



Master's thesis

Evaluating methods for predicting climatic conditions from pollen data

Elli Kiiski

Supervisors:
J. Sakari Salonen
Frederik Schenk
Indrė Žliobaitė

May 13, 2026

Title: Evaluating methods for predicting climatic conditions from pollen data

Author: Elli Kiiski

Date: May 13, 2026

Page count: 41 pages

Abstract:

Six climate reconstruction models using fossil pollen as a proxy to estimate climatic conditions are compared in terms of their design and performance and then applied to reconstruct the Holocene climate in northern Europe. The set of models includes three models used traditionally in climate reconstructions (MAT, WA, WAPLS) and three machine-learning based models newer to the field of paleoclimatology (RF, ERT, BRT). The study focuses particularly on how the models capture seasonal climate signals and whether differences between the reconstruction models influence reconstructed temperature trends.

Careful data preparation yielded pollen data from 211 sites in northern Europe, patterns of which were analyzed before running the climate reconstruction models. Principal component analysis revealed patterns interpreted as temperature-related and spatiotemporal variation supporting the suitability of fossil pollen as a climate proxy.

The reconstruction models were applied separately to reconstruct July and January mean temperatures in order to explore the possible role of seasonality in the Holocene Temperature Conundrum, a discrepancy between simulated and reconstructed temperature trends in the past twelve thousand years. While the climate reconstruction showed relatively consistent results for summer temperatures across all models, the winter reconstruction revealed greater variation between the models. Especially, WAPLS stood out from the rest of the models by producing systematically different temperature estimates compared to the other models.

The results suggest that model selection can significantly influence the temperature trends of climate reconstructions. This study demonstrates the importance of comparing multiple reconstruction models when analyzing long-term temperature trends and underlines the potential biases in seasonal climate reconstructions.

Keywords: Climate reconstruction models, fossil pollen, Holocene Temperature Conundrum, seasonal temperature reconstruction, MAT, WA, WAPLS, RF, ERT, BRT, PCA

Summary in Finnish

Menneiden ilmasto-olosuhteiden arviointi siitepölydatan avulla

Tutkiessamme ilmasto-olosuhteita tuhansien vuosien takaa emme voi turvautua mittaustuloksiin, joita ihmiset ovat pystyneet määrittämään vasta muutamien satojen vuosien ajan. Ympäristöön on kuitenkin jäänyt maapallon historian varrella jälkiä, joiden avulla voidaan selvittää millaiset olosuhteet maailmassa on vallinnut eri aikakausina.

Maan sedimenttikerrokseen ajan myötä kertyneet siitepölyhiukkaset ovat osa todistusaineistoa, joka on tallentunut luontoon menneisyyden olosuhteista. Ajoitettujen siitepölynäytteiden perusteella on mahdollista päätellä, mitkä kasvit ovat menestyneet maapallolla kunakin aikana, mistä voidaan edelleen arvioida millaiset ilmasto-olosuhteet ovat mahdollistaneet kyseisten lajien kasvamisen ja lisääntymisen. Näin ollen sedimentin siitepölykerrosten perusteella voidaan rekonstruoida menneitä ilmastoja eli muodostaa käsitys ilmasto-olosuhteista vuosituhansien takaa.

Laskennallisessa mielessä ilmaston rekonstruointi vastaa ohjatun oppimisen (supervised learning) mallin rakentamista, joka ennustaa ilmasto-olosuhteita esimerkiksi siitepölyaineiston perusteella. Mallien kouluttamiseen käytetään nykypäivän siitepölyhavaintoja ja niitä vastavia ilmastomuuttujien mitattuja arvoja, joiden väliset suhteet opittuaan mallit pystyvät ennustamaan ilmasto-olosuhteita pelkän siitepölydatan avulla.

Maisterintutkielmassani esitellään kuusi rekonstruktio menetelmää, jotka valitsevat holoseenin (viimeisten 12 000 vuoden) kausittaisia keskilämpötiloja pohjoisessa Euroopassa siitepölydatan avulla. Näitä malleja, joihin lukeutuu kolme perinteisesti ilmastorekonstruktioissa käytettyä mallia (MAT, WA, WAPLS) ja kolme paleoklimatologian alalla vielä uudempaa, koneoppimiseen perustuvaa mallia (RF, ERT, BRT), arvioidaan toimintaperiaatteen ja suorituskyvyn perusteella sekä käytetään datankeräyksen jälkeen holoseenin ilmaston rekonstruointiin.

Kansainvälisestä Neotoma-tietokannasta kerätty ja yhdenmukaiseksi harmonisoitu siitepölyaineisto koostuu pohjoisen Euroopan alueelta 211 eri sijainnista kerätyistä siitepölynäytteistä. Siitepölydatan sisäisiä rakenteita analysoidaan pääkomponenttianalyysillä (PCA), missä datasta paljastuu lämpötilaan liittyviä sekä aika-avaruudellisia ryhmittymiä, mikä tukee siitepölyn soveltuvuutta ympäristön lämpötilojen rekonstruointiin.

Lämpötilojen kausivaihtelua on esitetty ratkaisun avaimeksi paleoklimatologian keskuudessa kuuluisaan ristiriitaan (the Holocene Temperature Co-

nundrum), jossa ilmastosimulaatiot ja -rekonstruktiot arvioivat holoseenin vuosittaisten keskilämpötilojen kehityksen erilaisiksi. Tutkielmassani emme pyri suoraan ratkaisemaan tätä holoseenin lämpötilatrendeihin liittyvää ristiriitaa, vaan keskityimme tutkimaan mahdollisia mallista riippuvaisia vinoumia ilmastorekonstruktioissa.

Tutkiaksemme mahdollisia vuodenajasta riippuvaisia vaihteluita holoseenin ilmastossa, arvioimme ilmastorekonstruktioimalleilla erikseen heinäkuun ja tammikuun keskilämpötiloja. Kesälämpötilojen osalta mallit näyttävät tuottavan keskenään hyvin samankaltaisia tuloksia, kun taas talvirekonstruktioissa hajonta on huomattavaa. Erityisesti WAPLS erottuu joukosta tuottamalla muista malleista eroavia lämpötila-arvioita. Tulokset viittaavatkin siihen, että menetelmän valinnalla on merkittävä vaikutus lämpötilarekonstruktioihin, ja täten tutkielmani havainnollistaa useamman rekonstruktioimallin vertailun tärkeyttä ilmastoja rekonstruoitaessa.

Contents

1	Introduction	1
2	Background in paleoclimatology	3
2.1	Concepts and definitions	3
2.2	The Holocene Temperature Conundrum	5
3	Climate reconstruction models	7
3.1	Traditional models	7
3.1.1	Modern analogue technique	8
3.1.2	Weighted averaging	9
3.1.3	Weighted averaging partial least squares	10
3.2	Machine-learning based methods	10
3.2.1	Random forest	11
3.2.2	Extremely randomized trees	11
3.2.3	Boosted regression trees	11
3.2.4	Rejected models	12
3.3	Validation	12
3.4	Performance	13
4	Data and methods	15
4.1	Fossil pollen data	15
4.1.1	Pollen as a climate proxy	15
4.1.2	Characteristics of fossil pollen data	16
4.2	Data preparation	17
4.2.1	Datasets from the modern pollen database	17
4.2.2	Data harmonization	19
4.3	Climate variables	22
4.3.1	Choosing the variables	22
4.3.2	Primary vs. secondary variable	22
4.4	Workflow	23
4.4.1	Principal component analysis	23
4.4.2	Reconstructing the climate	26

4.5	Use of AI tools	27
5	Results and analysis	28
5.1	Patterns in pollen data	28
5.1.1	PCA for the pollen taxa	28
5.2	Reconstructed Holocene climate	30
5.2.1	Average reconstruction across all models	30
5.2.2	Comparing the models	30
6	Outcomes	36
6.1	Conclusions	36
6.2	Future research	37
	Bibliography	38
A	Code and data	41

Chapter 1

Introduction

Understanding how climate has changed in the past is essential in addressing present-day climate change and predicting how the climate will evolve in the future. At the same time, studying the mechanisms driving climate trends is important from a purely scientific perspective, as it deepens our understanding of the Earth’s geological history. Since environmental measurements such as temperature records extend back only a few centuries, past climatic conditions must be derived indirectly from ecological evidence.

Fossil pollen is one of the many traces that remain in nature from tens of thousands of years ago. Based on geologically dated pollen data it is possible to assess the vegetation of a particular region during a specific time. Furthermore, by understanding how climatic conditions influence vegetation, the past climatic conditions can be reconstructed from the composition of the pollen assemblages of that time.

From a computational perspective, reconstructing a climate means building a supervised learning model that predicts climatic conditions from observed fossil pollen data. Modern pollen records together with measured climate variables are used to train the models to learn the relationships between vegetation and climate, after which applying the models to the fossil pollen assemblages provide an estimation of the past climatic conditions.

In this thesis we study and compare the models used in climate reconstructions estimating past climatic conditions from fossil pollen data. Another central objective is to present the patterns found in the pollen data and analyze the climate trend of the past twelve thousand years in northern Europe based on a new reconstruction.

A major focus of this study is in the models themselves: to compare the formulation, assumptions and performance of three machine-learning based models to three traditional statistical approaches. More specifically, we ex-

amine how the models capture the effect of seasonality that has been identified as a possible key to a conundrum in paleoclimatology, a discrepancy between reconstructed temperature trends and climate model simulations. Rather than attempting to resolve the conundrum, we investigate potential model dependencies in seasonal temperature reconstructions for a limited region to shed light on how those dependencies may cause bias in the reconstructed climate.

In addition, data quality filtering and harmonization form an essential part of this study. Through careful selection of the datasets and preprocessing of the data, the quality and reliability of the reconstruction are secured.

We begin by introducing the paleoclimatological framework and the climate reconstruction models. We then discuss fossil pollen as a climate proxy, describe the workflow and methods, and interpret patterns in the pollen data using principal component analysis. Finally, we reconstruct past climate, analyze the reconstruction, and evaluate the behaviour and performance of the models.

Chapter 2

Background in paleoclimatology

People have been able to measure the temperature only for a few centuries, which is barely a blink of an eye in a geological context. To study the climate before such technology *paleoclimatology*, the research of past climates, relies on the ecological evidence preserved in the nature, such as pollen grains in the sediment layers. Depending on the timeframe of interest, other possible ecological factors include other microfossils like plankton or insect remains as well as tree rings, and ice cores [28]. Climate reconstructions utilize these signals in estimating the climatic conditions in the past.

Another technique to study past climates is simulations, that are based on known physical and environmental phenomena [19]. For example the Earth's orbital insolation and concentration of greenhouse gases in the atmosphere are factored in the model simulations. However, the simulations and reconstructions from the same period do not necessarily agree. As will be shown later, that is currently the case when analyzing the epoch of the past twelve thousand years.

Before discussing the conundrum demonstrating the modelling challenges related to climate reconstructions, we introduce the key concepts and definitions to paleoclimatology and fossil pollen analysis.

2.1 Concepts and definitions

Paleoclimate

A past climate predating the instrumental record, i.e. the climate is studied using something other than climate station readings. This is the research interest in paleoclimatology.

Paleotemperature

A reconstructed temperature dated to a past point in time.

Kiloanno (*ka*)

A unit of time in geology: one kiloanno equals one thousand years. In paleoclimatology, it is usually used to denote *thousand years before present*, where the “present year” is defined as 1950, when the first radiocarbon datings were just published [23] (the 76 years difference to the current year is insignificant on the scale of thousands of years).

Radiocarbon dating

The most widely used method for determining the age thousands of years back. It is based on the fact that a radioactive ^{14}C isotope of carbon is obtained by a living organism only during its lifetime, and from the moment of death the remaining ^{14}C starts to decay [23]. With a known half-life (about 5,730 years), the age of organic material can be traced back to roughly 50,000 years.

Calibrated radiocarbon year

A radiocarbon dating calibrated to correspond calendar years. Because the fluctuations in ^{14}C levels of the atmosphere have varied over time, the measured ^{14}C in an organism is adjusted with the known reference levels to derive a calibrated dating [23].

Holocene

The current epoch (i.e. our geological era). It began roughly 11.7 thousand years ago, when the latest ice age ended [9]. This is the time range of interest in this thesis, 11.7—0 ka.

Proxy

A measurable biophysical, geochemical or ecological indicator that provides information about past environmental conditions. The input variable of a climate reconstruction.

Climate variable

A measurable environmental property that describes a climatic condition. The output variable of a climate reconstruction.

Fossil pollen

Preserved pollen grains found in sediment layers [7]. They are used to study the vegetation of the past and further utilized as a proxy in climate reconstructions. (Although fossil pollen is the term commonly used, the pollen is not fossilized into rock and hence, should technically be called *subfossil pollen*.)

Taxon (*plur. taxa*)

A group of organisms classified at any rank in the taxonomic hierarchy (the

system of classifying organisms on different levels). Pollen is generally identified on the family level (Poaceae, grasses) or genus level (*Betula*, birch) (occasionally down to species level). Any of these labels is a taxon.

2.2 The Holocene Temperature Conundrum

One of the central challenges in analyzing climatic conditions of the past is that climate reconstructions sometimes disagree with climate model simulations. Such inconsistencies raise questions about the assumptions, limitations, and biases of the reconstruction models. These questions are particularly relevant for our study, which focuses on comparing a set of reconstruction models.

An infamous example of such conflict between reconstructions and simulations is the *Holocene Temperature Conundrum*, a term coined by Liu et al. (2014) [19] to describe the differing trends of the Holocene climate. When modelling the temperatures in the last twelve thousand years, the simulated graph shows a fairly steady warming trend, while the reconstructions suggest that the average annual temperature of the world peaked at one point during that time.

This discrepancy is visualized in Figure 2.1. As can be seen in the panel B, the overall consensus from the collection of simulations in Liu et al. (2014) is that the global temperature has generally increased during the Holocene. The paleoclimate reconstructions in the panel A, however, tell a different story: There seems to have been a maximum (often referenced as the *Holocene Thermal Maximum*) in the global temperature average somewhere in the mid-Holocene, according to the extensive reconstruction of Kaufman et al. (2020) [16].

Kaufman et al. (2020) [16] suggest that one possible key to the conundrum may be to consider the seasonal variation in temperatures around the globe. If the reconstructions are driven mainly by the summer temperature, perhaps reconstructing the winter temperature separately would reduce the peak.

In their paper *Machine-learning based reconstructions of primary and secondary climate variables from North American and European fossil pollen data* (2019) [24], Salonen et al. present models to reconstruct the Holocene climate in northern Europe separately for summer and winter temperatures. Six methods were concluded to be suitable for the task, all using fossil pollen as a proxy. The goal of this thesis is to apply these models for an extensive amount of fossil pollen data in order to examine if the models capture seasonal signals differently and to assess whether the Holocene Thermal Maximum still appears once we factor in the summer and winter temperatures separately.

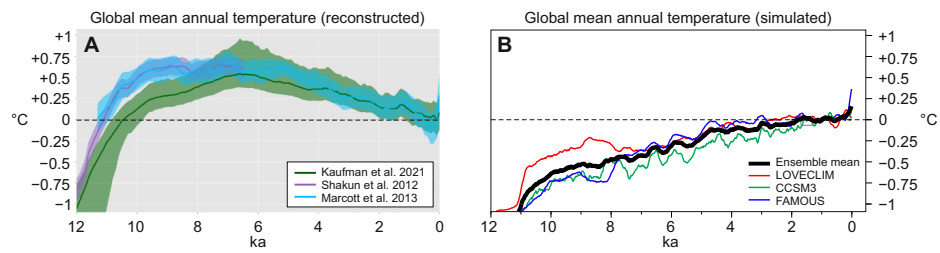


Figure 2.1: Comparison of Holocene paleoclimate reconstructions and model simulations. A and B show the Holocene global mean annual temperature in reconstructions (A; adapted from Kaufman et al. 2020 [16]) and model simulations (B; adapted from Liu et al. 2014 [19]), illustrating the Holocene Temperature Conundrum. (Figure put together by J. Sakari Salonen.)

Chapter 3

Climate reconstruction models

Reconstructing a climate means building a link from the proxy data to certain values of the climate variables. In this chapter, we study the models used in predicting the July and January temperatures from fossil pollen data and compare the classical reconstruction models to the newer approach presented in Salonen et al. (2019).

In the article, a total of eight models were trained to reconstruct the Holocene climate. The models, listed in Table 3.1 and presented in greater detail below, included classical methods (MAT, WA, WAPLS) in addition to newer machine-learning based methods, including regression-tree ensemble methods (RF, ERT, BRT) and neural network approaches (NNET, ELM).

3.1 Traditional models

Before the invention of machine-learning models, climate reconstructions were built solely using traditional methods based on statistical relationships. In principle, they all are variations of matching and regressing modern records to the fossil pollen data. Next, we cover the basic principles behind these methods.

To begin with, let us agree on notation. The climate variable (temperature) is denoted by x_i , where the index i represents the site at which it has been measured. When talking about a prediction of the climate variable (paleotemperature), we use $\hat{x}_i$. The predictor variables (the proportion data of the taxa) are denoted by y_{ik} , where i again represents the site and k the taxon. A variable with the superscript $*$ is always an estimate of the variable in question.

Table 3.1: Modelling tools used in Salonen et al. (2019) [24] and their parameterization.

Code	Method	Parameters
MAT	Modern analogue technique	Weighted mean of 5 closest analogues
WA	Weighted averaging	Monotonic deshrinking, tolerance down-weighting, square-root transformation of species data
WAPLS	Weighted averaging-partial least squares	3-component models, square-root transformation of species data
RF	Random forest	100 trees
ERT	Extremely Randomized Trees	Number of random cuts = 5
BRT	Boosted regression trees	Maximum number of trees = 3000, learning rate = 0.025, tree complexity = 4, bagging fraction = 0.5
NNET	Artificial neural network	Linear output units. Number of units in the hidden layer = 18 (T_{jul}), 19 (T_{jan}).
ELM	Extreme Learning Machine	Prediction with a mean of 5 networks. Positive linear activation function. Other parameter values specified.

3.1.1 Modern analogue technique

Perhaps the most widely used reconstruction method for fossil pollen data, the **modern analogue technique (MAT)** is a straightforward method based on the observation that similar composition of pollen taxa in a sample indicates similar environmental conditions [27]. Matching fossil pollen data to the most similar modern pollen samples with known temperature records (i.e. the modern analogue) provides an estimate of the corresponding paleotemperature. The numerical similarity of the modern and fossil sample can be defined by different metrics. In Salonen et al. (2019) they used the squared chord distance, a commonly used distance metric with pollen data [22]. Mathematically,

$$d_{ij}^2 = \sum_k (\sqrt{y_{ik}} - \sqrt{y_{jk}})^2. \quad (3.1)$$

The squared chord distance d_{ij}^2 between the modern sample i and fossil pollen sample j denotes their dissimilarity. The lower the value the closer the pollen compositions of the samples and, hence, the closer the climate conditions are at those sites and times. Usually the reconstructed temperature is determined by the weighted mean of a certain number of the closest analogues. In Salonen et al. (2019) this number was 5.

3.1.2 Weighted averaging

Weighted averaging (WA) is based on the assumption that each taxon has an optimal growth temperature, and the closer a site's actual temperature is to this optimum, the more abundant the pollen of that taxon will be at the site in question [27]. The optimum estimate u_k^* for taxon k is calculated from the training set as the average of the temperatures of all sites, weighted by the proportion of the taxon in the samples:

$$u_k^* = \frac{\sum_i y_{ik} x_i}{\sum_i y_{ik}}. \quad (3.2)$$

This is also called the regression step of WA.

The next step is calibration, where the process is turned around. To calculate the temperature estimate x_i^* , the average of the optimal temperatures of the taxa is weighted by the proportion of their occurrence in the sample:

$$x_i^* = \frac{\sum_k y_{ik} u_k^*}{\sum_k y_{ik}}. \quad (3.3)$$

This process of double averaging tends to shrink the range of the estimated x_i^* values compared to the actual temperature scale of x_i [1]. To fix this, an additional step of deshrinking regression is added to derive the factors by which the estimated temperatures are biased. The goal is to determine the coefficients a_0 and a_1 for updated taxa optima $u_k = a_0 + a_1 u_k^*$. The updated optima can then be used in reconstructing the paleotemperatures from the data. The coefficients are derived from the deshrinking regression formula

$$x_i = a_0 + a_1 x_i^*. \quad (3.4)$$

Regressing the set of the measured temperatures x_i on the estimated temperatures x_i^* over each site i solves the coefficients a_0 and a_1 . Now, using the formula $u_k = a_0 + a_1 u_k^*$ for each taxon k , we get the final prediction

$$\hat{x}_i = a_0 + a_1 x_i^* = a_0 + a_1 \frac{\sum_k y_{ik} u_k^*}{\sum_k y_{ik}} = \frac{\sum_k y_{ik} u_k}{\sum_k y_{ik}}, \quad (3.5)$$

which allows us to reconstruct the paleotemperatures from the dated taxa data.

3.1.3 Weighted averaging partial least squares

Adding to the WA method, **WAPLS (weighted averaging partial least squares)** conveys the residual structure of the fossil pollen data for further components. The final paleotemperature is a linear combination of those components. The first of these components is equivalent to the result of the WA method.

The second component is calculated from the residuals of the first prediction $r_i = x_i - \hat{x}_i$. The residuals are used to construct the second component that captures the remaining linear structure not explained by the first component. This is done with a procedure similar to WA:

1. calculate the optima for that second component by weighted averaging the residuals: $u_k^* = \frac{\sum_i y_{ik} r_i}{\sum_i y_{ik}}$,
2. calibrate the residuals: $r_i^* = \frac{\sum_k y_{ik} u_k^*}{\sum_k y_{ik}}$,
3. regress the residuals with $r_i = a_0 + a_1 r_i^*$, to get the prediction $\hat{r}_i$ of the second component.

The process can be iterated further to get the third and higher order components, each trying to capture any additional structures remaining in the higher order residuals. In Salonen et al. (2019) they used three components.

The final prediction for the paleotemperature in WAPLS is calculated by regressing the linear combination of all of the components:

$$\hat{x}_i = b_0 + b_1 \hat{x}'_i + b_2 \hat{r}_i + b_3 \hat{r}'_i, \quad (3.6)$$

where $\hat{x}'_i$ is the first component (the WA prediction) and $\hat{r}_i$ the second component (the first residual) and $\hat{r}'_i$ the third component (the second residual).

3.2 Machine-learning based methods

To complement the traditional assemblage matching and regression based approaches, there is nowadays a variety of machine-learning based models that make fewer assumptions about the ecological relationships in the data. In Salonen et al. (2019) they tested how three regression-tree ensemble models and two neural networks were able predict the paleotemperatures.

3.2.1 Random forest

Random forest (RF) is an ensemble model of decision trees [2]. A single decision tree is produced by an algorithm that consecutively splits the fossil pollen data in two groups: on each round, the split is made based on one taxon and threshold abundance value. The tree grows when another split is made for every group of the already split data of the previous round. The stopping condition for the splitting is determined beforehand.

When building a single tree, the split on each round is made using the best split possible, i.e. the algorithm chooses whichever taxon can be split in two parts best separating the target temperatures in the most homogeneous groups. However, in the case of RF, a subset of predictors is first chosen randomly and the best split then chosen among them [18]. This random element protects the algorithm against overfitting and reduces the computations made on each round.

Where one decision tree already is a method to predict the past temperatures, it is not very robust in handling outliers and is still prone to overfitting. To address these issues and to improve the overall accuracy, RF uses a multitude of decision trees, the average of which will be the final predicted temperature. Moreover, to add the random element, the trees are built by bootstrapping (each using a random mixture of the original data). For every tree, the training data is gathered by randomly sampling the rows of the original data with replacement. In other words, each dataset may have zero, one, or more instances of each row from the training data.

3.2.2 Extremely randomized trees

Extremely randomized trees (ERT) is inherently random forest with random splits [26]. In every node, instead of computing the best split for each taxon, the split threshold value is chosen by random. Only after that, the best taxon, based on which to split, is chosen. This added randomizing factor further reduces overfitting of the data.

3.2.3 Boosted regression trees

Boosted regression trees (BRT) is an ensemble method similar to random forest, where the trees are built sequentially rather than independently [6]. BRT iteratively utilizes the residuals of the previous tree and adjusts the prediction based on a chosen learning rate.

Having temperature values x , proportion data of the taxa y and learning rate η , BRT updates the prediction through the following steps:

1. Initialize the prediction as the mean of the temperature values $\hat{x} = \text{mean}(x)$.

2. Calculate residuals $r = x - \hat{x}$.
3. Fit a small decision tree to residuals. The new tree prediction $\hat{r}$ is the mean of the residuals in each leaf.
4. Update prediction $\hat{x}' = \hat{x} + \eta \cdot \hat{r}$ (and repeat with $\hat{x}'$ from step 2).

Steps 2-4 are repeated until a pre-determined stopping condition (usually maximum number of trees, set to 3,000 in Salonen et al. (2019)) is met. Similarly to the weighted averaging based models, the final prediction is a combination of the individual tree predictions.

3.2.4 Rejected models

In Salonen et al. (2019), they also tested two neural network algorithms for climate reconstructions. **NNET (traditional neural network implementation)** and **ELM (extreme learning machine)** process information through layers, each transforming the input. However, these models did not perform well in the study and, therefore, will not be covered here.

3.3 Validation

After training the models, their performance is validated by trying them out on modern data that has not been used in the training. Cross-validation is a common validation method, where the training data is sampled multiple times. Each time, the data is separated into training data that is used in calibrating the models and test data that is then used to test how well the model predicts the actual value.

For a climate reconstruction model, the nature of the training data has to be considered in the validation. Fossil pollen data is spatially correlated (details in section 4.1) and therefore the validation method needs to tackle the issue of spatial autocorrelation.

H-block cross-validation addresses autocorrelation by excluding the h observations preceding and following the target sample from training the model [3]. In Salonen et al. (2019), h-block cross-validation was implemented by having a buffer zone of radius h around the sample. This approach ensures that the prediction is made by using a model trained with data points located sufficiently far away to eliminate the influence of spatial correlation.

The problem with h-block cross-validation is to choose the range h optimally [29]. With too small a value of h the autocorrelation is not resolved. On the other hand, with too large h many observations are omitted in the training and performance decreases. In Salonen et al. (2019), the suitable value was sought by comparing bigger radii to $h = 0$ (the case equivalent to leave-one-out cross-validation). A varying radius (h values between 0-1,500 km were

tested in 100 km increments) was used to find the best size of the buffer. The optimal h -value was concluded to be in the range of approximately 200-600 km, where the observed error plateaued.

3.4 Performance

Conventionally, the traditional methods (MAT, WA, WAPLS) have been found to perform reliably in climate reconstructions. For example, Kaufman et al. (2020) gathered existing Holocene paleotemperature records [15], i.e. climate reconstructions, and conducted an extensive multi-method reconstruction for Holocene global mean temperature [16]. The majority of the pollen-based reconstructions were obtained from Marsicek (2018) [20], which used the modern analogue technique to reconstruct the Holocene paleotemperatures.

Weighted averaging based models tend to perform well with compositional data, with taxa that may be absent in many of the samples [1]. However, the simple WA method is known to overlook residual correlations because it regards each environmental variable separately. WAPLS, on the other hand, utilizes the residuals in the iterative steps, producing in principle more accurate predictions.

Before Salonen et al. (2019), the machine-learning based methods were a relatively new approach in the studies using fossil pollen data, and microfossil data in general. Goring et al. (2010) [8] tested random forest in modern pollen samples and concluded it to outperform WA and WAPLS (but not MAT) on one of the two chosen climate variables.

Also, boosted regression trees had been used in ecological modelling [5] with inconclusive results in pollen-climate calibration [25] but often outperforming WA in simulated fossil datasets [14], especially in cases where the taxa responded better to an ecologically less influential climate variable. To the knowledge of the authors of Salonen et al. (2019), extremely randomized trees had not been used for microfossil data before their study. Hence, any previous knowledge of the performance of ERT is not available.

In Salonen et al. (2019), most of the eight models reconstructed the Holocene climate with a reasonable degree of accuracy. Tables 3.2 and 3.3 show the models ranked by the error (root mean square error of prediction, RMSEP). For both July and January temperature predictions, the machine-learning based methods (RF, ERT, BRT) were the top three most accurate models.

The traditional methods (MAT, WA, WAPLS) did not fall too far behind the machine-learning based models and were found to give reliable predictions. The neural networks (NNET, ELM), however, did not perform consistently

well across the datasets. Therefore, they were not recommended for climate reconstruction.

Across all models, the performance was better in the July reconstructions. In terms of error and maximum bias, the January reconstructions had more uncertainty, which is to be expected since winter temperatures in general have less ecological influence and tend to show higher variability compared to summer. However, the same models seem to be performing the best in both summer and winter reconstructions.

Table 3.2: JULY Cross-validated performance metrics for the individual pollen–climate calibration models in Europe (Salonen et al. (2019) [24]).

Rank	Method	RMSEP	Max. bias
1	BRT	1.61	5.59
2	RF	1.61	7.75
3	ERT	1.63	7.81
4	WA	1.72	7.47
5	WAPLS	1.73	5.90
6	MAT	1.82	8.78
7	ELM	1.98	6.35
8	NNET	2.15	7.77

Table 3.3: JANUARY Cross-validated performance metrics for the individual pollen–climate calibration models in Europe (Salonen et al. (2019) [24]).

Rank	Method	RMSEP	Max. bias
1	BRT	2.92	7.16
2	ERT	3.08	9.36
3	RF	3.19	9.66
4	WAPLS	3.28	10.29
5	ELM	3.37	10.00
6	WA	3.46	13.46
7	MAT	3.47	6.73
8	NNET	4.03	7.99

Chapter 4

Data and methods

To reconstruct and analyze the Holocene climate, a workflow needs to be established all the way from gathering the fossil pollen data from sites in northern Europe to running the reconstruction models and picking the analysis methods for the results. In this chapter, we cover the characteristics of pollen data and define our output variables. We also explain the process of data preparation and introduce the methods used in analyzing the reconstruction.

4.1 Fossil pollen data

The reconstruction models utilize fossil pollen sequences as the input data. Fossil pollen has proven to be a suitable indicator for estimating climate conditions in the past, but its inherent characteristics also introduce uncertainties that may affect the quality and precision of the data.

4.1.1 Pollen as a climate proxy

What makes fossil pollen a good indicator of the past climate, is its sensitivity to changes in the environment, such as fluctuations in temperature and precipitation [4]. Since different plants require different conditions to survive and reproduce, the abundance of the pollen of each species in the sediment reveals how climate conditions have changed over time.

Another advantage of fossil pollen is its broad availability all over the world. In a year, a forest may produce dozens of kilograms of pollen per hectare [10]. The grains then spread around efficiently covering a large area, making the "pollen rain" not only abundant but also fairly homogeneous in a given environmental setting. Factoring in the chemical durability of a pollen grain, it is clear why fossil pollen is present in most types of sediment.

Due to its broad availability, fossil pollen can be found even in regions where humans have had the most impact, which is crucial for creating reliable global reconstructions. On the other hand, the wealth of pollen grains in sediment allows their occurrences to be calculated proportionally, rather than being limited to assessment only based on presence versus absence.

The ecological features of pollen, not limited to the mentioned ones, make it a good and widely available indicator of climate change. Hence, accounting for how the pollen proportions of different locations change over time is the basis of the climate reconstruction models. Of course, the set of the ancient species has to overlap significantly with those growing today for calibrating the models to be possible.

In our case, reconstructing the summer and winter temperatures separately, it is expected that the proxy is influenced by the temperatures of both growing season and winter time. This expectation is fulfilled by the fact that each taxon has its own "fundamental niche", a set of required environmental conditions to survive and reproduce [11]. Furthermore, the effects of changes in the temperatures of the opposite seasons should not have too similar effects on the pollen abundances. As will be presented later, this has also been ensured in the variable selection of Salonen et al. (2019).

4.1.2 Characteristics of fossil pollen data

Fossil pollen data is collected by drilling a sediment corer into the ground (usually the bottom of a lake or swamp for the soil to be untouched) and extracting several meters of sediment [10]. Samples extracted along the length of the sediment tube are then examined, and the proportion of each fossil pollen type is counted from that sample. Radiocarbon dating is performed for certain depths to establish chronological anchor points. The full chronology is then interpolated based on these datings to determine the age for each sample.

In a layer of sediment, there may be thousands of fossil pollen grains, all of which are impossible to count. Therefore, only a portion of the grains are counted and logged, introducing a risk for the more rare taxa to not get detected. Weng et al. (2006) [30] showed that a taxon with over 5% frequency in the sediment would be detected almost certainly in inspecting at least 100 grains. With a frequency of only 0.1%, to get even an 80% detection probability, would require identifying over a thousand grains from the sample.

One fossil pollen sample, i.e. a specimen collected from a specific site and depth, has at least the following attributes: calibrated radiocarbon age and the number of counted pollen grains of each occurring taxon. However, the taxonomic level, on which the pollen in the sample is differentiated, depends

mostly on the expertise and effort of the person doing the grain counting. Some recognize different pollen at species or genus level, and others may only mark down the family [10].

In short, the diligence (how many pollen grains do they count altogether) and competence (how precisely they classify the pollen) of the geologist sitting behind the microscope, affects how rare taxa we can expect to cover in the analysis. This clearly creates an opportunity for bias and ambiguity, especially if the data is not harmonized in advance.

Also, the quality of the radiocarbon dating varies across the sequences obtained from different sites by different people. Low resolution, i.e. the number of radiocarbon datings made per a length of a sediment sample, in a sequence introduces uncertainties in the validity of the chronology. Obviously, the less frequently measured anchor points exist, the more unreliable the interpolated datings become.

Fossil pollen data, as ecological data tend to be, is *spatially autocorrelated*. Site locations close to each other are more likely to yield similar data than sites on the opposite sides of Europe. Since the fossil pollen abundances can be, to some extent, predicted by the samples of nearby sites, those samples are not stochastically independent [17]. This creates challenges for the validation of the reconstruction models using the autocorrelated data. As covered in section 3.3, in Salonen et al. (2019) the models were validated with h-block cross-validation, precisely to combat the issue of autocorrelation.

4.2 Data preparation

Data preparation is the first challenge in building a high quality climate reconstruction. Although there is a good number of fossil pollen records available to us, a few steps are necessary before the data is ready to be input to the reconstruction models. In this section we cover the data preparation process starting with all the pollen sites recorded globally. The number of datasets (i.e. sites) remaining at each state of the data processing is presented in Table 4.1.

4.2.1 Datasets from the modern pollen database

In the international Neotoma database [21], there are paleoclimatological datasets from all over the world, including fossil pollen records from hundreds of sites in Europe. While the data is freely available, it is not standardized and varies in resolution and taxonomic level.

For the Holocene climate reconstruction of northern Europe, the sites were first filtered based on the desired location and age coverage. After deciding on

Table 4.1: Number of sites and different taxa labels at each state of data preparation starting from all the pollen sites available globally in Neotoma database.

State	Sites	Taxa labels
All Neotoma pollen sites	5582	≥ 1956
Sites in northern Europe	2052	≥ 1956
Sites covering the Holocene	400	≥ 1956
Sites with sufficient data resolution	211	1956
Harmonized site data	211	70

additional conditions in an attempt to ensure the data quality, we accepted the datasets that met the following criteria:

- Location of the site: from 47.5°N to 90°N in latitude and from 12°W to 67.5°E in longitude.
- Age: covers the years between 10—2 ka.
- Pollen sample resolution: at least 30 fossil pollen samples.
- Resolution of the chronology: at least 5 radiocarbon datings.

The location parameters were chosen to cover the Europe to the west of the Urals and north of the Alps, where the climate circumstances are fairly uniform and similar to the ones in the calibration datasets of the reconstruction models. The age range, in turn, was selected so that all the datasets would extend to every 2,000-year section of the Holocene.

The minimum numbers of pollen samples and radiocarbon datings were put in place to filter out the sites with the poorest resolution, while preserving an adequate number of sites to input for the reconstruction. With at least 30 pollen samples, there are likely to exist datapoints dated on every kiloinno. In addition, for the data to be reliably dated, there must be datings frequently enough for the ages determined for each sample to be reliable. For the reconstruction, the minimum of 5 datings was agreed on. This was not so much because the number was thought to be explicitly sufficient, but to retain enough usable datasets.

With this criteria, a Python program (see appendix A) was implemented to fetch geographically and chronologically pre-filtered datasets from the Neotoma database and further filtering those by data resolution. The final number of pollen sites satisfying these criteria was 211, including seven sites provided by the thesis supervisors.

The geographic distribution of the sites passing the filter is not entirely uniform. In Figures 4.1 and 4.2 each dot, though different in size, represents

one fossil pollen site. We see that there are a lot of sites in the British Isles, whereas in Finland, Sweden and Russia, for example, there are regions totally uncovered. This lack of coverage in some regions is not ideal but, fortunately, the overall number and distribution of the sites seems sufficient.

4.2.2 Data harmonization

After ensuring the quality of the data, the issue with the taxonomic classification of the pollen samples still remains. The names of the different pollen types are not uniform across the datasets. The pollen types may be identified at different taxonomic rank or they may be written by different abbreviations. Therefore, the data are harmonized under the labels that are used by the reconstruction models.

The samples were labeled by matching the taxon name reported in the Neotoma database to one of the 70 labels used by the reconstruction models. A tentative pairing was executed by a Python algorithm comparing the name to existing name-label pairs (see appendix A) and the final matching was confirmed by J. Sakari Salonen.

Finally, the absolute count of pollen grains was converted into the relative percentages of the 70 types in the sample. After this, we have the data in a standardized form for the climate reconstruction models. For birch and alder, this data is visualized in Figures 4.1 and 4.2.

The harmonized fossil pollen data was combined in one big CSV file, in which the site records are stacked consecutively. This data along with the metadata of the sites is available online (instructions in appendix A).

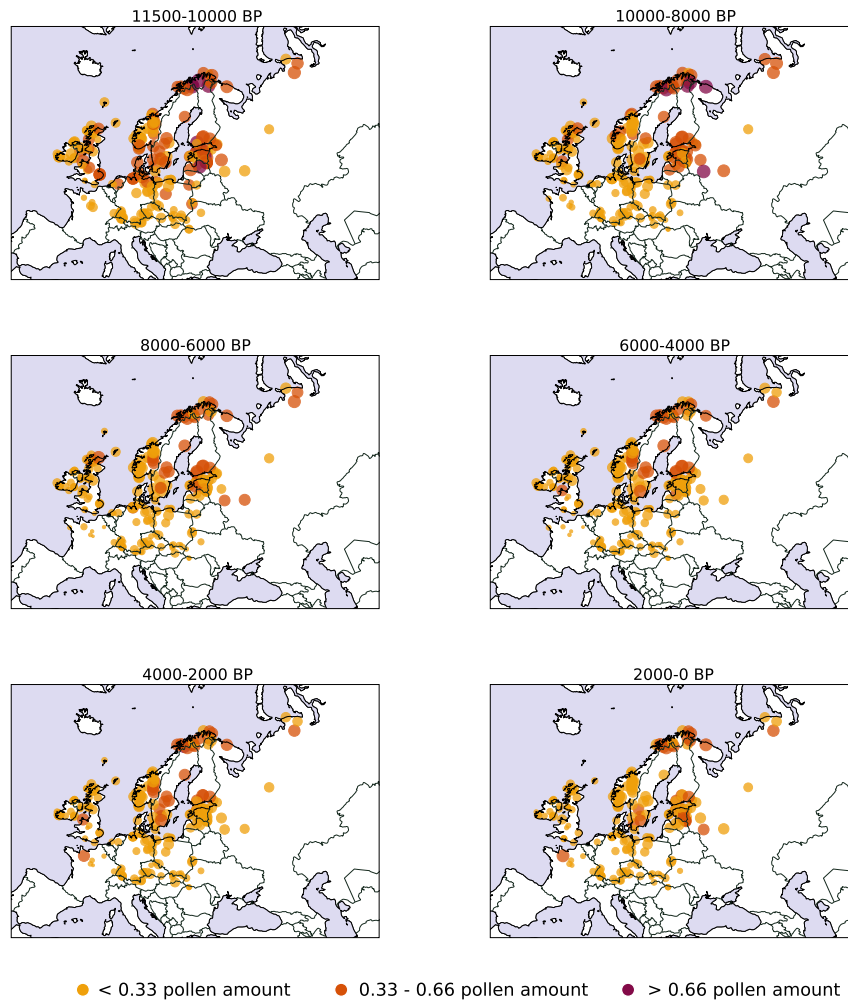


Figure 4.1: Birch (lat. *Betula*) pollen in northern Europe during Holocene. Birch pollen has been very abundant in early Holocene, especially in the northernmost Europe (over 66% of the grains in some sites). Though it is still very common, we can see that its relative incidence has dropped towards the current day.

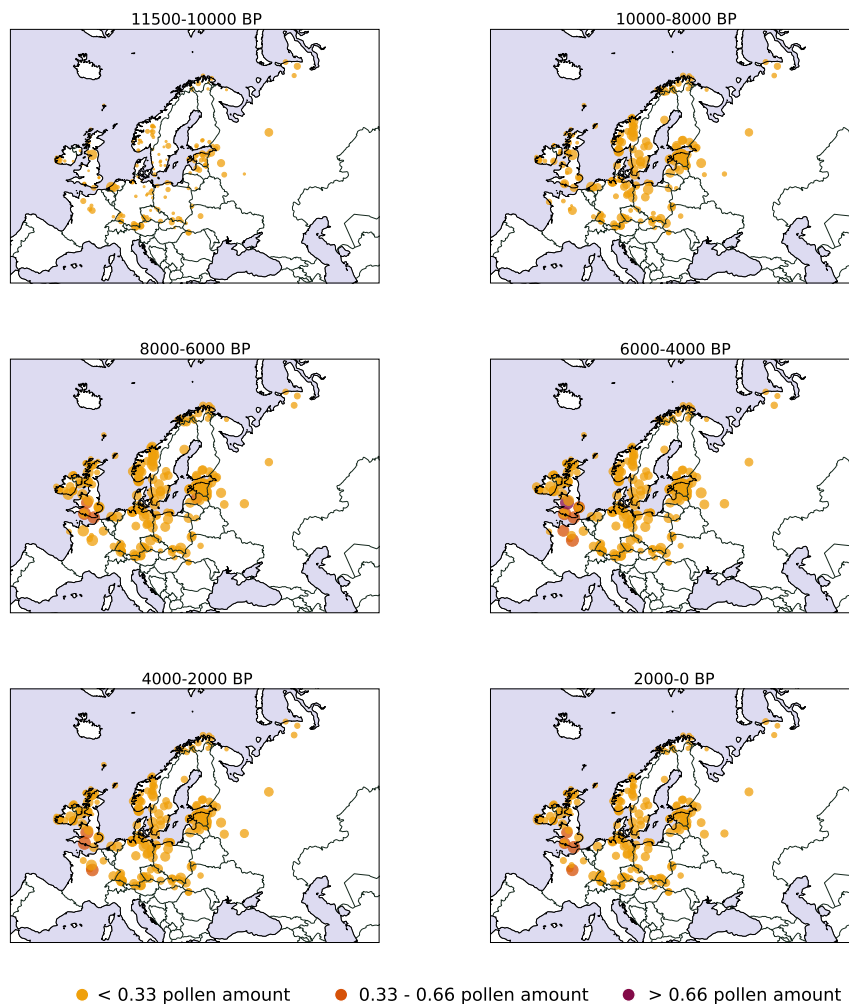


Figure 4.2: Alder (lat. *Alnus*) pollen in northern Europe during Holocene. On the contrary compared to birch, alder has increased its incidence in the early Holocene, then establishing a share of around 30% in central Europe.

4.3 Climate variables

We have established fossil pollen as the proxy for the climate reconstruction, but what are the variables to be reconstructed? Many recent approaches have presented reconstructions for annual average temperature, although commonly synthesized from individual local reconstructions for a variety of seasonal climate variables, and as a result showcased the Holocene Temperature Conundrum. To detect the possible effect of seasonality, we need to take another perspective. However, the variables can not be simply picked arbitrarily, instead they need to meet certain conditions to ensure the sensibility and quality of the reconstructions.

4.3.1 Choosing the variables

Since we are interested in the seasonal temperature variability as a possible key to the Holocene Temperature Conundrum, the two ideal climate variables for our reconstructions are the average summer and winter temperatures. Thus, we find the reconstruction methods presented by Salonen et al. (2019) suitable for our purposes, since their models use the mean July and January temperatures as variables for the European data.

The selection of a climate variable is dependent on the proxy. That is, the proxy has to be impacted by the climate variable being reconstructed. Moreover, the variable should not be strongly correlated with other ecologically significant variables, as it may cause the proxy to predict those variables rather than the variable being reconstructed [13]. Especially when reconstructing multiple variables, the combined effect of those variables needs to be considered.

In Salonen et al. (2019), the process of choosing the climate variables for the reconstructions started with calculating a Spearman correlation matrix for 20 potential variables with ecological influence. The subset of variables with less than 0.7 absolute correlation with each other were then ranked by how well their variation was explained by the modern calibration data. After excluding any variables with no ecological credibility or independent effect, the best two remaining variables, mean July and January temperatures, were chosen.

4.3.2 Primary vs. secondary variable

The climate variables to be reconstructed are ranked according to their impact on the proxy. The variable with more explanatory power is defined as the primary, whereas the rest as secondary, tertiary, and so on. In Salonen et al. (2019) two climate variables were used, corresponding to a primary and a secondary variable.

In their variable selection, Salonen et al. appointed the mean July temperature (T_{jul}) as the primary variable because in northern Europe the summer temperature is known to be ecologically more influential. The mean January temperature (T_{jan}) was actually found to explain the most variance out of all of the possible variables (with the greatest R^2 value). Having sufficiently low correlation with T_{jul} , it was chosen as the secondary climate variable.

Overall, the distinction between primary and secondary climate variables helps capture the complexity of the influence that changes in the climate have on the vegetation. It prevents oversimplifying the conditions impacting the growth of different species. For fossil pollen data, inherently containing data from multiple species with different ecological responses, reconstructing several climate variables utilizes the ecological information contained in the data more comprehensively.

4.4 Workflow

This study consists of three phases: data processing (presented in section 4.2), data pattern analysis, and reconstruction and analysis of the Holocene climate. The full workflow is visualized in Figure 4.3 and the latter two processes, data pattern inspection with principal component analysis and climate reconstruction by the reconstruction models, are explained in this section.

4.4.1 Principal component analysis

After data preparation and before running the climate reconstructions, we looked for possible patterns of the fossil pollen data. To reveal the underlying patterns the data entails, we aimed to reduce the complexity of the multivariate data using principal component analysis (PCA). We visualized the results to whether there are clusters of taxa behaving similarly and interpret, based on what property the taxa have clustered.

However, standard PCA is not directly suitable for compositional data such as our fossil pollen data of taxa proportions that sum up to one in each sample. Therefore, we applied a centered log-ratio (CLR) transformation to eliminate the dependency caused by the constant sum of each sample. More detailed descriptions of both mentioned methods, PCA and CLR, are provided below.

PCA was implemented with Python using the *scikit-learn* library (see appendix A for code) after utilizing *pandas* and *NumPy* for data handling and CLR transformation. For plotting, *matplotlib* was used.

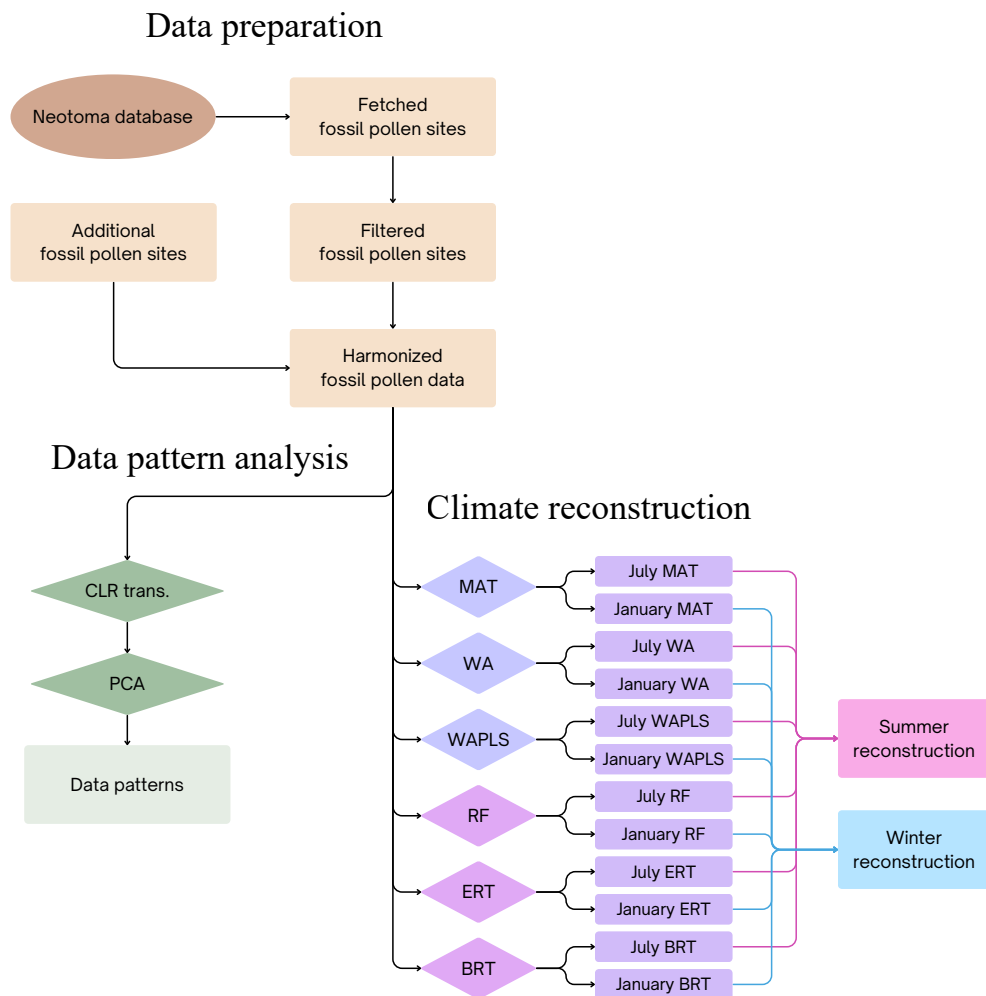


Figure 4.3: The workflow visualized as a flowchart.

Principal component analysis

PCA is a technique to reduce complex data with multiple dimensions to a smaller number of variables called principal components (PCs) [12]. These components present the data with fewer dimensions while preserving maximum amount of the original variation. Each principal component is a linear combination of the original variables and defines an axis in a transformed coordinate system. The components are formed so that PC_1 (the first principal component) captures the most variation within the original data, PC_2 the next largest amount, and so on. Mathematically, PCA can be formulated as follows.

Consider a data matrix $X \in \mathbb{R}^{n \times p}$, each column having zero mean. Each row represents an observation that has p features. For example, each observation could be a fossil pollen sample, while the features correspond to the abundances of p different taxa.

The covariance matrix

$$S = \frac{1}{n-1} X^\top X \quad (4.1)$$

describes how the features vary together.

By calculating the eigenvectors v_i and corresponding eigenvalues λ_i of the covariance matrix

$$Sv_i = \lambda_i v_i, \quad (4.2)$$

we get the directions of the principal components, each eigenvector defining a linear combination of the original features. The corresponding eigenvalues tell how much of the variance the principal component defined by the corresponding eigenvector explains.

Hence, the first principal component points to the direction of the eigenvector that corresponds to the largest eigenvalue. When the eigenvectors are ordered to correspond to the descending order of the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p$, the score (transformed data) of each principal component is calculated as the projection of the data X onto the corresponding eigenvector. For example, the first principal component is

$$PC_1 = Xv_1. \quad (4.3)$$

The further components are similarly defined by the subsequent eigenvalues. Because the covariance matrix is symmetric, its eigenvectors corresponding to distinct eigenvalues are orthogonal. Consequently, all PC axes are orthogonal to each other.

Centered log-ratio transformation

For CLR transformation we have to first calculate the geometric mean of each row x_i

$$g_i = \left(\prod_{j=1}^p x_{ij} \right)^{\frac{1}{p}}, \quad (4.4)$$

where p is the number of the taxa. The CLR transformation

$$\text{clr}(x_{ij}) = \log\left(\frac{x_{ij}}{g_i}\right) \quad (4.5)$$

returns values that remove the constraint of compositional data caused by the constant sum.

Since the logarithm is undefined for zero, a pseudocount of 10^{-6} is added to all zero entries before the transformation.

4.4.2 Reconstructing the climate

The best performing six models (MAT, WA, WAPLS, RF, ERT, BRT) from Salonen et al. (2019) were used to reconstruct the Holocene summer and winter temperatures. The carefully prepared data (same as used in the PCA for data patterns) was input to the script running the models one by one, each outputting a series of paleotemperature values.

The reconstruction models were run in R, using the implementation described in Salonen et al. (2019). The output of the script is a massive array of paleotemperatures, each row containing both July and January temperature predicted by each model (12 temperatures in total) from a specific site location and time. A CSV file with the output and corresponding site information is available online (instructions in appendix A).

To analyze the reconstruction, the July and January paleotemperatures were separated to reconstruct the summer and winter climate independently. The individual paleotemperatures were then combined to analyze the climate trend across all models, as well as each model separately (these results are presented in section 5.2).

Furthermore, in the analysis of the climate trends, the paleotemperatures are compared to today's temperatures in the same locations. For the modern day comparison, we use average reconstruction values from the last two millennia. *The modern reference summer and winter temperatures* for each site are defined as the mean of the last 2,000 years' reconstructed temperatures across

all six models. We call the difference between a given paleotemperature and the modern day reference temperature an *anomaly*.

The reason we do not use measured temperatures as the modern day reference is to minimize bias. Comparing the paleotemperatures to the reconstructed average temperatures closer to the modern day rather than actual temperatures of today (or even the pre-industrial time) reduces inconsistencies the reconstructed data may have compared to actual observations.

The data visualizations (including fossil pollen sites already presented in Figures 4.2 and 4.1) were created with Python (see appendix A), utilizing libraries such as *pandas* and *NumPy* for data handling, *matplotlib* for plotting, and *cartopy* for geospatial visualization. Also, locally weighted regression (LOESS) smoothing was applied using the *loess* package to show the temperature trend in the reconstruction data.

4.5 Use of AI tools

In writing this thesis, language models were used for background research and proofreading. NotebookLM was used in summarizing potential source material in order to assess whether they included relevant information for the study. ChatGPT was used to find spelling and grammar errors in the manuscript.

The programs developed for this study were created with assistance of ChatGPT in software configuration, debugging and identifying suitable Python libraries.

Chapter 5

Results and analysis

Finally, we present the results and interpretation of the pollen data patterns along with the Holocene climate reconstruction for northern Europe. In this chapter, the output data are visualized and analyzed relying on the paleoclimatological expertise of the thesis supervisors.

5.1 Patterns in pollen data

From a paleoecological perspective, certain plant species tend to survive and reproduce under similar climatic conditions. To explore such patterns in our complex, high-dimensional datasets, we present the data in a smaller number of dimensions to visually interpret those patterns.

5.1.1 PCA for the pollen taxa

To understand patterns in the pollen assemblages over the Holocene, PCA was run on the pollen data prepared for the reconstruction models. In the input data, rows represented a single fossil pollen sample in a specific location (site) and time (radiocarbon dating) and the value in each column was the abundance proportion of one taxon in the sample.

The first principal components explain 18.1% (PC_1), 8.7% (PC_2), 6.8% (PC_3) and 5.0% (PC_4) of the total variance. In ecological datasets, explained variance of 18% for the first component is considered relatively high. The first two principal components are then visualized, as they capture the most notable structure in the data.

Figure 5.1 shows two-component PCA with CLR transformation applied to the fossil pollen data. According to J. Sakari Salonen, one of the thesis supervisors and an expert in paleoclimatology, PC_1 can be interpreted as a core warming signal. In the right side of the first axis, we find a group of taxa

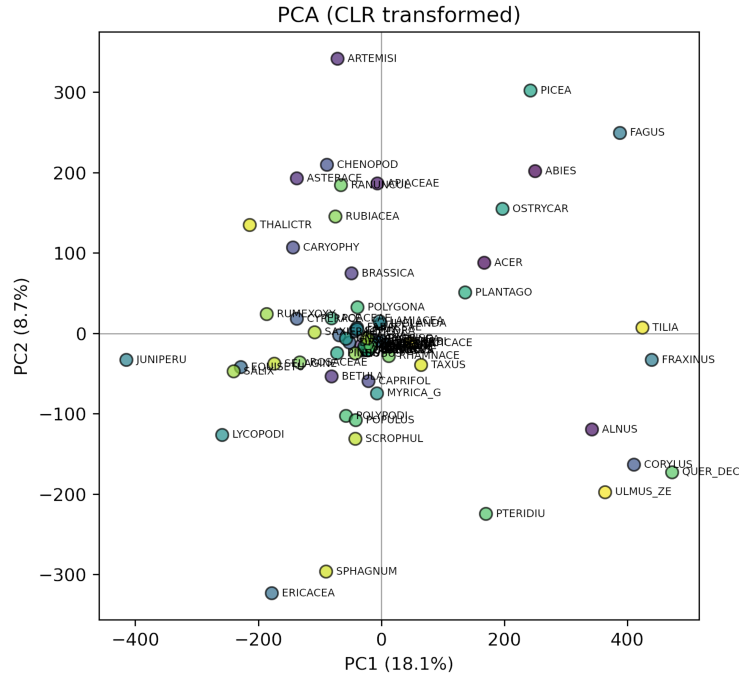


Figure 5.1: Principal component analysis on fossil pollen data. PC_1 (explains 18.1% of the variance) seems to divide the taxa based on their temperature optimum, while PC_2 (explained variance 8.7%) shows signs of having a spatiotemporal nature.

(*Fagus*, *Tilia*, *Fraxinus*, *Alnus*, *Corylus*, deciduous *Quercus*, *Ulmus*/*Zelkova*) that are common warmth indicators. Taxa found on the left (*Juniperus*, *Salix*, *Rumex*/*Oxyria*), on the other hand, are taxa that tolerate cold extremely well.

The second principal component does not seem to provide as clear an explanation for the distribution of the taxa along the second axis. Interestingly, three taxa in the upper right corner (*Picea*, *Abies*, *Fagus*) increased in abundance in the late Holocene. Since also the uppermost taxon, *Artemisia*, has become more common in association with the human influence on the landscape, PC_2 could reflect the spatiotemporal trends of the vegetation. In other words, the taxa with high PC_2 values look like chronological "late arrivals", either due to being typically dominant late in the natural forest succession or due to the simultaneous increase in human impact on vegetation.

The patterns in the pollen data, especially the PC_1 interpreted as the warming signal, are the basis of the functionality of the climate reconstruction models. The reconstruction models have been calibrated to utilize such patterns to predict the paleotemperatures connected to each sample.

5.2 Reconstructed Holocene climate

As explained in the section 4.4, the reconstructed climate is presented as temperature anomalies (temperature difference relative to the modern day) at each location and time. To interpret the reconstruction visually, we present the data in an intuitive way to get a sense of the trends of the summer and winter temperatures across the Holocene, as well as assess the difference between the outputs of each model.

5.2.1 Average reconstruction across all models

As the first overview of the Holocene climate, we combine the results of all of the models. The climate change in different parts of the northern Europe is best visualized using a sequence of maps that present the mean temperature anomaly at each site during 2,000-year time windows across all six reconstruction models.

Figure 5.2 shows how the reconstructed summer (July) and winter (January) temperatures change as we approach the current day. Each dot represents a fossil pollen site, and the color of each site indicates the temperature anomaly from the modern day reference temperature. When the maps are examined in the context of the Holocene Temperature Conundrum, a thermal maximum around mid-Holocene shows up quite clearly. However, the January reconstructions show more regional variety. For example, the British Isles seem to be exhibiting warming trends totally different from the continental Europe.

The standard deviation of the temperature anomalies across all models quantifying the spread between model estimates for each time window is presented in Figure 5.3. There does not seem to be much temporal variation in the standard deviation but the difference is clear between July and January reconstructions. Since the January paleotemperatures have more deviation, anomalies in winter maps of Figure 5.2 have a larger spread.

5.2.2 Comparing the models

The average maps across all of the models present us with a good starting point for the overall trend, yet they do not reveal any information about how the paleotemperatures predicted by each model relate to each other. For example, based on them we cannot say, whether the large standard deviation across the January reconstruction is caused by different models not agreeing or because the temperatures vary a lot at different times during the time period.

We visualize the differences between the models in two ways. First, every temperature value given by every model (one for every pollen sample across

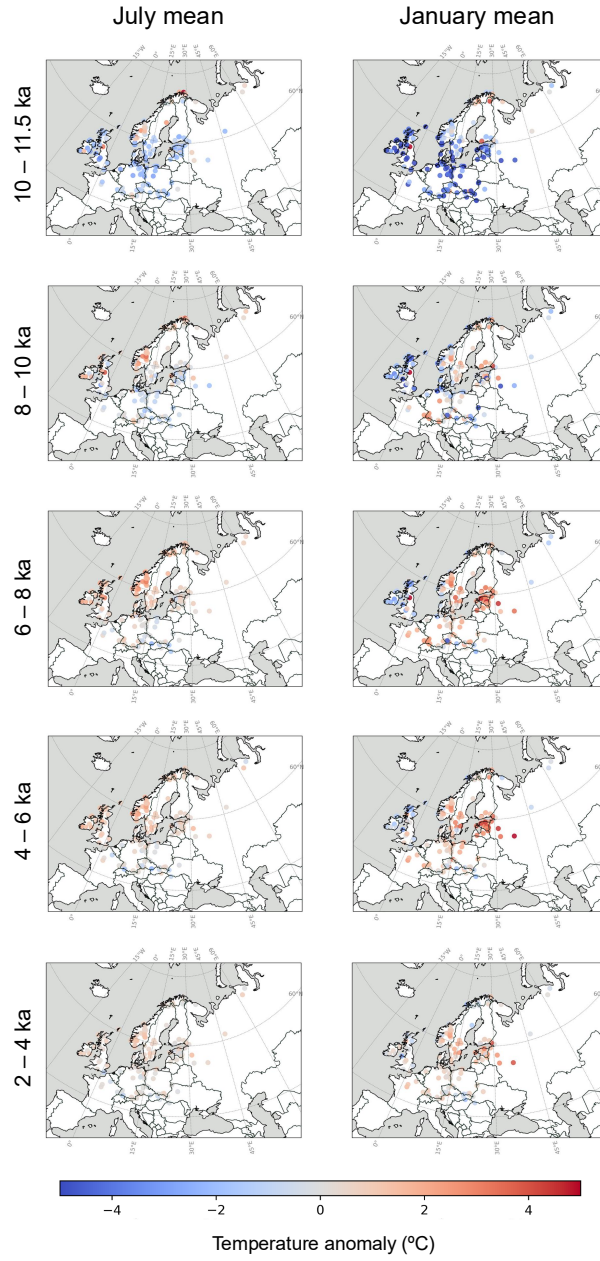


Figure 5.2: Mean across all six models (WA, WAPLS, MAT, RF, ERT, BRT) compared to the modern reference temperature and visualized in 2 ka intervals.

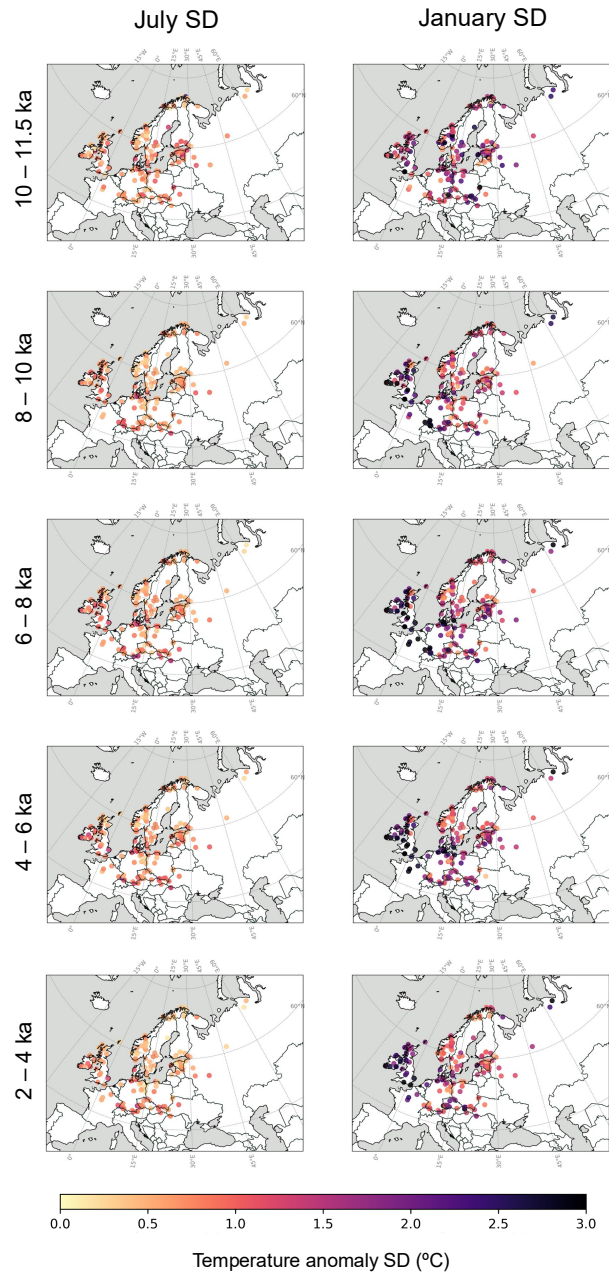


Figure 5.3: Standard deviation across all six models (WA, WAPLS, MAT, RF, ERT, BRT) compared to the modern reference temperature and visualized in 2 ka intervals.

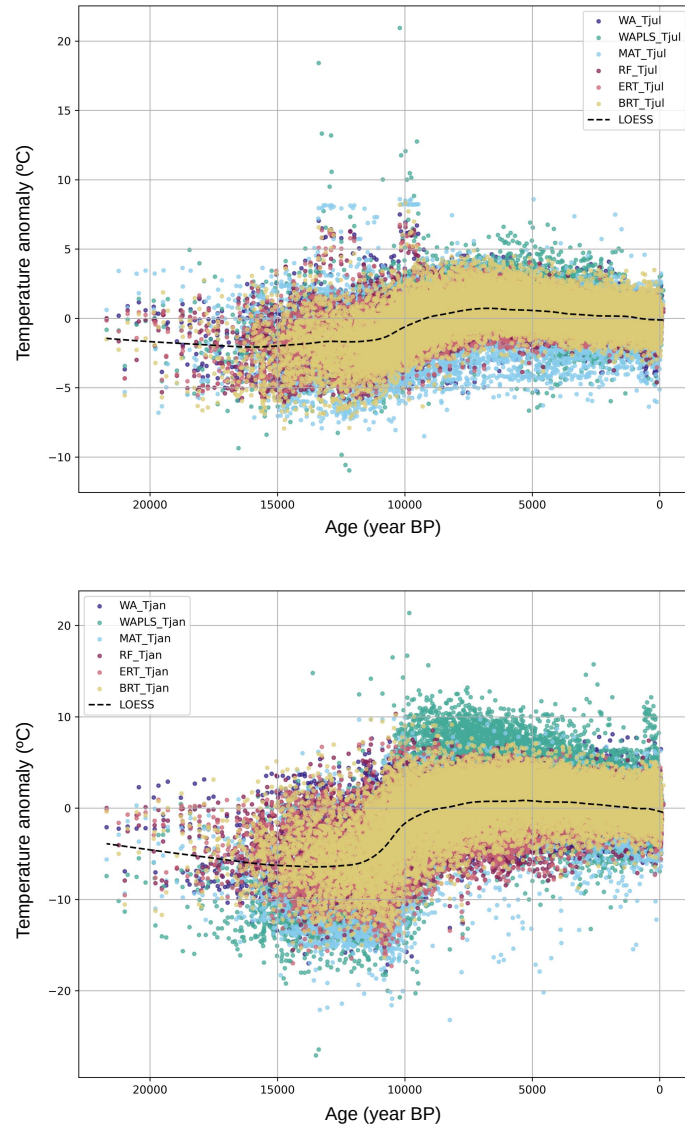


Figure 5.4: Summer (upper panel) and winter (lower panel) temperatures for every fossil pollen sample from every used reconstruction model.

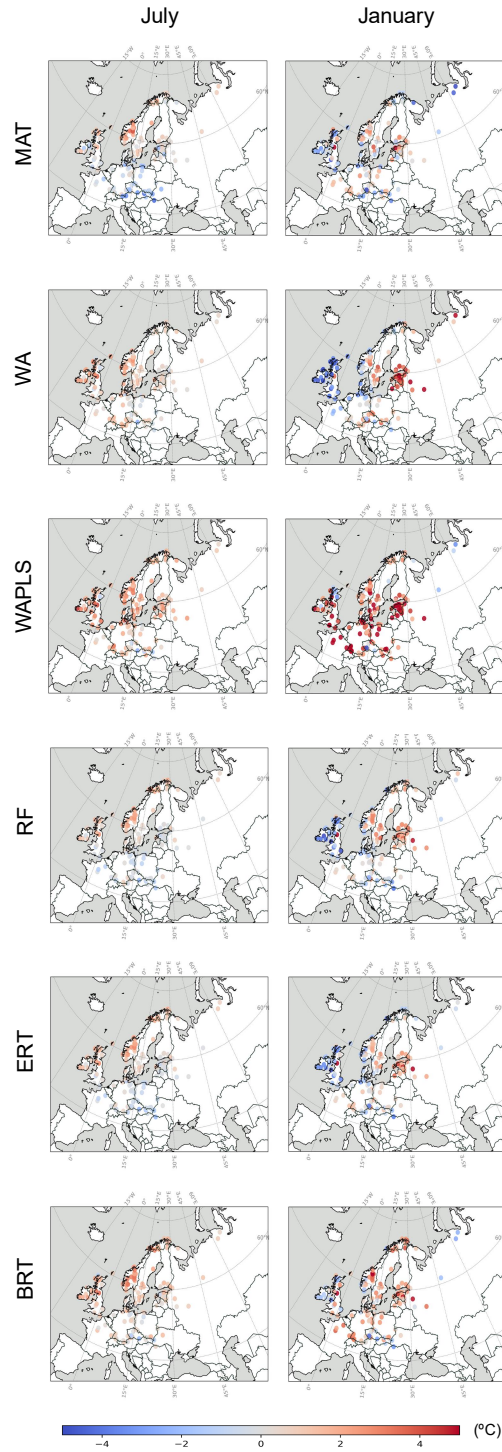


Figure 5.5: Temperature anomalies from each reconstruction model for the overall warmest period, 8-6 ka.

all sites) is plotted in Figure 5.4. The models are distinguished by color. In Figure 5.5, in turn, the overall warmest time range (8-6 ka) is mapped for each model.

The plots of Figure 5.4 (note that the Holocene starts around 11.7 ka!) show that the reconstruction models do not agree on the winter temperature values as consistently as they do for the summer temperatures. It seems like WAPLS may be an outlier, dragging the mean upwards by predicting higher temperatures than the rest of the models.

The same observation can be made by looking at the maps in Figure 5.5, where the January map of WAPLS shows radically different colors compared to the others. The rest of the maps, on the other hand, seem very similar, suggesting further that WAPLS is an outlier. For summer temperatures there does not seem to be any major conflict between the models, though WAPLS somewhat stands out again.

Assessing the reconstructions in the context of the performance each model had in Salonen et al. (2019), we may suggest a greater weight on the results of the best performed models, while noting that good performance in calibration and validation does not guarantee similar performance for data dated in the past. According to Table 3.3, WAPLS did outperform both WA and MAT, which does not support disregarding WAPLS from the reconstructions. However, if we choose to consider only BRT, ERT, and RF, the top three best performing models for both July and January, it seems the deviation across the models would drop significantly.

The visualizations alone are not sufficient to draw any comprehensive conclusions. More detailed analysis has to be made to understand the regional variability and differences in model performance. In doing so, not only the statistical but also ecological and geological context needs to be considered. Nevertheless, the substantially greater variation between models in the winter reconstruction compared to the summer reconstruction suggests that differences in model structure and in how seasonal climate signals are captured can significantly influence reconstructed temperature trends.

Chapter 6

Outcomes

In this study, we compared six models used in climate reconstructions and used fossil pollen as a proxy to reconstruct the July and January paleotemperatures with those models. Based on the resulting climate reconstruction, we analyzed the Holocene climate trend in northern Europe. This study consisted of model comparison, data preparation, principal component analysis of the proxy data, climate reconstruction, and visual analysis of the data and the reconstruction.

The reconstruction models were assessed based on the performance review of Salonen et al. (2019), as well as literature before the article. Data preparation and the visualization of the data patterns and reconstructed temperatures were implemented by Python programs presented in appendix A. The interpretation of the results was heavily reliant on interviewing J. Sakari Salonen.

6.1 Conclusions

The performance of the models chosen for this study (MAT, WA, WAPLS, RF, ERT, BRT) did not yield striking differences between them. In Salonen et al. (2019), all were found reliable in predicting the past climate and the results of the new reconstruction did not give grounds to conclude otherwise, even though WAPLS results were not aligned with the other models in the climate reconstruction.

Careful data preparation played a key role in successfully reconstructing the Holocene climate. The temporal and geographical constraints were clear and well justified to correspond the timeframe of interest (Holocene) and the data used in calibrating the reconstruction models.

Paleoclimatologically, the existence of the Holocene Thermal Maximum was

supported by the July temperature reconstructions. The January temperatures, on the other hand, showed more variety across the reconstruction models and opened potential ways to reassess the Holocene Temperature Conundrum. All in all, within the extent of our analysis, the question of the existence of a thermal maximum remains unanswered and would require a deeper dissection of the paleotemperatures given by each reconstruction model and across different regions.

6.2 Future research

Where this study ends, there starts the interesting part for the paleoclimatologists. The climate reconstruction provides massive amount of paleotemperature data to be analyzed beyond our first visualizations. From a methodological perspective, the distinct behaviour of the WAPLS model (especially in the winter reconstruction) would be an interesting question for further research. In general, considering biases originating from model choice could improve the reliability of future climate reconstructions.

Excluding WAPLS as an outlier from the January temperature reconstruction could be a natural next step to see if the average temperature line would turn convex and counter the peak in summer temperatures, potentially explaining the Holocene Temperature Conundrum. Another perspective could be to examine the climate trends of different regions that seem to have very different climate trends over the Holocene.

In terms of the quality of the proxy data, a stricter filter for the data and especially the radiocarbon datings would improve the data integrity and could yield somewhat different results. However, an attempt to maximize the data quality while preserving sufficient number of datapoints may not produce any better reconstructions than what we already have here.

Bibliography

- [1] Cajo J. F. ter Braak and Steve Juggins. “Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages”. In: *Hydrobiologia* (1993). URL: <https://doi.org/10.1007/BF00028046>.
- [2] Leo Breiman. “Random Forests”. In: (2001). URL: <https://doi.org/10.1023/A:1010933404324>.
- [3] Prabir Burman, Edmond Chow, and Deborah Nolan. “A cross-validated method for dependent data”. In: *Biometrika* (1994). URL: <https://doi.org/10.1093/biomet/81.2.351>.
- [4] Manuel Chevalier et al. “Pollen-based climate reconstruction techniques for late Quaternary studies”. In: *Earth-Science Reviews* (2020). URL: <https://doi.org/10.1016/j.earscirev.2020.103384>.
- [5] Glenn De’ath. “Boosted trees for ecological modeling and prediction”. In: *Ecology* (2007). URL: <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/0012-9658%282007%2988%5B243%3ABTFEMA%5D2.0.CO%3B2>.
- [6] J. Elith, J. R. Leathwick, and T. Hastie. “A working guide to boosted regression trees”. In: *Journal of Animal Ecology* (2008). URL: <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2656.2008.01390.x>.
- [7] K. Faegri et al. *Textbook of Pollen Analysis*. Blackwell Press, 2000.
- [8] Simon Goring et al. “Are pollen-based climate models improved by combining surface samples from soil and lacustrine substrates?” In: *Review of Palaeobotany and Palynology* (2010). URL: <https://www.sciencedirect.com/science/article/pii/S0034666710001338>.
- [9] *Holocene*, *Wikipedia*.
Retrieved: 19.11.2024. URL: <https://en.wikipedia.org/wiki/Holocene>.
- [10] Hannu Hyvärinen. *Siitepölyanalyysi ja kvartäärin siitepölystatigrafia Luoteis-Euroopassa*. Study material in the Geology department of University of Helsinki. 1986.

- [11] Stephen Jackson and Jonathan Overpeck. “Responses of plant populations and communities to environmental changes of the Late Quaternary”. In: *Paleobiology* (2000). DOI: 10.1017/S0094837300026932.
- [12] I. T. Jolliffe. *Principal Component Analysis*. Springer, 2002. ISBN: 978-1-4757-1904-8.
- [13] Steve Juggins. “Quantitative reconstructions in palaeolimnology: new paradigm or sick science?” In: *Quaternary Science Reviews* (2013). URL: <https://www.sciencedirect.com/science/article/pii/S0277379112005422>.
- [14] Steve Juggins, Gavin L Simpson, and Richard J Telford. “Taxon selection using statistical learning techniques to improve transfer function prediction”. In: *The Holocene* (2015). URL: <https://doi.org/10.1177/0959683614556388>.
- [15] Darrell Kaufman et al. “A global database of Holocene Data Descriptor paleotemperature records”. In: *Scientific Data* (2020). URL: <https://doi.org/10.1038/s41597-020-0445-3>.
- [16] Darrell Kaufman et al. “Holocene global mean surface temperature, a multi-method reconstruction approach”. In: *Scientific Data* (2020). URL: <https://doi.org/10.1038/s41597-020-0530-7>.
- [17] Pierre Legendre. “Spatial autocorrelation: trouble or new paradigm?” In: *Ecology* (1993).
- [18] Andy Liaw and Matthew Wiener. “Classification and Regression by randomForest”. In: *R News* (2002). URL: <https://CRAN.R-project.org/doc/Rnews/>.
- [19] Zhengyu Liu et al. “The Holocene temperature conundrum”. In: *Proceedings of the National Academy of Sciences* (2014). URL: <https://www.pnas.org/doi/abs/10.1073/pnas.1407229111>.
- [20] Jeremiah Marsicek et al. “Reconciling divergent trends and millennial variations in Holocene temperatures”. In: *Nature* (2018). URL: <https://doi.org/10.1038/nature25464>.
- [21] *Neotoma database*. URL: <https://www.neotomadb.org/>.
- [22] J. T. Overpeck, T. Webb, and I. C. Prentice. “Quantitative Interpretation of Fossil Pollen Spectra: Dissimilarity Coefficients and the Method of Modern Analogs”. In: *Quaternary Research* 23 (1985). DOI: 10.1016/0033-5894(85)90074-2.
- [23] *Radiocarbon dating*, *Wikipedia*. Retrieved: 19.11.2024. URL: https://en.wikipedia.org/wiki/Radiocarbon_dating.
- [24] J. Sakari Salonen et al. “Machine-learning based reconstructions of primary and secondary climate variables from North American and European fossil pollen data”. In: *Scientific Reports* (2019). URL: <https://doi.org/10.1038/s41598-019-52293-4>.
- [25] J. Sakari Salonen et al. “Reconstructing palaeoclimatic variables from fossil pollen using boosted regression trees: comparison and synthesis with other quantitative reconstruction methods”. In: *Quaternary*

- Science Reviews* (2014). URL: <https://www.sciencedirect.com/science/article/pii/S0277379114000183>.
- [26] Jaak Simm, Ildefons Magrans de Abril, and Masashi Sugiyama. “Tree-Based Ensemble Multi-Task Learning Method for Classification and Regression”. In: *IEICE Transactions on Information and Systems* (2014).
 - [27] John P. Smol, H. John B. Bricks, and William M. Last. *Tracking Environmental Change Using Lake Sediments*. Kluwer Academic Publishers, 2001.
 - [28] Erik Stokstad. “Myriad Ways to Reconstruct Past Climate”. In: *Science* (2001). URL: <https://www.science.org/doi/abs/10.1126/science.292.5517.658>.
 - [29] R.J. Telford and H.J.B. Birks. “Evaluation of transfer functions in spatially structured environments”. In: *Quaternary Science Reviews* (2009). URL: <https://www.sciencedirect.com/science/article/pii/S0277379108003806>.
 - [30] Chengyu Weng, Henry Hooghiemstra, and Joost Duivenvoorden. “Challenges in estimating past plant diversity from fossil pollen data: Statistical assessment, problems, and possible solutions”. In: *Diversity and Distributions* (2006). DOI: 10.1111/j.1366-9516.2006.00230.x.

Appendix A

Code and data

The scripts used for the data preparation, principal component analysis, and data visualization were implemented utilizing the existing Python libraries and tools. The reconstruction models were run with R as described in Salonen et al. (2019).

The code written for this thesis can be found in Github: <https://github.com/ellikiiski/Fossil-pollen-climate-reconstruction-tools/>.

In the same repository (link above), there are CSV files available for

- Harmonized fossil pollen data from all sites
- Metadata of the fossil pollen sites
- Reconstructed temperature predictions of all sites and reconstruction models.