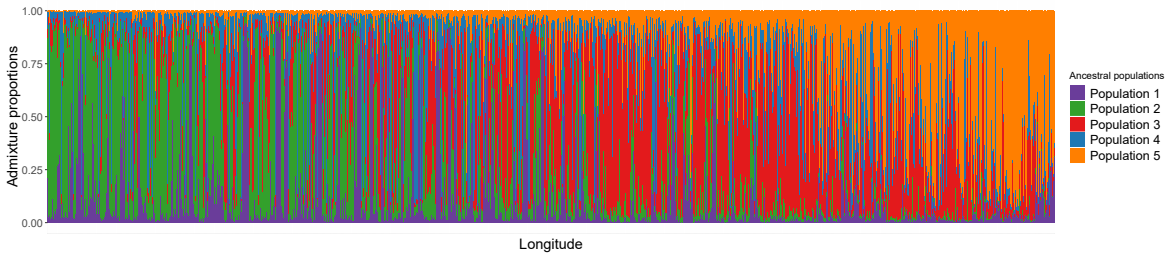


Figure 4.2: Municipality admixture proportions ordered by longitude

In the longitudinal direction (Figure 4.2), we can see that population 2 is mostly concentrated in the west part and gradually decreases from West to East; moreover, it seems to highly contribute for at least 1/4 of the municipalities. Reversely, population 5 gradually increases from West to East. Population 3 seems to appear in the center of Switzerland, and it is noteworthy that there is a clear cline from population 3 to population 5. Population 1 and 4 do not seem to follow a clear longitudinal gradient.

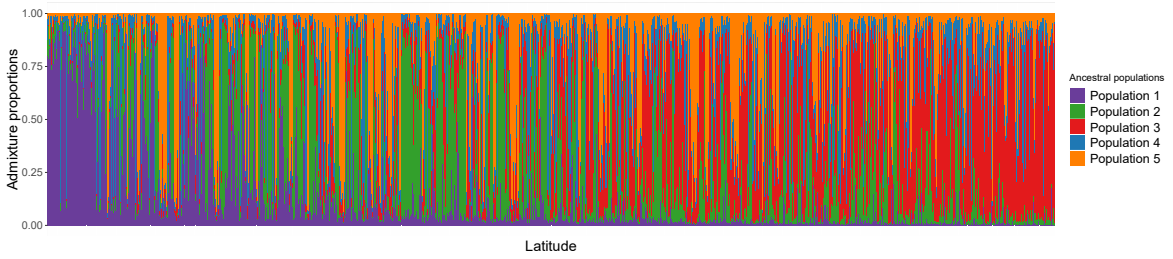


Figure 4.3: Municipality admixture proportions ordered by latitude

In the latitudinal direction (Figure 4.3), population 1 seems to contribute in the South part of Switzerland and it is not detectable in the North part; population 3 follows the opposite latitudinal direction than population 1, but does not seem to have a clear gradual variation. Population 5 does not seem to have a clear latitudinal variation.

4.2.1.2 Maps

Ancestral population 1 (Figure 4.4) contributes mostly to the VS canton and gradually diminishes in the North direction until the southern part of BE, part of the Swiss German speaking of FR and some municipalities of the central part of Switzerland. Moreover, it spreads in the East direction along the OW, NW, SZ and UR cantons until reaching the GR.

Ancestral population 2 (Figure 4.5) covers most of the municipalities in the western part. The most significant contributions mostly concern canton BE and FR. From the western part of canton BE and the canton FR, it spreads gradually in all directions: until BS in the North direction, in the South direction until the VS border and in the Eastern part until ZH and SZ.

Population 3 (Figure 4.6) finds its highest admixture proportions in the northern part of Switzerland and gradually varies in the South direction. If we take the central point in canton SH, the admixture proportions diminish continuously in all directions reaching canton SO, LU, UR, OW, GL, SG, AI, and AR.

The spatial variation of ancestral population 4 (Figure 4.7) does not show a clear pattern; most of the municipalities seem to originate to this population. This population disperses randomly, and we cannot recognize a continuous pattern. Furthermore, a unique center from where this population could hypothetically originate cannot be recognized.

Ancestral population 5 (Figure 4.8) finds its highest concentration in the southern-East part of Switzerland. It mostly concerns canton GR and gradually disperses in all directions reaching cantons UR, LU, AG, ZH, SH and TG.

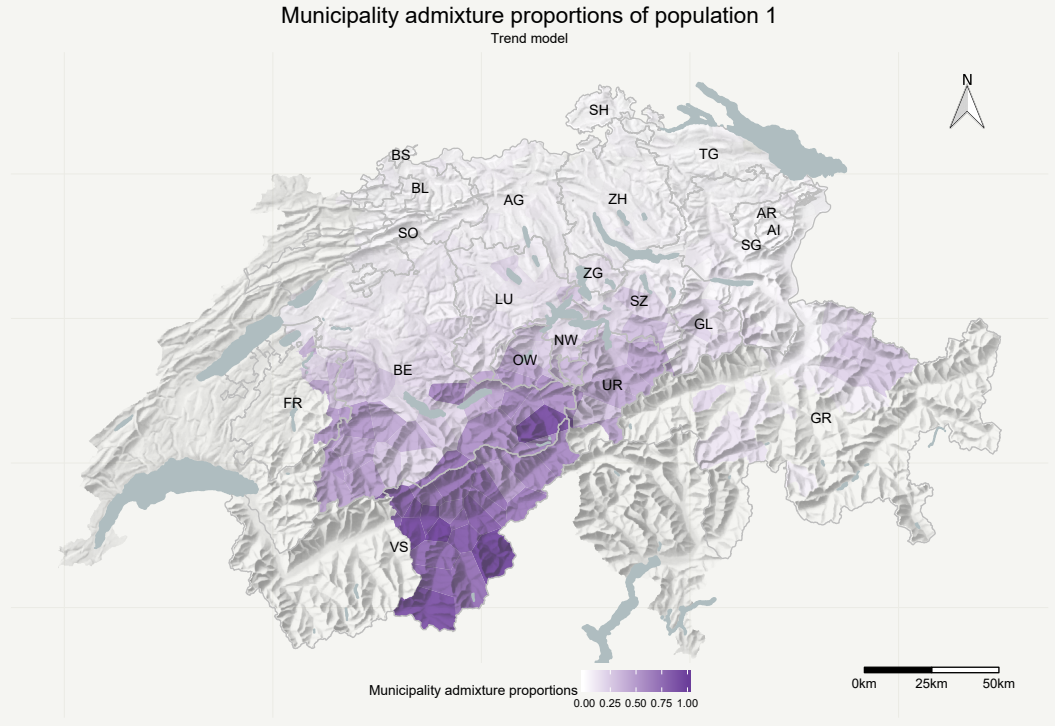


Figure 4.4: Spatial distribution of ancestral population 1

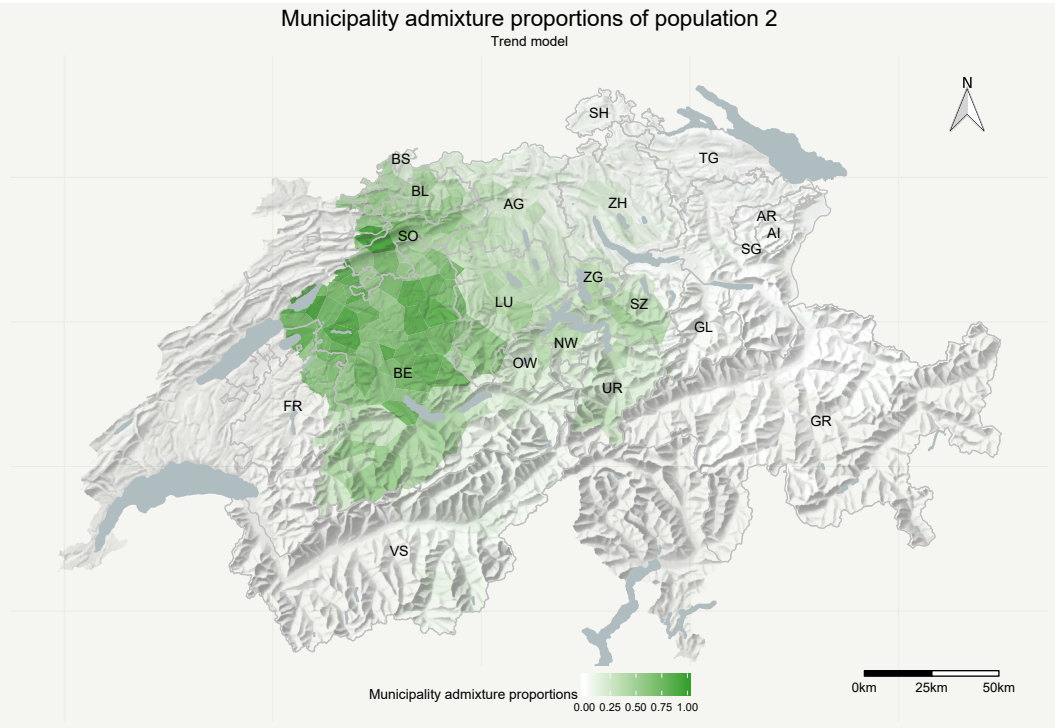


Figure 4.5: Spatial distribution of ancestral population 2

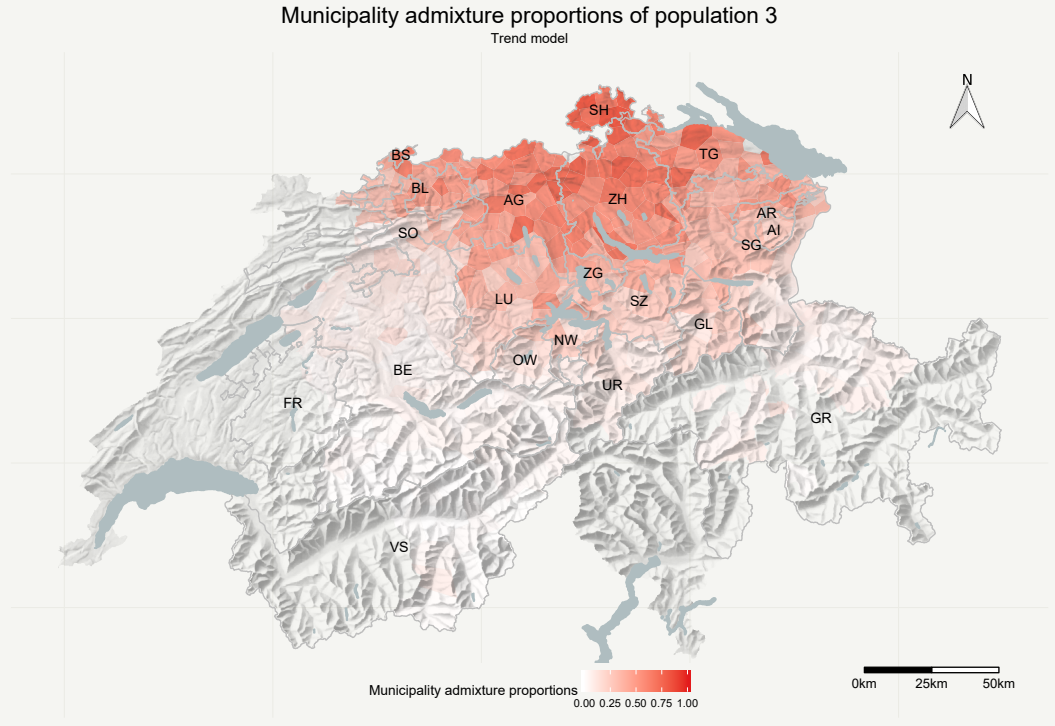


Figure 4.6: Spatial distribution of ancestral population 3

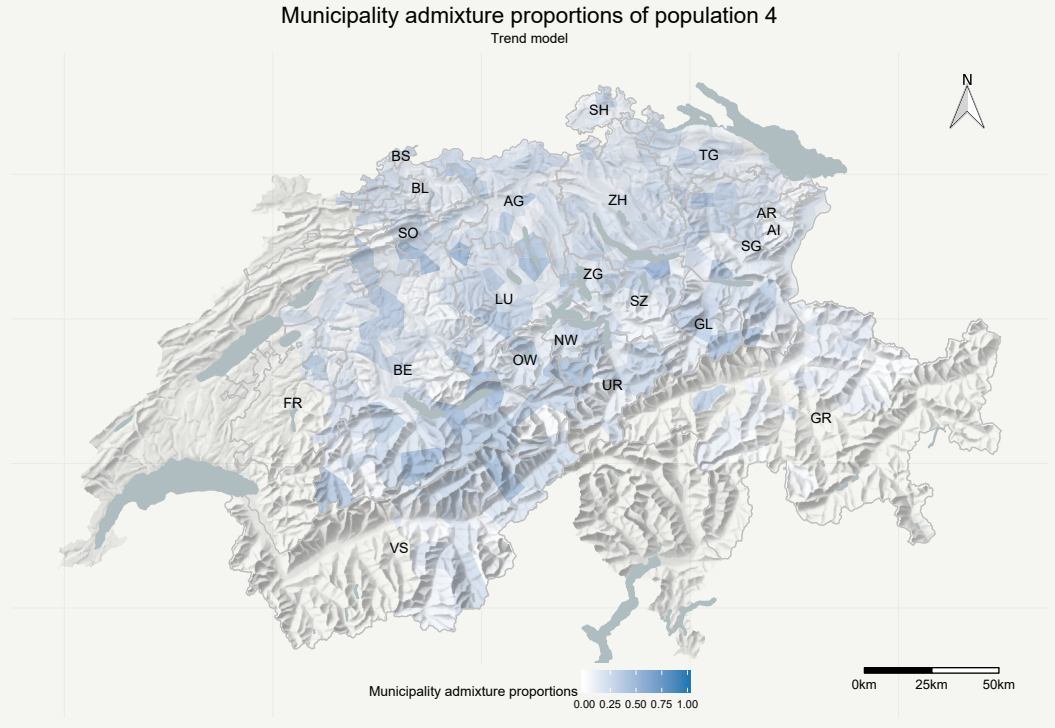


Figure 4.7: Spatial distribution of ancestral population 4

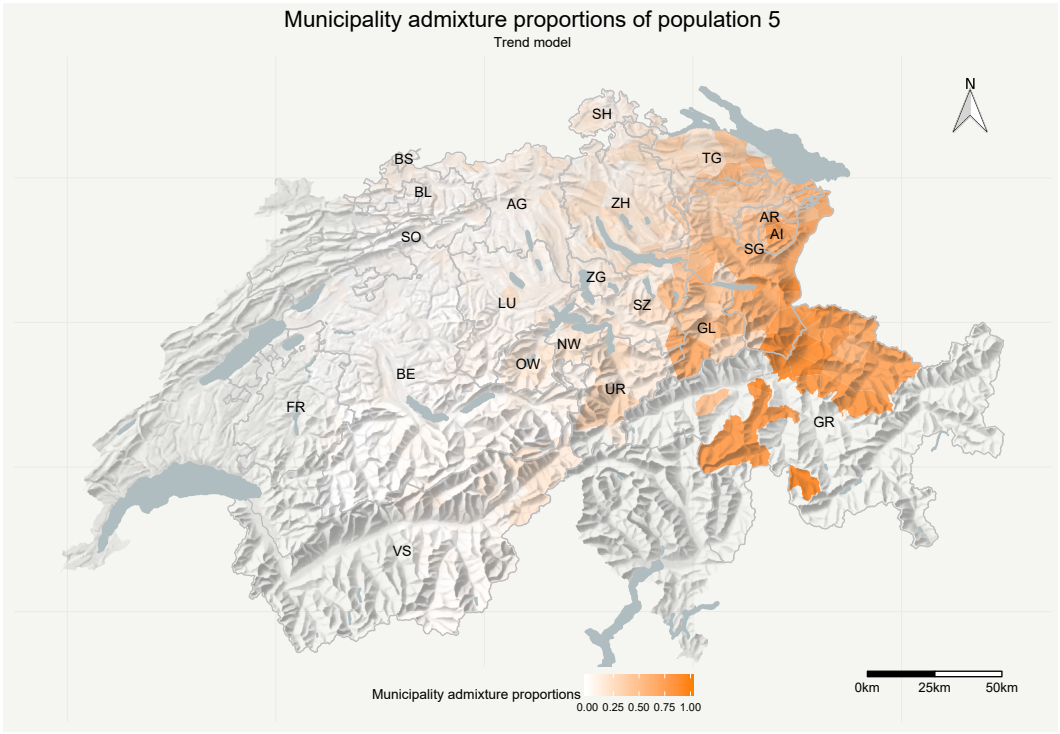


Figure 4.8: Spatial distribution of ancestral population 5

4.2.2 K_{max}=6, Non-spatial model

For the non-spatial model, only the new populations detected are presented. The other maps can be found in the appendix D. Ancestral population 3 of the trend model (Figure 4.6) is no longer detected. Hence, with $K_{max} = 6$, two more populations are detected and presented hereafter.

The first additional population detected in the non-spatial model is mostly situated in the central part of Switzerland (Figure 4.9). The highest admixture proportions are found in municipalities of canton LU, AG, NW, ZG and SZ. However, we can find lower admixture proportions of this population nearly in all municipalities.

The second population involves the North-Eastern part of Switzerland (Figure 4.10), mostly cantons AR, AI, TG, SH and ZH. The highest admixture proportions are found in AR, AI and TG. Furthermore, admixture proportions gradually decrease from East to West.

4.3 Spatial trend

With the aim of quantifying the global spatial effect, the β coefficients and their relative CI of the trend surfaces are presented in table 4.1. Positive $\beta_{long,k}$ represent West to East

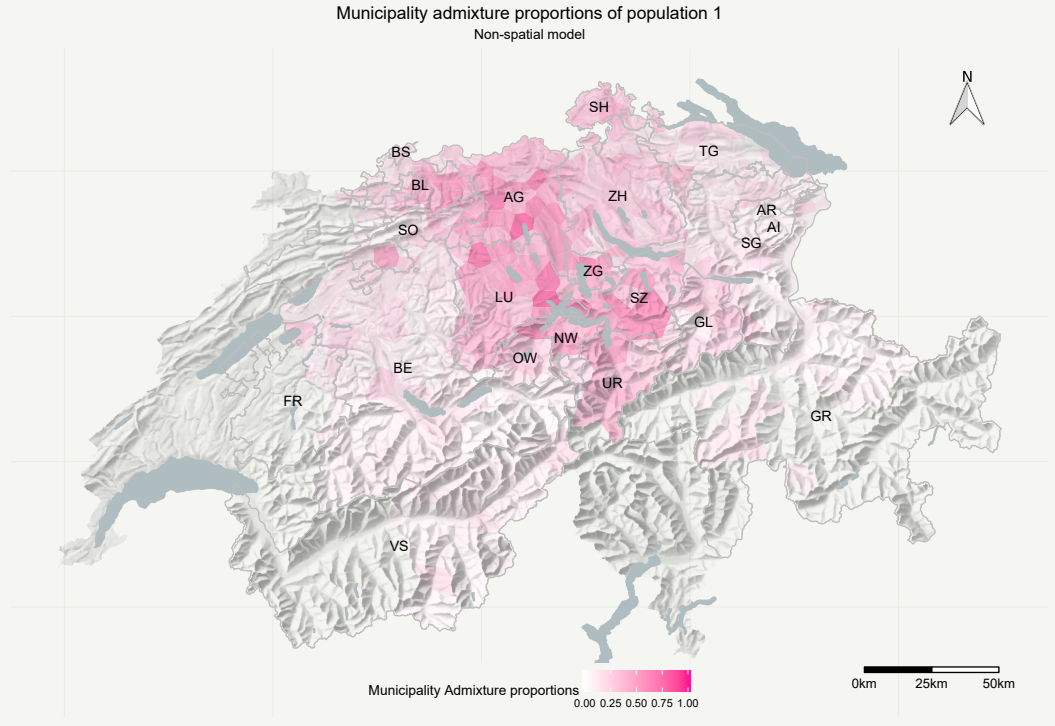


Figure 4.9: Non-spatial model: spatial distribution of ancestral population 1

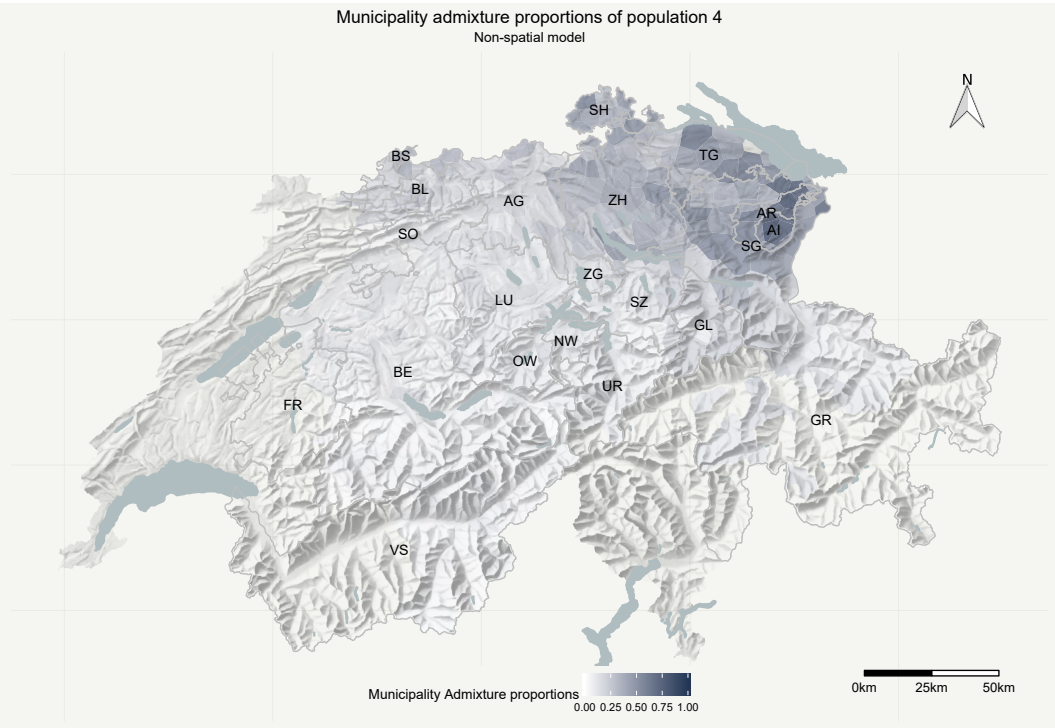


Figure 4.10: Non-spatial model: spatial distribution of ancestral population 4

trends, namely that the admixture proportions increase eastwards and vice-versa. Positive $\beta_{lat,k}$ represent South to North trends, meaning that the admixture proportions increase northwards and vice-versa.

	β_{long}	β_{lat}	$\beta_{long}CI_{95\%}$	$\beta_{lat}CI_{95\%}$
Population 1	-0.16	-5.39	[-0.83, 0.06]	[-5.84, -4.54]
Population 2	-2.83	0.27	[-3.05, -1.97]	[-0.05, 0.92]
Population 3	0.42	4.43	[-0.07, 0.66]	[3.1, 4.86]
Population 4	-0.21	0.12	[-0.41, 0.30]	[-0.53, 0.67]
Population 5	3.17	-1.20	[2.79, 3.38]	[-1.60, -0.58]

Table 4.1: β estimates and relative CI

Population 1 shows only a significant southwards latitudinal trend (CI of $\beta_{long,1}$ includes 0). Contrarily to population 1, population 2 admixture proportions can be modeled with a westwards longitudinal trend surface rather than a latitudinal one. Population 3 is in line with population 1 but in the opposite direction; in fact, its latitudinal trend goes from South to North. Population 4 does not show any spatial trend. Both β estimates are near 0 and their relative CI intervals include both 0. Ancestral population 5 shows significant latitudinal and longitudinal trends, even if $\beta_{lat,5}$ is close to 0.

4.4 Shannon diversity

Figure 4.11 shows the dialectal diversity map obtained by performing the Shannon diversity index for each spatial unit. The most diverse dialectal varieties are located in the central part, including cantons UR, NW, SZ, OW and GL. The least diverse spatial units are found in canton VS, GR, the northern part of BE and SH, all located to the extremities of the country. The other spatial units seem to have similar mixing proportions.

4.5 Color interpolation

Figure 4.12 shows the color interpolation approach results. The results are read from South West in the clockwise direction. The central part of Switzerland is described last.

Population 1 is dominant in the South-West part. The transition from population 1 to population 2 is generally discrete, except in the southern-West part of canton BE, where no evident dominant population can be identified. Going northwards, a transition area is found between population 2 and population 3, located in canton BL and the western part of canton AG. Canton ZH and SH present similar colors and already contain some population 5 residual. Canton TG, AR and the northern part of SG are located in a transition area between population 3

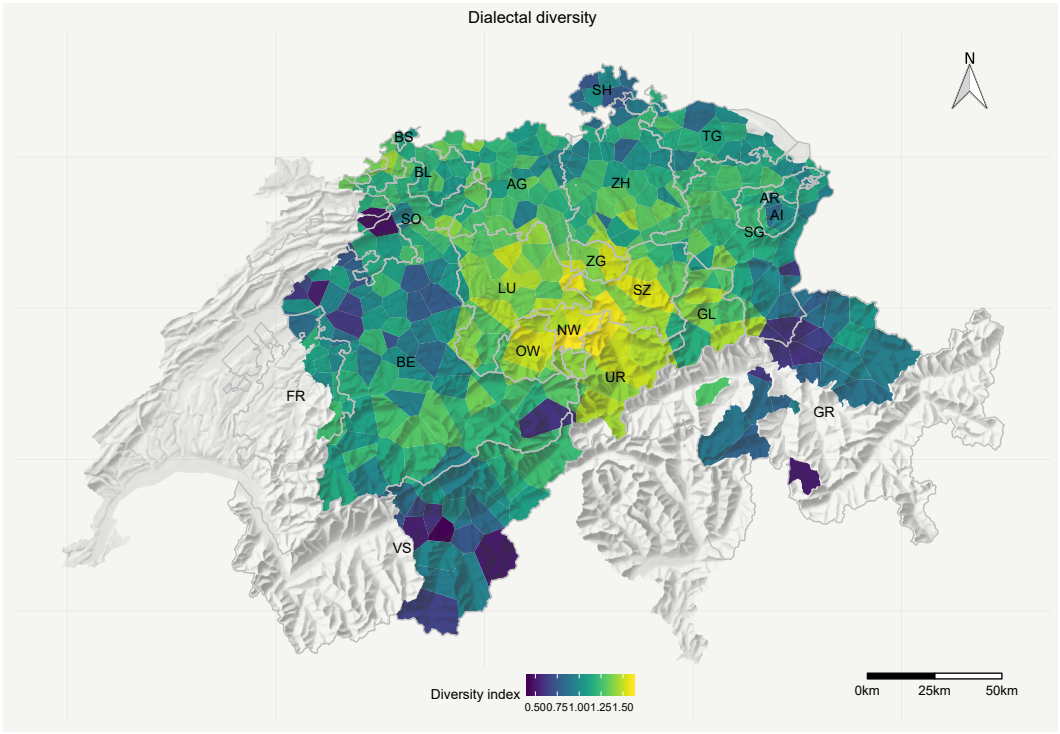


Figure 4.11: Dialectal diversity

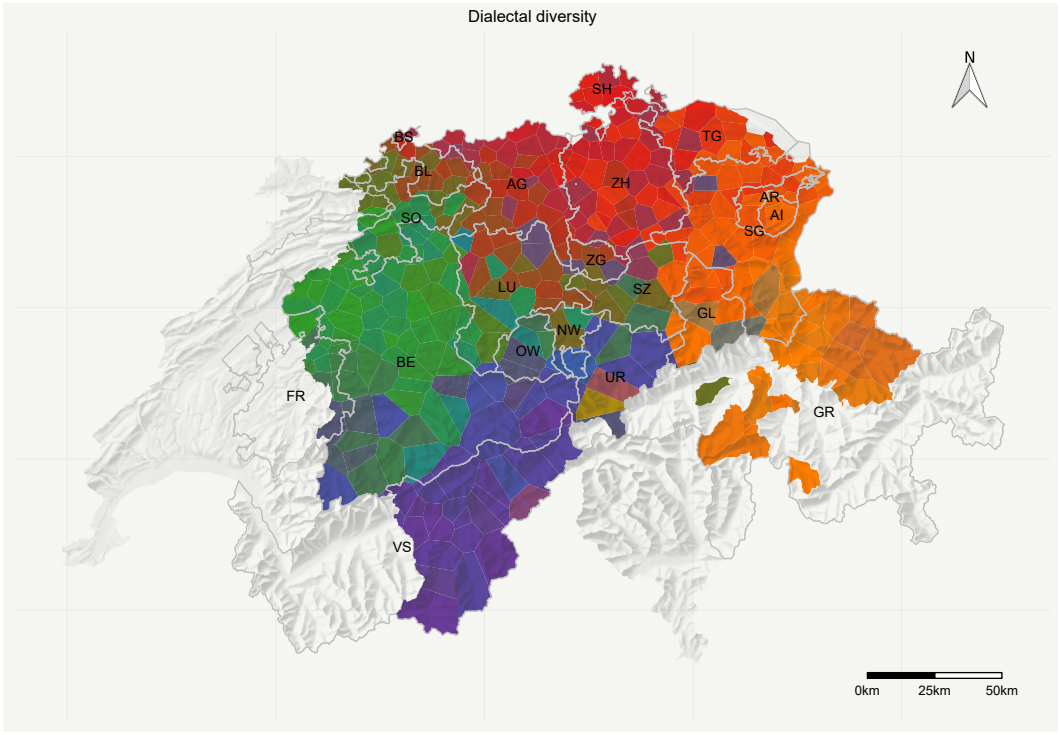


Figure 4.12: Color interpolation

and population 5. GR canton belongs mostly to population 5, even if we can detect some other ancestral population traces that cannot be clearly identified. GL seems to have similar dominant populations such as AR and northern part of SG, except for some municipalities which seem to belong to population 4 and population 2 partly. The central-southern part of Swiss German municipalities including cantons UR, NW, SZ and SZ do not show any evident dominant population covering the whole cantons.

Chapter 5

Discussion

The results gave several insights on the spatial variation of morphosyntax. First of all, the choice of the number of ancestral dialectal families at the origin of Swiss German morphosyntax will be discussed since further analysis depends on this choice. Secondly, the geographic distribution of these ancestral populations will be analyzed since it is the primary goal of this work. Moreover, the power of methods from evolutionary biology to draw insights into historical contact due to ancient migrations is evaluated. Furthermore, the evaluation of inter-population transitions is discussed in order to evaluate if the dialect continuum framework is well adapted for Swiss German morphosyntax. In fine, the limitations of TESS model will be examined.

5.1 Number of ancestral populations and models performance

The comparison of different spatial models allowed us to choose the model that best fits the SADS data and the suitable number of ancestral populations. DICs values suggest that the suitable number of ancestral populations is $K_{max} = 5$ for spatial models (trend and full-trend) and $K_{max} = 6$ for the non-spatial model (Figure 4.1). The fact that the non-spatial model performs better with higher K_{max} is not surprising; in fact, the more ancestral populations are identified, the more the spatial structure is fragmented and it is less likely that admixture proportions can be modeled with a proper spatial model. The spatial variation of STRUCTURE results confirms the results obtained with the non-spatial model (Section E) and strengthens the general model robustness.

The full-trend model outperforms the other models only for $K_{max} = 4$, which is not a suitable number of ancestral populations given that the plateau is reached with $K_{max} = 5$. This means that spatial autocorrelation is not necessary for the model with $K_{max} = 5$. Spatial autocorrelation would be useful to account for short-range isolation-by-distance scenarios [25]; given the spatial structure of the dialectal families, we are more in a long-range isolation-by-

distance scenario and it may be the reason why spatial autocorrelation does not have a significant effect on the model performances.

In any case, all the models detect an evident spatial structure of the ancestral dialectal families and some of them are detected with both K_{max} (see appendices D,C).

5.2 Spatial structure of ancestral dialectal families

Although the geographic distribution of morphosyntax is generally considered less salient than other linguistic levels, most of the ancestral dialectal families show an evident spatial structure. Ancestral populations cores can be identified for most of the dialectal families, and their structure is in line with the wave model explained in section 2.5, which affirms that a language trait gradually spreads from a central point in its neighborhood. In this case we can identify this phenomenon at an aggregate level; in fact, population 1, 2, 3 and 5 (Figures 4.4, 4.5, 4.6 and 4.8) show clear central points from which they disperse in a diffusional process. This could be hypothetically interpreted as the geographical origin of these dialectal families. The cores of these ancestral populations are located at the extremities and the diversity analysis (Figure 4.11) confirms this organization; in fact, the least diverse municipalities correspond to the same spatial units representing the core of the ancestral populations. Furthermore, the β coefficients (Table 4.1) representing the spatial trends of the admixture proportions are always oriented in the direction of the cores: southwards for population 1, westwards for population 2, northwards for population 3, and eastwards for population 5.

Since the cores of the cluster are located at the extremities of the study area, the admixture proportions of each population gradually decrease towards the center. Hence, this gradual decreasing towards the center implies that municipalities have an increase of diversity in the same direction (Figure 4.11). This implies an arduous interpretation of the color interpolation map (Figure 4.12) since no evident dominant population can be identified in the center. Indeed, the color interpolation method does not seem to be well adapted for this kind of spatial organization since highly diverse municipalities do not have two evident dominant populations and the re-calculation of the proportion becomes meaningless. A manner to tackle this problem would be to take into account this diversity in the calculus of the new proportions; for instance by creating new color categories representing the equal mixing of several dialectal families. Furthermore, ancestral population 1 in the non-spatial model (Figure 4.9) covers the same area of the highly diverse municipalities in figure 4.11 confirming then the bad performance of the non-spatial model in detecting clines.

Population 4 of the trend model (Figure 4.7) does not show any particular spatial trend. Different cores are detected, but they assume a random spatial organization, and as expected,

the β coefficients (Table 4.1) of this population do not show any evident spatial trend, neither latitudinal nor longitudinal. This population could be interpreted as the ground Swiss German since a large part of the study area seems to originate from this dialectal family.

We find an evident spatial structure and the cores of the ancestral families spread gradually from a central point in their neighborhood. This could be signal of a real diffusional process occurring when dialectal varieties come into contact and gives insights on the historical meaning of this spatial structure.

5.3 Morphosyntactic traits as heritable units

The application of spatial population genetics in dialectometry was not only assessed to quantify the Swiss German morphosyntactic continuum. The assumption of treating dialect traits as biological characters is not only a methodological constraint; it underlies that language traits are heritable units such as biological characters [94]. Population 1 of the trend model (Figure 4.4) demonstrates this parallel; in fact, as mentioned in section 2.7, one the most important migrations concerning the Swiss population is the Walser migration. The spatial pattern of the Walser migration (Figure 2.3b) correspond almost exactly to the spatial organization of population 1 (Figure 4.4). First of all, if we interpret the core of this cluster as the ancestral population origin point, this could be therefore the origin of the Walser population. From this point, we can almost detect the migration corridor reaching municipalities of the cantons UR and GR. However, this ancestral population spreads northwards as well, reaching municipalities of cantons BE, OW, SZ and FR. This could confirm the diffusional processes concerning dialects in contact, due to both migration and geographical adjacency. This cluster might finally represent the Highest Alemannics division (Figure 2.3a), since it concerns the cantons VS, GR, BE and FR as well.

In the same way, population 3 (Figure 4.6) could correspond to the High Alemannics dialectal family (Figure 2.3a), that concerns the northern part of Switzerland. However, this dialectal division would concern also canton BE, which is not entirely included in population 3. This could due to the fact that canton BE has been one of the most influential cantons in history [50,81], and the algorithm detects canton BE as a core dialectal family (Figure 4.5).

5.4 Clines of Swiss German morphosyntax

The color interpolation map (Figure 4.12) was used to detect geographical clines between two specific populations. Indeed, if we start with population 1 and progress in the clockwise direction, we can find different types of transitions from one population into another. First of all, a clinal variation is found in the southern-west part of canton BE between popula-

tion 1 (Figure 4.4) and population 2 (Figure 4.5). The stacked barplot ordered by latitude (Figure 4.3) confirms partly what the color interpolation approach finds. There is a gradual variation of population 1 in the latitudinal direction, but both population 2 (Figure 4.5) and population 5 (Figure 4.8) are found. This is due to the fact that the stacked barplot gives a singular dimension, and if a cline occurs in both latitudinal and longitudinal direction, this visualization is not well adapted.

Canton BL is located in a transition area between population 2 and population 3. The stacked barplot ordered by longitude (Figure 4.2) confirms the presence of this cline. In fact, we can note a general gradual decrease of population 2 and a gradual increase of population 3 in the longitudinal direction. The most striking result is between population 3 and population 5. Both color interpolation map (Figure 4.12) and stacked barplot ordered by longitude (Figure 4.2) show an evident geographical cline; in fact, thanks to the color map interpolation we can find a clear gradual variation from red to orange colors in the northeast part of Switzerland. This is found in the right part of the stacked barplot ordered by longitude (Figure 4.2) where there is a gradual decrease of population 2 against a gradual increase of population 5. This transition area resembles to the population 4 detected by the non-spatial model (Figure 4.10). This confirms that spatial Bayesian clustering is a more adapted method to detect clines; indeed, when the cline is detected ($K_{max} = 4$), the trend model does not improve significantly (Figure 4.1). In contrast, the non-spatial model does not detect this gradual transition and performs better by classifying this transition in another ancestral dialectal family. The same applies to ancestral population 1 of the non-spatial model (Figure 4.9); it detects the mixing of several dialectal families as a single population. Clines have been detected and the dialect continuum framework seems to be well adapted to define the dialectal boundaries of Swiss German morphosyntax.

5.5 Limitations of the model

For a matter of time and hardware disposal, it was not possible to compute further runs, but one should run at least 100 runs per K_{max} to have more robust Bayesian estimates [24]. However, all the runs performed in this study converged to the same parameter estimates, and we expect that further runs would strengthen these conclusions. Other population genetics methods exist to infer population structure and are considerably faster than TESS and STRUCTURE. TESS has been chosen because of its similarity to STRUCTURE, which is the most acknowledged algorithm to infer population structure. Furthermore, the comparison of the models is straightforward. However, it should be pertinent to test other algorithms than Bayesian clustering to see whether we yield the same results.

Furthermore, all conclusions that can be drawn from the results depend directly on the choice

of the number of populations. Even if the work-flow of the choice of the number of populations is statistically rigorous, the DIC sometimes does not lead to the best choice [9]. Hence, further runs or having a more detailed historical background would be beneficial to assess the model performances.

Chapter 6

Conclusion

We used the spatial Bayesian clustering algorithm TESS to assess the structure of Swiss German morphosyntax and infer likely ancestral populations that gave rise to these dialects' spatial configuration. TESS, primarily developed for population genetics analysis, has been for the first time used to infer population structure from only linguistic data as genetic evidence.

Several models have been tested by changing the prior information while modeling the admixture proportions of each municipality. This prior information relies on the modeling of the admixture proportions in space with the use of trend surface analysis. Three models have been compared: one assuming no spatial dependence (non-spatial model), one integrating the global effect of geography (trend model) and one taking into account both global and local effects with the integration of a spatial autocorrelation term (full-trend model).

Most of the inferred populations show a clear spatial structure. The cores of these populations are located at the extremities of the Swiss German part of Switzerland, and their admixture proportions gradually vary towards the center. Spatial clines have been detected in the northwest and northeast part of Switzerland and dialect continuum seems to be the most adapted spatial organization of Swiss German morphosyntax. Furthermore, the Walser migration patterns have been discerned, and morphosyntactic traits could be therefore considered a good signal of historical relatedness.

What is the total number of morphosyntactic families at the origin of Swiss German dialects?

The spatial models and the non-spatial model predict a different number of morphosyntactic families. However, the spatial models detect spatial clines and with bigger K_{max} there is no significant improvement. The non-spatial model does not detect spatial clines and classifies

these clines into distinct populations. Hence, $K_{max} = 5$ is the most suitable number of morphosyntactic families.

Is geography an important factor in explaining Swiss German morphosyntactic variation?

The trend model best fits the SADS data, and since this model predicts the admixture proportions by taking into account the global effect of space, geography is an important factor to explain Swiss German morphosyntactic variation. Furthermore, an evident spatial structure is detected.

How intra-population admixture proportions vary across space?

Except population 4, the other populations show a clear spatial structure. Core populations have been detected and their admixture proportions gradually diffuse from the cores to their neighborhood.

Are inter-population transitions characterized by discrete or gradual changes?

Generally, gradual changes are found between the ancestral dialectal families. The clearest spatial cline has been detected in the northwest part of Switzerland between the northern population and the southeast population. Another cline is found in the northwest part of Switzerland between Bern population and northern population. We found less gradual changes between the Walser population and the Bern population. Overall, we can conclude that the dialect continuum framework is well suited to model the Swiss German morphosyntax.

Chapter 7

Perspectives

The results obtained with Tess give new insights on the spatial variation of morphosyntax and can be the basis of further analysis.

First of all, even if the spatial information has been integrated into the model, no topographic information has been taken into account. Therefore, it could be interesting to integrate some natural barriers such as rivers, lakes or mountains. This is possible to perform in Tess by changing the spatial network connecting the sample sites. Moreover, it could be a more realistic spatial framework given the Swiss topography. Furthermore, in this scenario, it should be pertinent to see whether full-trend model performs better than the other models.

The dialectal diversity of Swiss German municipalities (Figure 4.11) shows an evident spatial structure that strengthens the hypothesis of language transfer between dialects; in fact, the centrality of the most diverse municipalities could be the consequence of high contact due to their central location. Hence, it would be pertinent to assess centrality and betweenness analysis specific to graph theories, in order to assess the relationship between these values and the dialectal diversity. Since pairwise travel time distances dating back to 1850 are available for Switzerland [33], one could compare the relation results between dialectal diversity and connectivity measures for different time periods.

To strengthen the conclusion about the consideration of morphosyntactic traits as heritable units, one could perform the same analysis with real genetic data and compare the inferred populations in both scenarios. Another interesting manner to combine these two types of input data could be the use of both biological and linguistic traits in the same analysis.

Given the computational expensiveness of Tess and its struggle to deal with large datasets, it would be interesting to infer population structure with fewer individuals per sample site. This could be useful to understand the stability of the inferred dialectal families.

Bibliography

[1] Quentin D Atkinson and Russell D Gray. Curious parallels and curious connections—phylogenetic thinking in biology and historical linguistics. *Systematic biology*, 54(4):513–526, 2005.

[2] Peter Auer and Jürgen Erich Schmidt. *Theories and methods*, volume 30. Walter de Gruyter, 2010.

[3] David J Balding and Richard A Nichols. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica*, 96(1-2):3–12, 1995.

[4] Sjef Barbiers, Hans Bennis, and Gunther De Vogelaer. *Syntactic Atlas of the Dutch Dialects (Sand): Pronouns, Agreement and Dependencies*, volume 1. Amsterdam University Press, 2005.

[5] Gabriela Bart, Robert Weibel, Pius Sibler, and Elvira Glaser. Analysis of swiss german syntactic variants using spatial statistics. *Alvarez Pérez, Xosé Afonso, Ernestina Carrilho & Catarina Magro (red.). Current Approaches to Limits and Areas in Dialectology. Newcastle upon Tyne: Cambridge Scholars Publishing*, 2013.

[6] Mark A Beaumont and Bruce Rannala. The bayesian revolution in genetics. *Nature Reviews Genetics*, 5(4):251, 2004.

[7] Julian Besag. Statistical analysis of non-lattice data. *The statistician*, pages 179–195, 1975.

[8] Claire Bowern. The riddle of tasmanian languages. *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb20121842, 2012.

[9] SP Brooks, Jim Smith, Aki Vehtari, Martyn Plummer, Mervyn Stone, Christian P Robert, DM Titterington, JA Nelder, Anthony Atkinson, AP Dawid, et al. Discussion on the paper by spiegelhalter, best, carlin and van der linde. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 64(4):616–639, 2002.

[10] Claudia Bucheli and Elvira Glaser. The syntactic atlas of swiss german dialects: empirical and methodological problems. *Syntactic microvariation*, 2:41–73, 2002.

[11] Luca L Cavalli-Sforza. The basque population and ancient migrations in europe. *Munibe*, 6:129–137, 1988.

[12] Jack K Chambers and Peter Trudgill. *Dialectology*. Cambridge University Press, 1998.

[13] Chibiao Chen, Eric Durand, Florence Forbes, and Olivier François. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Resources*, 7(5):747–756, 2007.

[14] Wikimedia Commons. File:alemannic-dialects-map-english.png — wikimedia commons, the free media repository, 2014. [Online; accessed 6-August-2018].

[15] Jukka Corander, Jukka Sirén, and Elja Arjas. Bayesian spatial modeling of genetic population structure. *Computational Statistics*, 23(1):111–129, 2008.

[16] Jukka Corander and Jing Tang. Bayesian analysis of population structure based on linked molecular information. *Mathematical biosciences*, 205(1):19–31, 2007.

[17] H Cramer. Mathematical methods of statistics (princeton university, princeton, 1946). *Google Scholar*, page 118, 1989.

[18] Noel Cressie. Statistics for spatial data. *Terra Nova*, 4(5):613–617, 1992.

[19] William Croft. *Explaining language change: An evolutionary approach*. Pearson Education, 2000.

[20] Charles Darwin. On the origin of species by means of natural selection. *Murray, London*, 1859.

[21] Carsten F Dormann. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global ecology and biogeography*, 16(2):129–138, 2007.

[22] Michael Dunn. Language phylogenies. *The Routledge handbook of historical linguistics*, pages 190–211, 2015.

[23] Michael Dunn, Stephen C Levinson, Eva Lindström, Ger Reesink, and Angela Terrill. Structural phylogeny in historical linguistics: methodological explorations applied in island melanesia. *Language*, 84(4):710–759, 2008.

[24] E Durand, C Chen, and O François. Tess version 2.3 reference manual. *Available at: memberstimc. imag. fr/Olivier. Francois/tess. html*, 2009.

[25] Eric Durand, Flora Jay, Oscar E Gaggiotti, and Olivier François. Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution*, 26(9):1963–1973, 2009.

[26] Sheila Embleton. Multidimensional scaling as a dialectometrical technique: Outline of a research project. In *Contributions to quantitative linguistics*, pages 267–276. Springer, 1993.

[27] Ethnologue. Languages of the world. https://www.ethnologue.com/statistics. Accessed: 2018-08-15.

[28] Daniel Falush, Matthew Stephens, and Jonathan K Pritchard. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4):1567–1587, 2003.

[29] Arthur Fibicher. *Walliser Geschichte*. Kantonales Erziehungsdepartement, 1983.

[30] Alexandre François. Trees, waves and linkages. *The Routledge Handbook of Historical Linguistics. London: Routledge*, pages 161–189, 2015.

[31] Alexandre François. Méthode comparative et chaînages linguistiques: Pour un modèle diffusionniste en généalogie des langues, 2017.

[32] Olivier François, Sophie Ancelet, and Gilles Guillot. Bayesian clustering using hidden markov random fields in spatial population genetics. *Genetics*, 174(2):805–816, 2006.

[33] Philipp Fröhlich, Thomas Frey, Serge Reubi, and Hans Ulrich Schiedt. Entwicklung des transitverkehrs-systems und deren auswirkung auf die raumnutzung in der schweiz (cost 340): Verkehrsnetz-datenbank. *Arbeitsberichte Verkehrs-und Raumplanung*, 208, 2005.

[34] Hong Gao, Scott Williamson, and Carlos D Bustamante. An mcmc approach for joint inference of population structure and inbreeding rates from multi-locus genotype data. *Genetics*, 2007.

[35] Elvira Glaser. Area formation in morphosyntax. *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives*, pages 195–221, 2013.

[36] Hans Goebl. *Dialektometrie*, volume 157. Verlag der Österreichischen Akademie der Wissenschaften, 1982.

[37] Hans Goebl. Le laboratoire de dialectométrie de l'université de salzbourg. un bref rapport de recherche. *Zeitschrift für französische Sprache und Literatur*, pages 35–55, 2008.

[38] Hans Goebl. Dialectometry and quantitative mapping. *Language and space. An international handbook of linguistic variation*, 2:433–457, 2010.

[39] Hans Goebl, Karl Jaberg, and Jules Gilliéron. *Dialektometrische Studien: Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, volume 2. M. Niemeyer, 1984.

[40] Charlotte Gooskens and Wilbert Heeringa. Perceptive evaluation of levenshtein dialect distance measurements using norwegian dialect data. *Language variation and change*, 16(3):189–207, 2004.

[41] Russell D Gray, David Bryant, and Simon J Greenhill. On the shape and fabric of human history. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1559):3923–3933, 2010.

[42] Jack Grieve, Dirk Speelman, and Dirk Geeraerts. A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change*, 23(2):193–221, 2011.

[43] Gilles Guillot, Frédéric Mortier, and Arnaud Estoup. Geneland: a computer package for landscape genetics. *Molecular ecology notes*, 5(3):712–715, 2005.

[44] Richard W Hamming. Error detecting and error correcting codes. *Bell System technical journal*, 29(2):147–160, 1950.

[45] Daniel L Hartl and Andrew Clark. *Principles of population genetics*, volume 116. Sinauer associates Sunderland, 1997.

[46] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.

[47] Wilbert Heeringa and John Nerbonne. Dialect areas and dialect continua. *Language Variation and Change*, 13(3):375–400, 2001.

[48] Wilbert Jan Heeringa. *Measuring dialect pronunciation differences using Levenshtein distance*. PhD thesis, Citeseer, 2004.

[49] Clive J Hoggart, MARK D Shriver, Rick A Kittles, David G Clayton, and Paul M McKeigue. Design and analysis of admixture mapping studies. *The American Journal of Human Genetics*, 74(5):965–978, 2004.

[50] Rudolf Hotzenköcherle, Niklaus Bigler, Robert Schläpfer, and Rolf Börlin. *Die Sprachlandschaften der deutschen Schweiz*, volume 1. Sauerländer, 1984.

[51] Mattias Jakobsson and Noah A Rosenberg. Clumpp: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14):1801–1806, 2007.

[52] Péter Jeszenszky and Robert Weibel. Measuring boundaries in the dialect continuum. In *Proceedings of the AGILE*, 2015.

[53] Péter Jeszenszky and Robert Weibel. Modeling transitions between syntactic variants in the dialect continuum. In *The 19th AGILE International Conference on Geographic Information Science, Helsinki (Finnland), 14 June 2016-17 June 2016*, 2016.

[54] Mark Jobling, Matthew Hurles, and Chris Tyler-Smith. Human evolutionary genetics: origins, peoples & disease. pages 443–444, 2013.

[55] Brett Kessler. Computational dialectology in irish gaelic. In *Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics*, pages 60–66. Morgan Kaufmann Publishers Inc., 1995.

[56] Bernd Kortmann. Areal variation in syntax. *Language and Space: An International Handbook of Linguistic Variation. Berlin/New York: Walter de Gruyter*, pages 837–864, 2010.

[57] Vladimir I Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710, 1966.

[58] Jeremy W Lichstein, Theodore R Simons, Susan A Shriner, and Kathleen E Franzreb. Spatial autocorrelation and autoregressive models in ecology. *Ecological monographs*, 72(3):445–463, 2002.

[59] Giuseppe Longobardi and Cristina Guardiano. Evidence for syntax as a signal of historical relatedness. *Lingua*, 119(11):1679–1706, 2009.

[60] Ernst Mayr. *Populations, species, and evolution: an abridgment of animal species and evolution*, volume 19. Harvard University Press, 1970.

[61] Paolo Menozzi, Alberto Piazza, and L Cavalli-Sforza. Synthetic maps of human gene frequencies in europeans. *Science*, 201(4358):786–792, 1978.

[62] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[63] Matthew W Mitchell, Brian Rowe, Paul R Sesink Clee, and Mary Katherine Gonder. Tess ad-mixer: A novel program for visualizing tess q matrices. *Conservation Genetics Resources*, 5(4):1075–1078, 2013.

[64] Luay Nakhleh, Donald A Ringe, and Tandy Warnow. Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language*, 81(2):382–420, 2005.

[65] John Nerbonne, Wilbert Heeringa, and Peter Kleiweg. Edit distance and dialect proximity. *Time Warps, String Edits and Macromolecules: The theory and practice of sequence comparison*, 15, 1999.

[66] John Nerbonne and Peter Kleiweg. Toward a dialectological yardstick. *Journal of Quantitative Linguistics*, 14(2-3):148–166, 2007.

[67] John Nerbonne and William A. Kretzschmar, Jr. Dialectometry++. *Literary and Linguistic Computing*, 28(1):2–12, 2013.

[68] John Nerbonne and William A Kretzschmar Jr. Dialectometry++. *Literary and Linguistic Computing*, 28(1):2–12, 2013.

[69] Johanna Nichols and Tandy Warnow. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass*, 2(5):760–820, 2008.

[70] OpenStreetMap contributors. Swiss Lakes. http://www.mapcruzin.com/free-switzerland-arcgis-maps-shapefiles.htm, 2018.

[71] Mark Pagel. Human language as a culturally transmitted replicator. *Nature Reviews Genetics*, 10(6):405, 2009.

[72] Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11, 2006.

[73] Jonathan K Pritchard, Matthew Stephens, and Peter Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.

[74] Jelena Prokić and John Nerbonne. Recognising groups among dialects. *International journal of humanities and arts computing*, 2(1-2):153–172, 2008.

[75] Simon Pröll, Simon Pickl, and Aaron Spettl. Latente strukturen in geolinguistischen korpora. *Elmentaler, Michael/Markus Hundt/Jürgen Erich Schmidt (Hg.): Deutsche Dialekte. Konzepte, Probleme, Handlungsfelder. Stuttgart: Steiner*, 2014.

[76] QGIS Development Team. *QGIS Geographic Information System*. Open Source Geospatial Foundation, 2009.

[77] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[78] Ger Reesink, Ruth Singer, and Michael Dunn. Explaining the linguistic diversity of sahul using population models. *PLoS biology*, 7(11):e1000241, 2009.

[79] Brian D Ripley. *Statistical inference for spatial processes*. Cambridge university press, 1991.

[80] Brian D Ripley. *Pattern recognition and neural networks*. Cambridge university press, 2007.

[81] Yves Scherrer and Philipp Stoeckle. A quantitative approach to swiss german–dialectometric analyses and comparisons of linguistic levels. *Dialectologia et Geolinguistica*, 24(1):92–125, 2016.

[82] August Schleicher. *Die Darwinsche theorie und die sprachwissenschaft: Offenes sendschreiben an herrn Ernst Häckel*, volume 2. Böhlau, 1871.

[83] Johannes Schmidt. *Die verwantschaftsverhältnisse der indogermanischen sprachen*. Böhlau, 1872.

[84] Jean Séguy. *La dialectométrie dans l'Atlas linguistique de la Gascogne*. Société de linguistique romane, 1973.

[85] Guido Seiler. On three types of dialect variation and their implications for linguistic theory. evidence from verb clusters in swiss german dialects. *Dialectology meets Typology. Dialect Grammar from a cross-linguistic Perspective*, pages 367–399, 2004.

[86] Robert G Shackleton Jr. English-american speech relationships: A quantitative approach. *Journal of English Linguistics*, 33(2):99–160, 2005.

[87] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[88] Pius Sibler. *Visualisierung und geostatistische Analyse mit Daten des Syntaktischen Atlas der Deutschen Schweiz (SADS)*. PhD thesis, Geographisches Institut der Universität Zürich, 2011.

[89] Pius Sibler, Robert Weibel, Elvira Glaser, and Gabriela Bart. Cartographic visualization in support of dialectology. *Proceeding AutoCarto 2012*, 2012.

[90] David J Spiegelhalter, Nicola G Best, Bradley P Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

[91] Marco René Spruit et al. *Quantitative perspectives on syntactic variation in Dutch dialects*, volume 174. LOT, 2008.

[92] Swiss Federal office of Topography. Canton borders . https://shop.swisstopo.admin.ch/fr/products/landscape/boundaries3D, 2018.

[93] Swiss Federal office of Topography. Digital height map . https://shop.swisstopo.admin.ch/en/products/height_models/dhm25200, 2018.

[94] Kaj Syrjänen, Terhi Honkola, Jyri Lehtinen, Antti Leino, and Outi Vesakoski. Applying population genetic approaches within languages. *Language Dynamics and Change*, 6(2):235–283, 2016.

[95] Benedikt Szmrecsanyi. Corpus-based dialectometry: a methodological sketch. *Corpora*, 6(1):45–76, 2011.

[96] Benedikt Szmrecsanyi. Methods and objectives in contemporary dialectology. *Contemporary Approaches to Dialectology: The Area of North, Northwest Russian and Belarusian Vernaculars*, 2014.

[97] Sarah G Thomason. Contact explanations in linguistics. *The handbook of language contact*, pages 31–47, 2010.

[98] Sarah Grey Thomason and Terrence Kaufman. *Language contact*. Citeseer, 2001.

[99] Waldo R Tobler. A computer movie simulating urban growth in the detroit region. *Economic geography*, 46(sup1):234–240, 1970.

[100] Rudolf Trüb. *Der Sprachatlas der deutschen Schweiz (SDS): ein Grossatlas für einen Kleinraum*. Niemeyer, 1989.

[101] Peter Trudgill. Dialects in contact. pages 86–97, 1986.

[102] Esteve Valls, John Nerbonne, Jelena Prokic, Martijn Wieling, Esteve Clue, and Maria Rosa Lloret. Applying the levenshtein distance to catalan dialects: A brief comparison of two dialectometric approaches. *Verba: anuario galego de filoloxía*, 39, 2012.

[103] P Vounatsou, T Smith, and AE Gelfand. Spatial modelling of multinomial data with latent structure: an application to geographical mapping of human gene and haplotype frequencies. *Biostatistics*, 1(2):177–189, 2000.

[104] William S-Y Wang and James W Minett. Vertical and horizontal transmission in language evolution. *Transactions of the Philological Society*, 103(2):121–146, 2005.

[105] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

[106] Martijn Wieling and John Nerbonne. Advances in dialectometry. 2015.

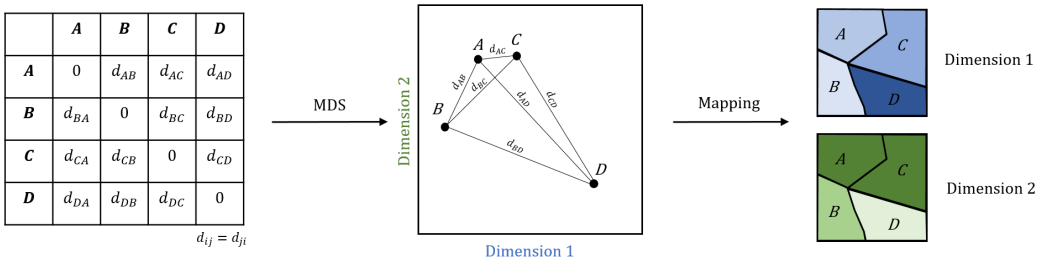
[107] Yu Zhang. Tree-guided bayesian inference of population structures. *Bioinformatics*, 24(7):965–971, 2008.

Appendices

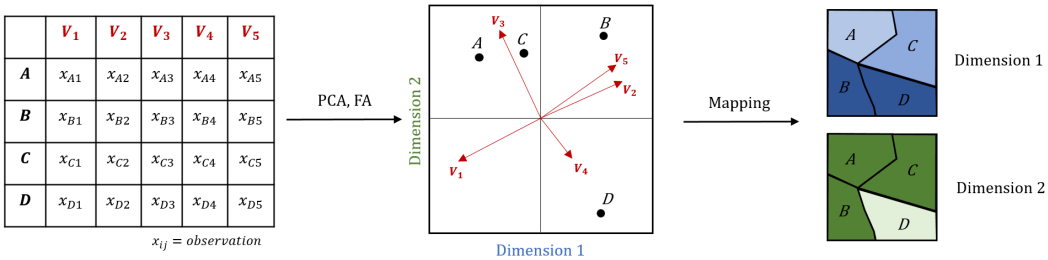
Appendix A

Dimensionality reduction

MDS takes distance matrices as input data and aims to re-project the spatial units in a lower dimension space such that distances are preserved (Figure A.1a). PCA and FA deal with multivariate data and aim to reduce the initial number of variables such that variables in the lower-dimensional space capture best the variation of the dataset (Figure A.1b). Hence, using these methods, one would focus the attention on the units coordinates in the lower-dimensional space[1].



(a) MDS takes distance matrices as input data and aims to re-project the spatial units in an lower dimensional space such that distances are preserved



(b)

Figure A.1: Dimensionality reduction techniques

[1] For instance the projections of the units on the new principal components

Appendix B

Markov chain Monte Carlo (MCMC) methods

Let θ the parameter of our model and $\pi(\theta)$ the target distribution. Each new sample $\theta^{(new)}$ is drawn from a distribution characterized by the previous sample $\theta^{(t-1)}$

$$\theta^{(new)} \sim \pi(\theta^{(t-1)}) \tag{B.1}$$

Several MCMC algorithms have been proposed to deal with the decision of which value of θ is accepted or rejected. TESS model makes use of the Metropolis-Hastings (MH) [46, 62] algorithm and the Gibbs sampler.

B.1 Metropolis-Hastings

MH generates a random number from a specific proposal distribution $g(\theta^{(new)}|\theta^{(t-1)})$, which is the conditional probability of proposing $\theta^{(new)}$ given $\theta^{(t-1)}$. After that, the acceptance probability α, which refers to the probability to accept $\theta^{(new)}$, is calculated as follow

$$\alpha(\theta^{(new)}, \theta^{(t-1)}) = min\left(1, \frac{\pi(\theta^{(new)})}{\pi(\theta^{(t-1)})} \frac{g(\theta^{(t-1)}|\theta^{(new)})}{g(\theta^{(new)}|\theta^{(t-1)})}\right) \tag{B.2}$$

In order to decide whether $\theta^{(new)}$ is accepted or rejected for the next iteration, a uniform random number u that ranges between 0 and 1 is generated and compared with $\alpha(\theta^{(new)}, \theta_{t-1})$. The parameter value used in the next iteration, denoted as $\theta^{(t)}$, will be

$$\theta^{(t)} = \begin{cases} \theta^{(new)}, & \text{if } \alpha(\theta^{(new)}, \theta^{(t-1)}) \geq u \\ \theta^{(t-1)}, & \text{otherwise} \end{cases} \tag{B.3}$$

Each step is repeated until the sample converges, namely that the Markov chain has stationary distribution $\pi(\theta)$. There are two issues with MH. Firstly, the samples are dependent on the starting values. A way to get around this problem is to remove the first part of the sample, called *burn-in period*, which is the time needed for the chain to stabilize. In this study, a burn-in period of 10000 iterations has been chosen. Secondly, since MH deals with a Markov chain process (Equation B.1), the samples may be autocorrelated. Increasing the MCMC sample size, a process called *thinning*, is a way to cope with this problem. Hence, after several tests, a thinning period of 30000 iterations has been chosen.

B.2 Gibbs sampler

A particular case of MH, is Gibbs sampling. This method has been considered a useful tool for clustering issues [73], due to its facility in dealing with interdependent parameters during the inference. In fact, Gibbs sampling updates the parameters conditional on the previous values of the other parameters. Let's assume that we want to sample two distinct parameters θ_1 and θ_2 as follows:

$$\theta_1, \theta_2 \sim \pi(\theta_1, \theta_2) \tag{B.4}$$

We begin with initial values $(\theta_1^{(0)}, \theta_2^{(0)})$, and the Gibbs sampler for an iteration t is given by

$$\theta_1^{(t)} \sim \pi(\theta_1 | \theta_2^{(t-1)})$$
$$\theta_2^{(t)} \sim \pi(\theta_2 | \theta_1^{(t)})$$

With $\theta = (\theta_1, ..., \theta_k)$, the generalized Gibbs sampler is then given by

$$\theta_1^{(t)} \sim \pi(\theta_1|\theta_2^{(t-1)}, ..., \theta_k^{(t-1)})$$
$$\theta_2^{(t)} \sim \pi(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, ..., \theta_k^{(t-1)})$$
$$\vdots$$
$$\theta_k^{(t)} \sim \pi(\theta_k|\theta_1^{(t)}, ..., \theta_{k-1}^{(t)}, \theta_{k+1}^{(t-1)}, ..., \theta_k^{(t-1)})$$

Here again, the main aim is to generate enough iterations such that the Markov chain $\theta^1, \theta^2, ...\theta^T$ has stationary distribution $\pi(\theta)$.

B.3 MCMC workflow

TESS MCMC algorithm starts by randomly initializing the vector Z ($Z^{(0)}$). By assuming $t = 1, 2...T$ iterations, the following steps are performed:

1. Sample $P^{(t)}, Q^{(t)} \sim Pr(P, Q|X, Z^{(t-1)})$ with Gibbs sampler
2. Sample $Z^{(t)} \sim Pr(Z|P^{(t)}, Q^{(t)})$ with Gibbs sampler
3. Update α using MH

For the update of the parameters, the reading of supplementary materials of Pritchard [73] and Durand *et al.* [25] is suggested.

.

Appendix C

Full-trend model results

C.1 Admixture proportions

Here are the maps representing the spatial variation of the average admixture proportions of the 5 lowest DIC runs.

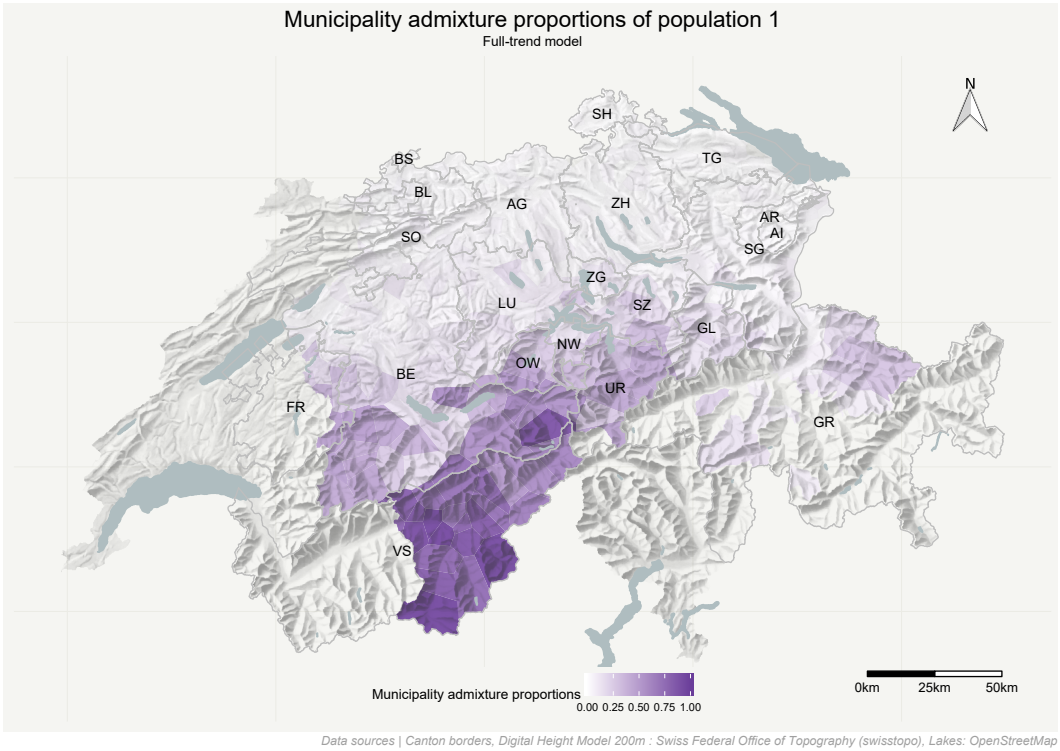


Figure C.1: Full-trend model: Spatial distribution of ancestral population 1

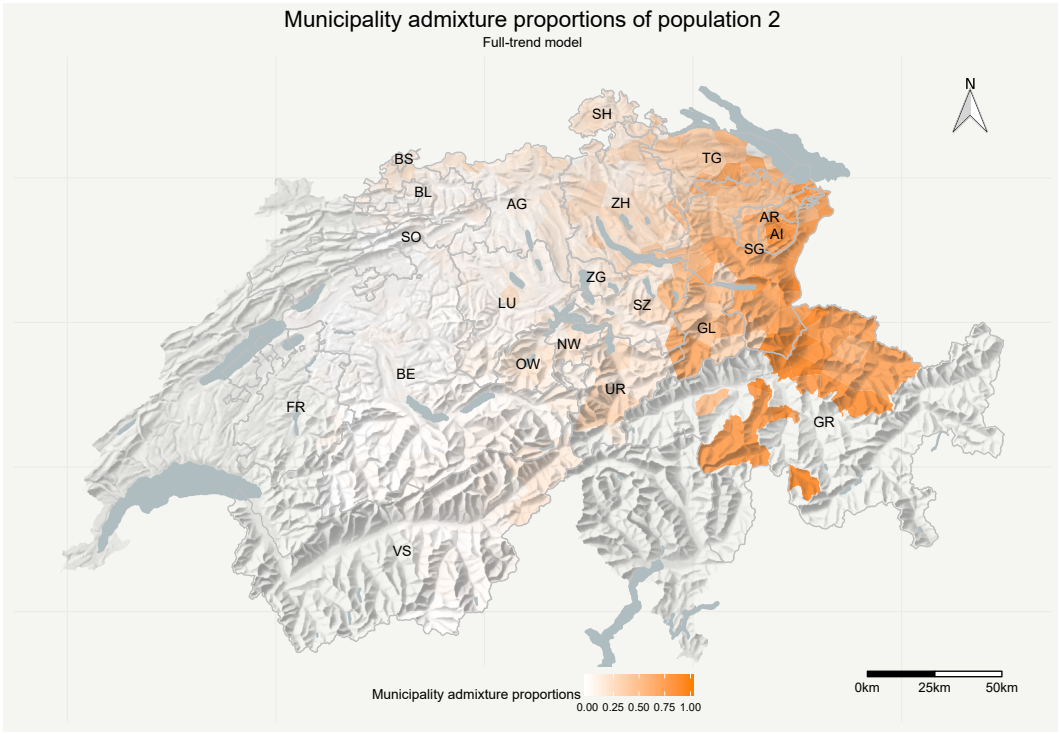


Figure C.2: Full-trend model: Spatial distribution of ancestral population 2

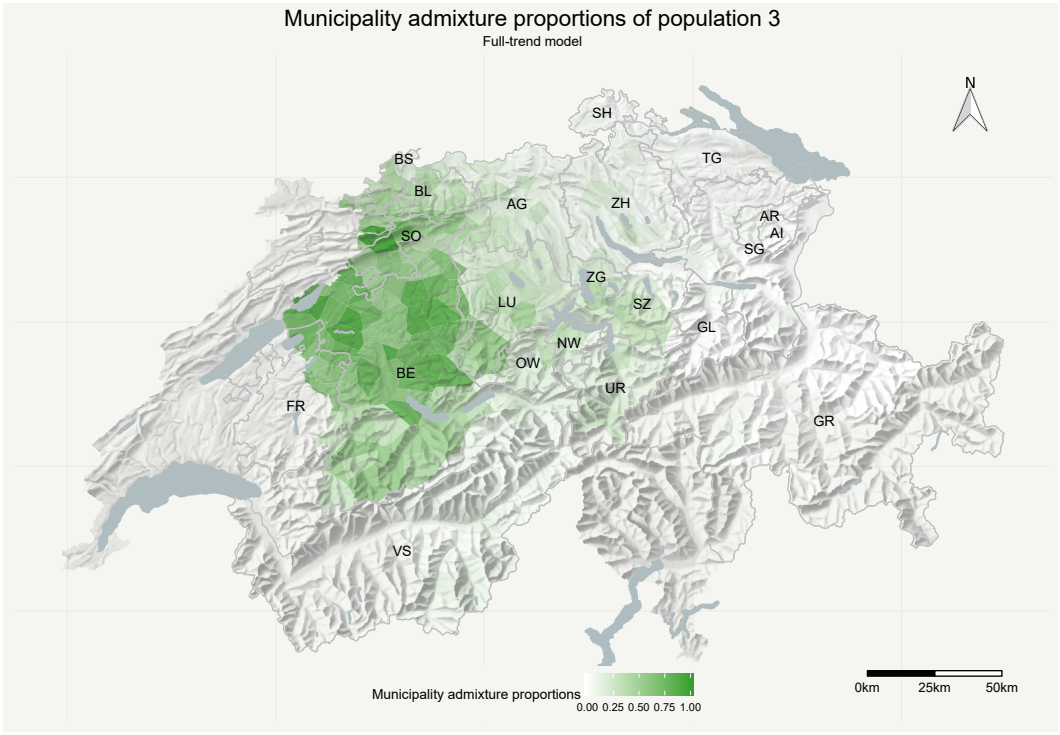


Figure C.3: Full-trend model: Spatial distribution of ancestral population 3

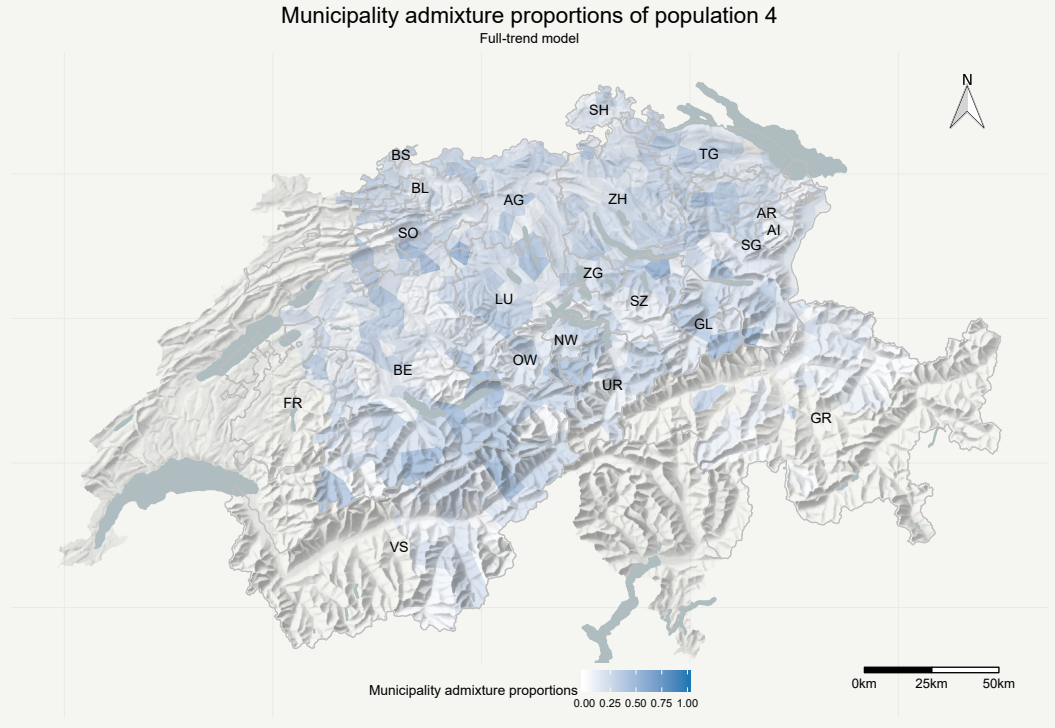


Figure C.4: Full-trend model: Spatial distribution of ancestral population 4

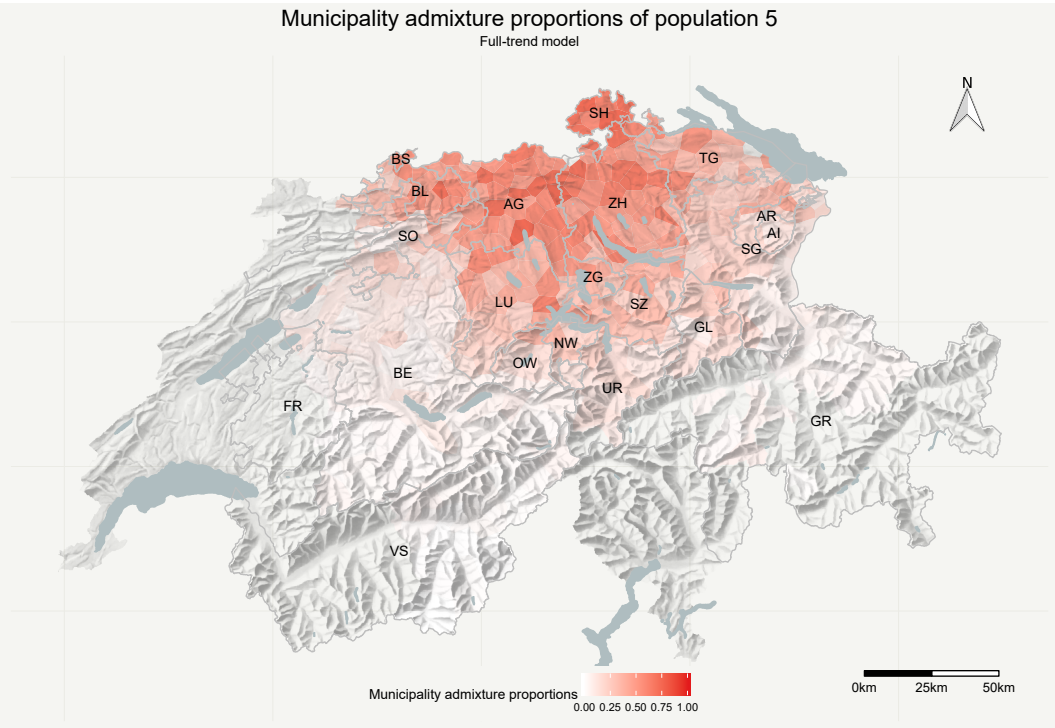


Figure C.5: Full-trend model: Spatial distribution of ancestral population 5

C.2 Spatial trend and autocorrelation

Here are the spatial trend and the autocorrelation estimates for $K_{max} = 5$ and the lowest DIC run.

	β_{long}	β_{lat}	$\beta_{long}CI_{95\%}$	$\beta_{lat}CI_{95\%}$	ρ
Population 1	-0.18	-5.71	[-0.70, 0.20]	[-6.05, -4.24]	0.02
Population 2	-2.73	-0.53	[-2.94, -2.16]	[-0.95, -0.07]	0.02
Population 3	0.01	4.34	[-0.22, 0.23]	[2.93, 4.82]	0.04
Population 4	0.00	-0.32	[-0.26, 0.29]	[-0.72, 1.32]	0.01
Population 5	3.23	-1.17	[2.79, 3.46]	[-1.85, -0.56]	0.02

Table C.1: Full-trend: β estimates, relative CI and ρ estimates

Appendix D

Non-spatial model results

D.1 Admixture proportions

Here are the maps representing the spatial variation of the average admixture proportions of the 5 lowest DIC runs.

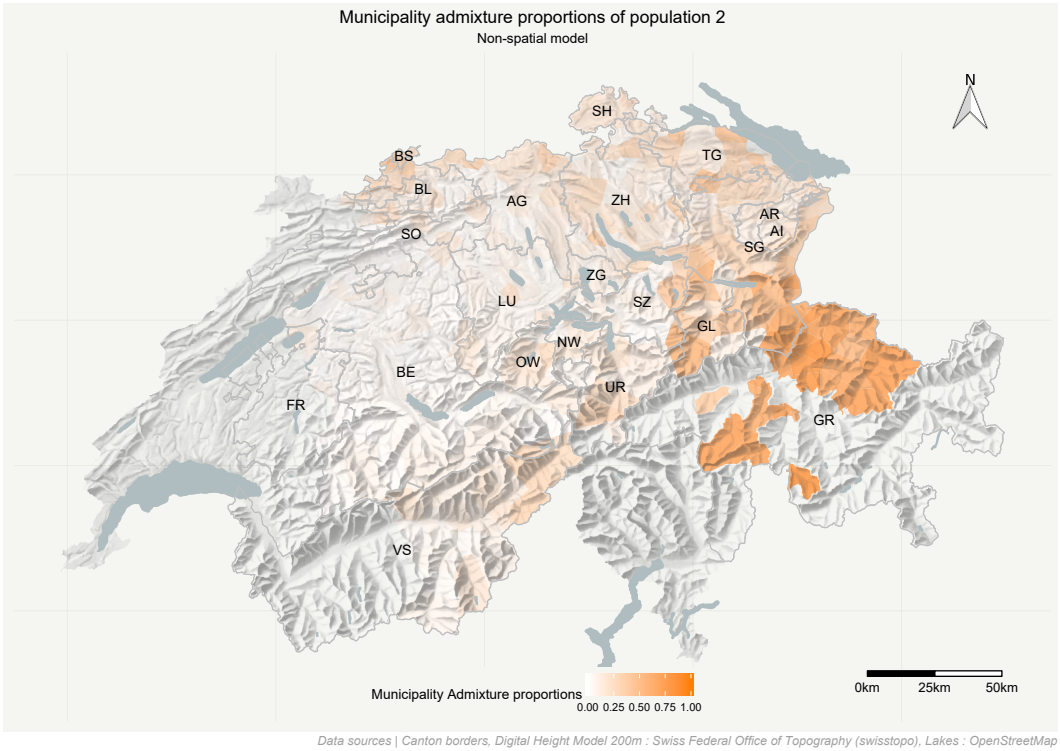


Figure D.1: Non-spatial model: spatial distribution of ancestral population 2

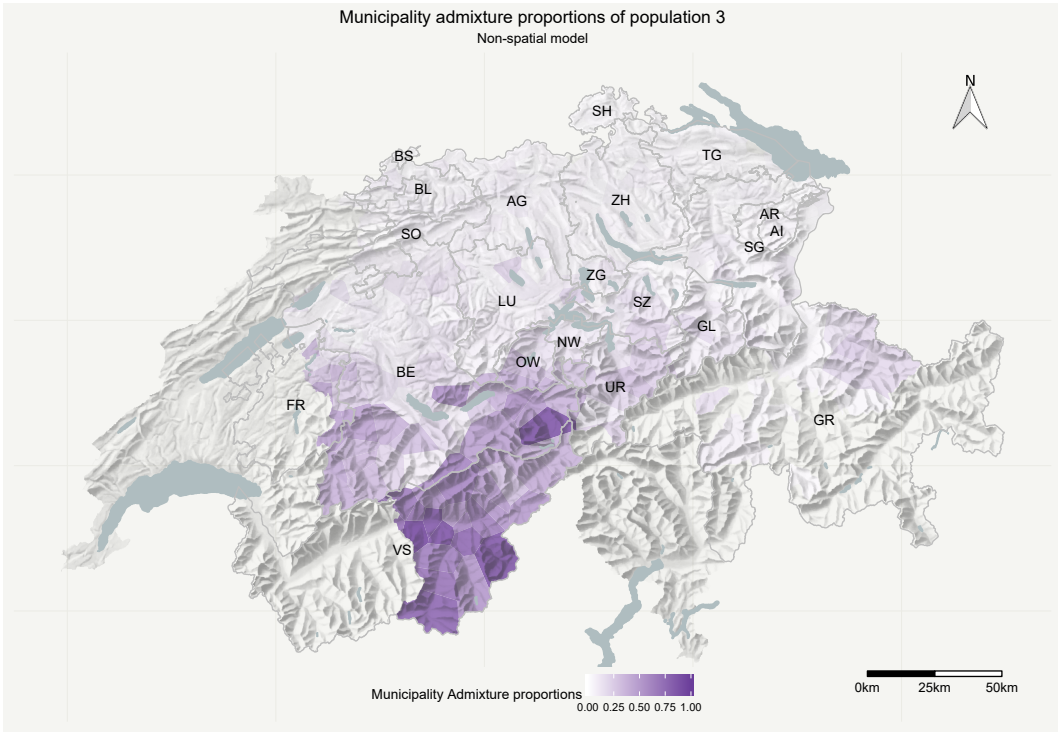


Figure D.2: Non-spatial model: spatial distribution of ancestral population 3

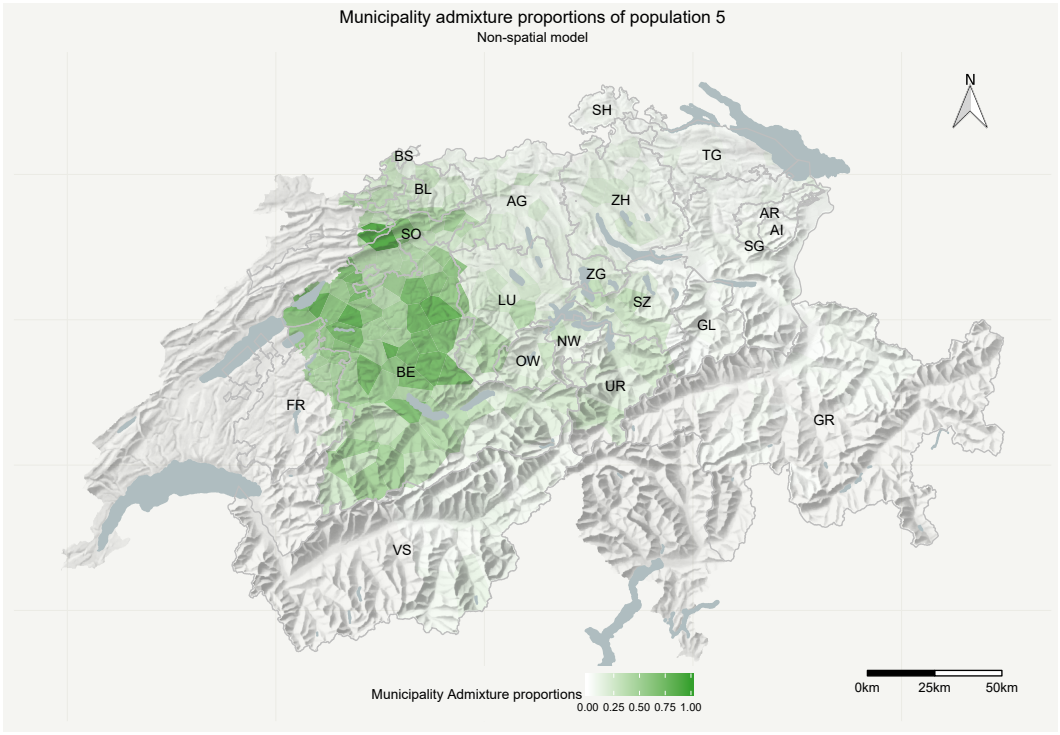


Figure D.3: Non-spatial model: spatial distribution of ancestral population 5

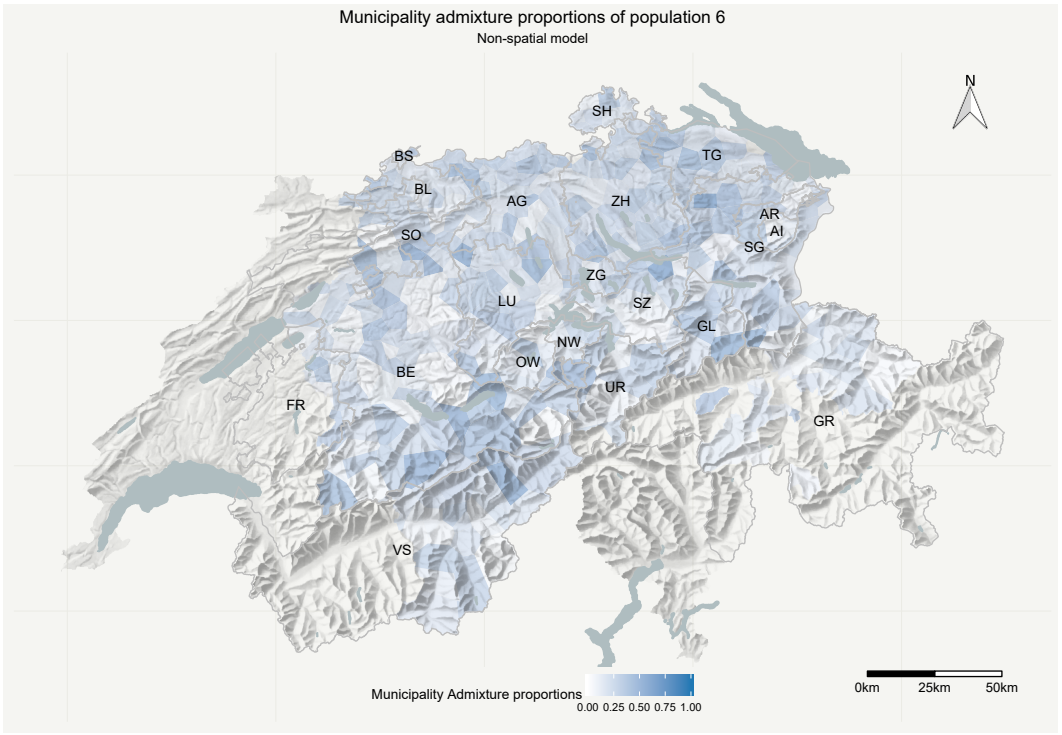
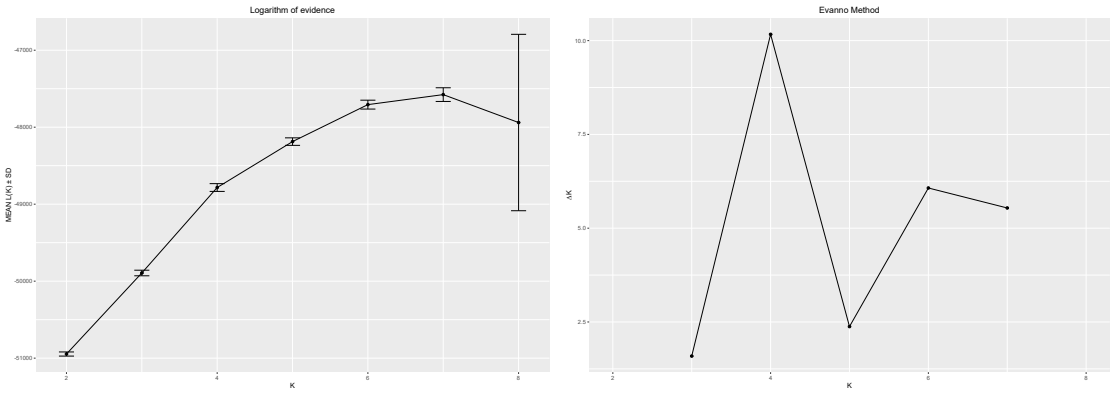


Figure D.4: Non-spatial model: spatial distribution of ancestral population 6

Appendix E

STRUCTURE results

E.1 Choice of the number of populations



(a) Logarithm of evidence **(b)** Evanno method

Figure E.1: Choice of the number of ancestral populations

E.2 Admixture proportions

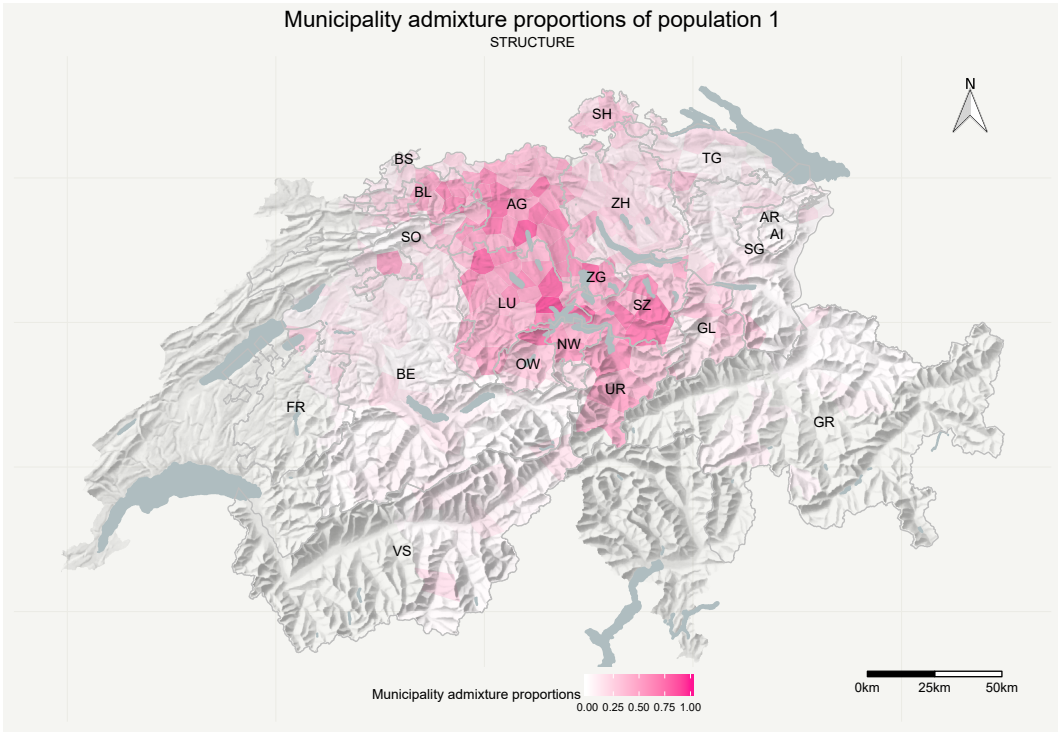


Figure E.2: STRUCTURE: Spatial distribution of ancestral population 1

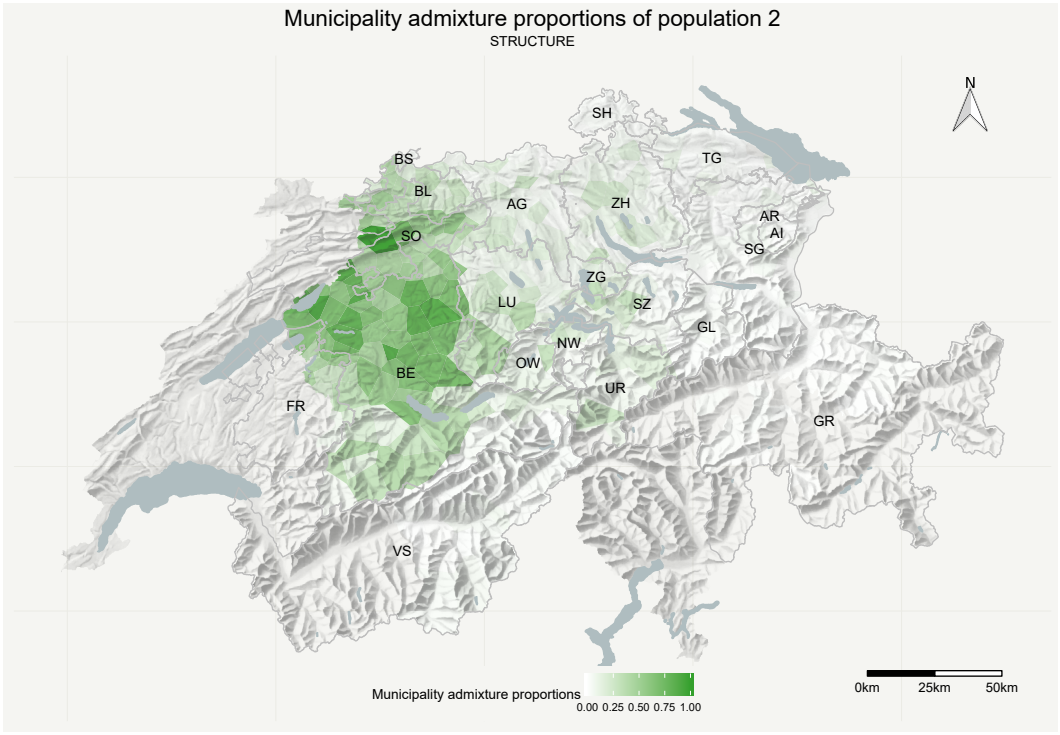


Figure E.3: STRUCTURE: Spatial distribution of ancestral population 2

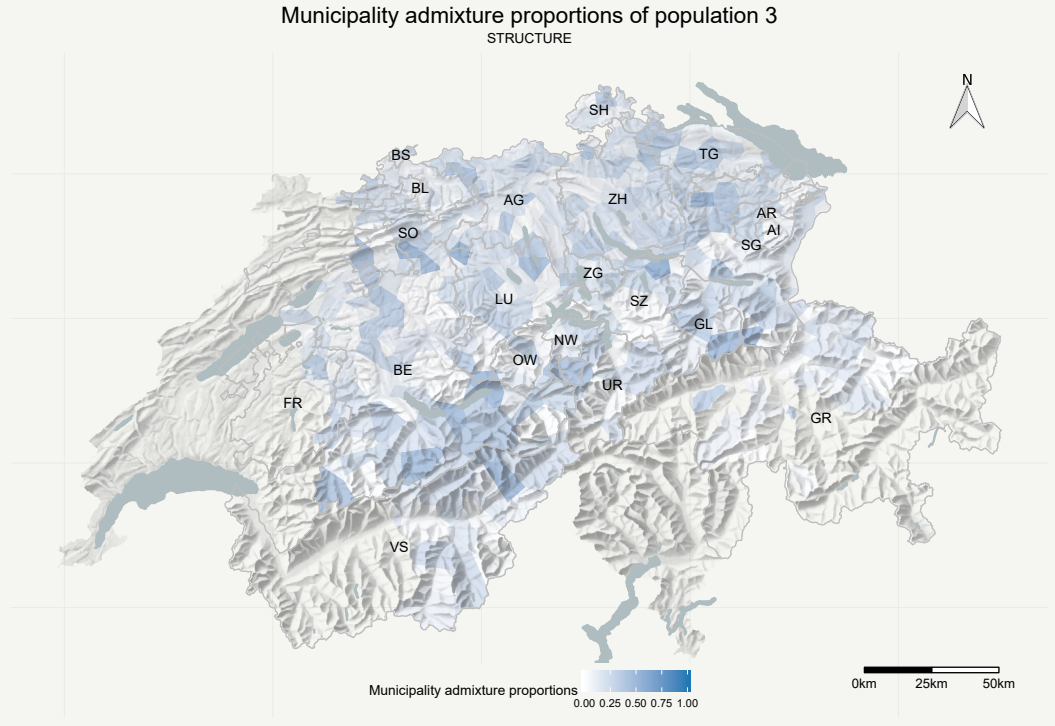


Figure E.4: STRUCTURE: Spatial distribution of ancestral population 3

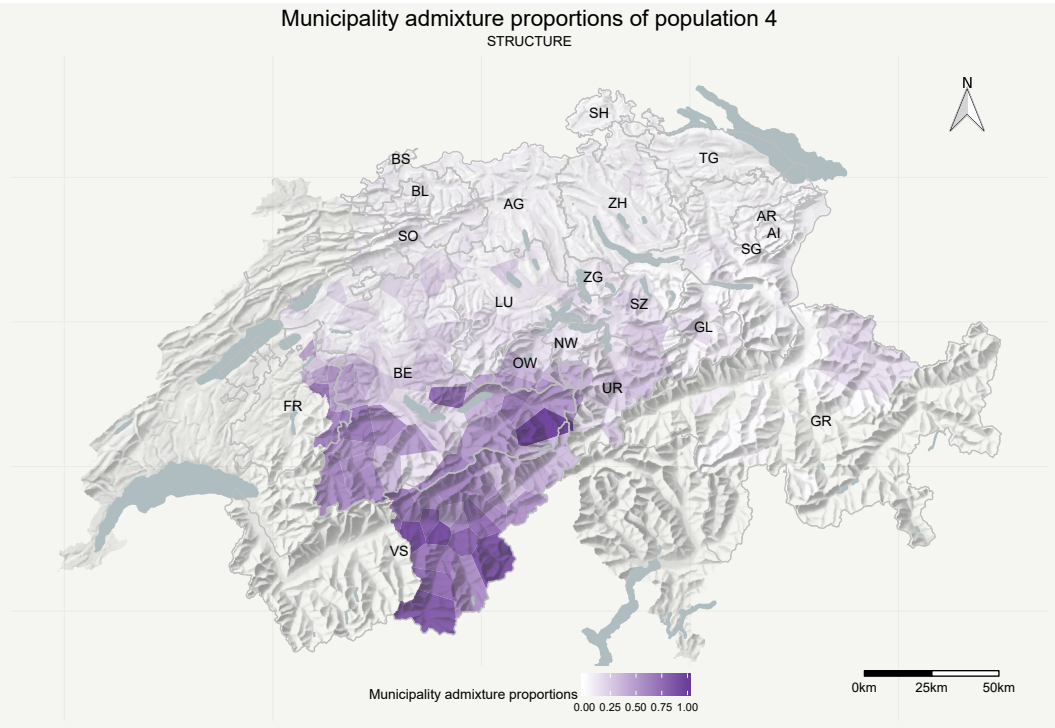


Figure E.5: STRUCTURE: Spatial distribution of ancestral population 4

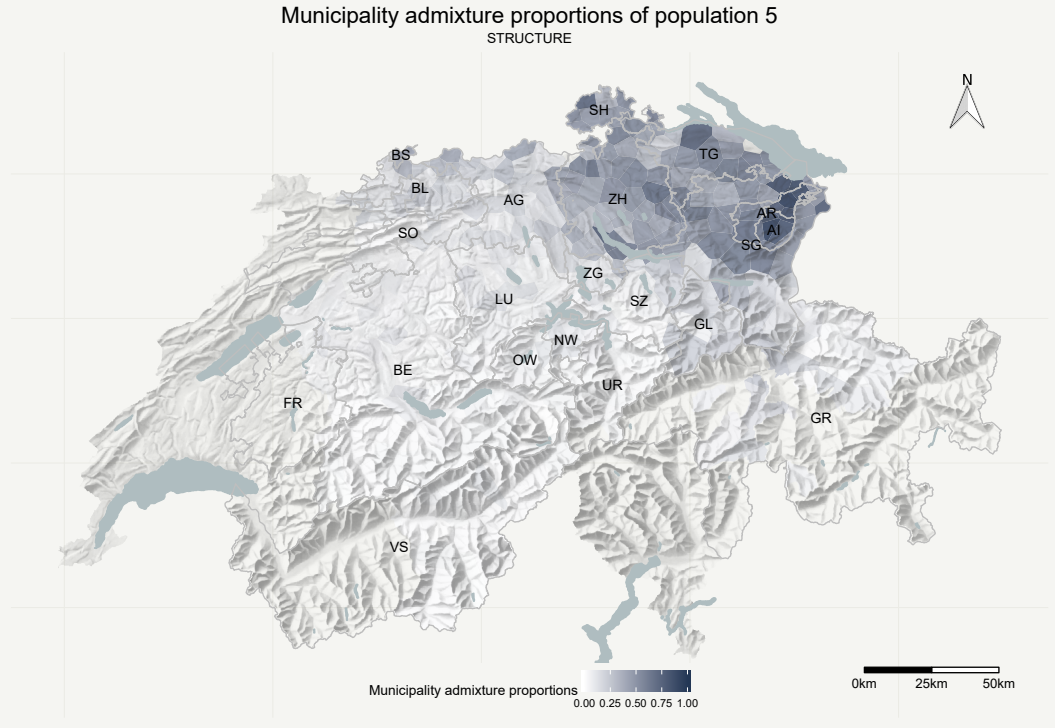


Figure E.6: STRUCTURE: Spatial distribution of ancestral population 5

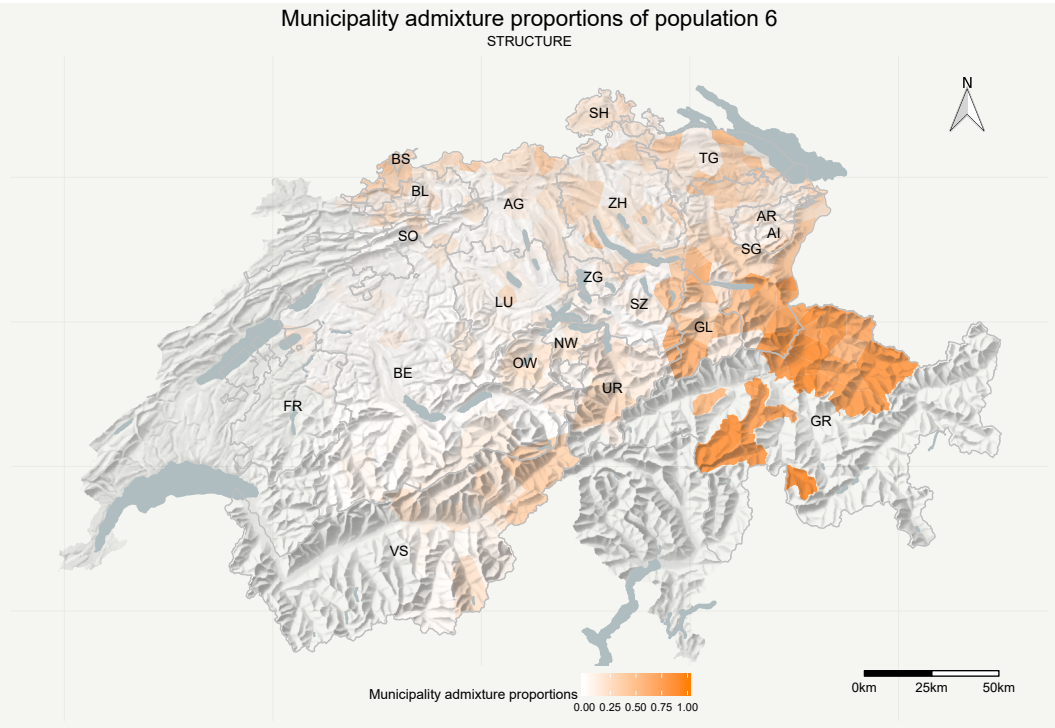


Figure E.7: STRUCTURE: Spatial distribution of ancestral population 6