A quantile regression analysis on the impact of climate change on the seasonal pollen release in Sweden

Max Brehmer

2023-04-20

Abstract

Preface

Introduction

Over the past couple of decades it has been made clear that our climate is changing rapidly in various ways. Most notably the global average temperature has risen by ca $0.2°C$ per decade since the mid seventies, this constitutes an almost $1°C$ increase over the past half century [@hansen2006global]. In the same time frame Sweden has also seen significant shifts in its otherwise stable and temperate climate. Several studies in recent decades draw the conclusion that plant phenology is impacted by this increase in temperature. [@van2002influence] discusses in their study of the seasonal pollen shift in the Netherlands that an advance in the start of the pollen season by 3-22 days took place in the latter third of the 20th century. Likewise this paper strives to understand what seasonal changes have occurred to the pollen season in Sweden.

As mentioned in [@lind2016pollen] the results may differ for various species of pollen. More precisely they found a stark difference in duration among arboreal plant species compared to herbaceous ones, with the former trending towards an earlier end date, while the latter was pushed to a further date and thus have a longer seasonal duration. Grass pollen, being herbaceous, is the leading cause of pollen allergy in many developed countries, meaning a lot of people suffer from these seasonal changes for an extended time [@garcia2017poaceae]. In Sweden and other parts of northern Europe however, due to differences in temperature and overall climate, the arboreal types like birch (betula) are the most common cause of pollinosis [@d2017allergenic].

Continous monitoring of pollen conducted by the Swedish Museum of Natural History (NRM) began in 1973 at the Palynological laboratory in Stockholm. Since then multiple other stations have been included in the scope of NRM's continous pollen monitoring program. As of 2022 there are 20 active stations involved [@nrm2022pollen], monitoring the release of several unique species of pollen. In this paper we consider 7 of the most allergenic species, these are the arboreal pollen of alder (alnus), birch (betula), hazel (corylus), oak (quercus), willow (salix), elm (ulmus) and the herbaceous species of grass pollen (poaceae).

In this paper we will attempt to determine the shift in dates of the start and end of the pollen season in Sweden as an effect of global warming of Earth's climate. We will consider global warming as a linear trend over the researched time period as to simplify the process of analyzing pollen patterns. We can do this as research has shown acceptable fitting of linear models over anthropogenic climate change, in regard to temperature [@hansen2006global]. An analysis will be conducted based on two separate frequentist quantile regression models, namely linear regression on empirical quantiles (EQ) and non-parametric quantile regression (QR). In a study of seasonal shifts of migratory birds [@karlsson2022comparison] perform both these methods, this paper covers the majority of the theory in regards to the construction of the statistical models. In the case of QR, we also make good use of [@takeuchi2006nonparametric] for a more in depth description of the method.

By its conclusion this research paper aims to have built a statistical model that can explain the historic shift in pollen seasons for each of the 7 species and also possesses the ability to predict expected further changes

in the Swedish pollen season.

This paper begins with the `Prior research` section by presenting the findings of what we consider to be the most relevant studies of similar character to what we envisage our own research to look like.

The `Data` section is where all aspects of the data used in the research is presented. We cover these aspects in three smaller sections. Firstly `Data collection` where a light description of both the physical and digital methods used to collect the data. An extensive description of how we use the data is later presented in the section `Understanding the data`. Here the reader is granted a look at the structure of the data in the form of easily-to-read tables and which variables are to be considered in the research. We conclude the `Data` section with `Data selection` where an explanation of how we deal with missing values is conducted.

Section X `Theory` encompasses all the theoretical definitions of the methods included in this paper. In order to strengthen the significance of our results, we have used two independent methods. Linear regression on empirical quantiles and Nonparametric quantile regression. Linear regression on empirical quantiles, which we refer to as EQ, is a simple method commonly used to perform statistical analyses on different quantile levels of data, while nonparametric quantile regression (QR) is a more complex method which substitutes the conditional mean of the loss function for the conditional median and requires numerical approaches to the resulting optimization problem. Our sections on EQ and QR respectively explain the theoretical aspects of these methods, while simultaneously providing descriptions of how we have applied them in our models.

A section containing the most important results of our models is given in `Results`. This part of the paper is divided in to 2 sections for increased readability. `Plots` contains visual descriptions of how the pollen season behaves over time, for selected combinations of location and species. The remaining figures are available in the appendix. `Tables` is where the key values are presented in the form of ANOVA tables and alike. These values include regression coefficients, p-values and R^2 for each model. Again, as for the plots, results for all combinations of data are not presented in this section. Rather a selection of the most significant, representative and most extreme results are, while remaining tables can be found in an appendix.

The `Discussion` is the section in which we attempt to make sense of the results observed and visualized in the `Results` section. For each of the analyzed results we may be able obtain a definitive conclusion, or not, in which case a description of what we are missing for us to draw a conclusion follows. Since there are limits to what we can expect to answer in this thesis, we have set aside a section (`Further improvements`) on how, and which parts of our work can be improved upon.

This paper concludes with an `Appendix` section. Additional or complementary information which the reader may find interesting, but not vital for the results of which we form our conclusions. Each appendix is reffered to by, and refers back to a result or explanation.

Litterature review

Data

Data collection

As we mentioned earlier in this paper, the monitoring of pollen in the Stockholm region is conducted by the Palynological laboratory at NRM. The laboratory in question uses a Burkard Seven Day Volumetric Spore Trap to capture pollen and spores from the air through a small entrance meant to resemble the human airways. Thus approximately 10 liters of air passes through the machine each minute, which is what humans tend to consume. In order to capture the pollen particles carried by the passing air, a sticky tape is mounted to a drum rotating at 2 mm per hour. As only a small portion of the tape is exposed to the air at each point in time, this method grants us a good indication of the volume of pollen in the passing air at any given moment. Each captured pollen is individually counted with regular intervals using microscopes. It must be noted that not all stations possess the same equipment. In particular, differing microscope sizes are used across the country. Consequently, the measured values of the pollen counts are biased towards the larger microscopes, thus showing a somewhat inaccurate representation of the true pollen counts [@nrm2022microscope] [@nrm2022pollen2]. However considering the structure of the dataset and the

consequent data analysis being relativistic, for which a descriptive presentation follows in the `Data` section, this phenomenon has been ignored.

Understanding the data

The dataset that we have at our disposal contains 5 unique variables: `date`, `station`, `name`, `count` and `factor`. Of which all but the `factor` variable are used in this research paper. We have also added a `latitude` variable since it is known that higher latitudes contribute to more extreme climate changes [alecrim2023higher]. This variable is however entirely dependent on `station`. A short description of the meaning of each variable is shown in figure ?. Data is recorded during the predicted pollen season, based on historic results.

Variable	Type	Decription
Station	categorical	Geographic location of the pollen monitoring station.
Species	categorical	Genus of the recorded pollen counts.
Date	continous	Gregorian calendar date on which the airborne pollen were registered.
Count	continous	Number of individual pollen counted.
Factor	continous	Reference variable for the size of the microscope used.
Latitude	continous	Northern latitudinal cooardinates of said station.

Viewing the years of availability in figure ? we conclude that not all data points are present in the data set. The stations had differing opening dates and not all species tend to be available to begin with. If no consideration for the location of said data points are made, we may observed skewed results, as the geographic distribution of monitored pollen changes over time due to availability. Thus analyzing the data in geographic categories of where they were collected is a necessary consideration.
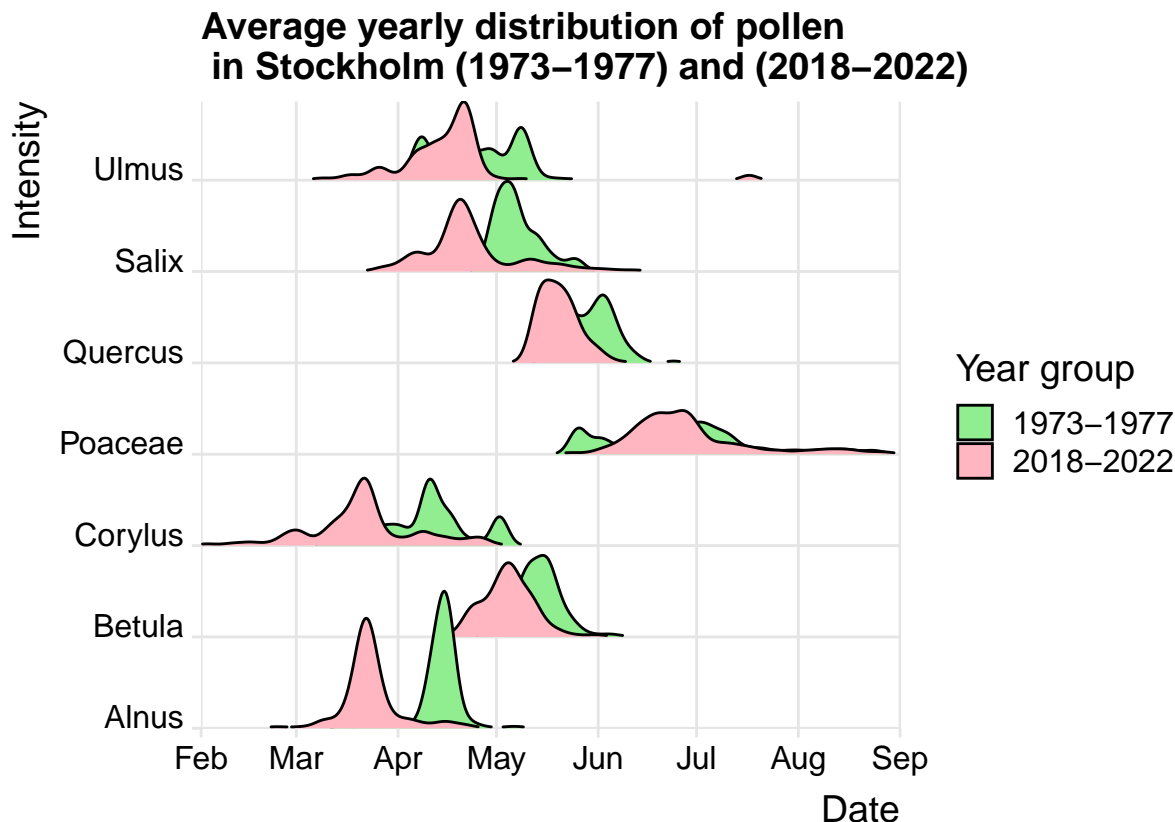
Station	Latitude	Pollen genus (since ...)
Umeå	62.83	Alnus, Betula, Poaceae, Ulmus (1979), Salix (1981), Corylus (1987), Quercus (1995)
Eskilstuna	59.37	Alnus, Betula, Corylus, Poaceae, Quercus, Salix, Ulmus (1976)
Stockholm	59.33	Alnus, Betula, Corylus, Poaceae, Quercus, Ulmus (1973), Salix (1977)
Norrköping	58.59	Alnus, Betula, Corylus, Poaceae, Quercus, Salix, Ulmus (1987)
Jönköping	57.78	Alnus, Betula, Poaceae, Quercus, Salix, Ulmus (1988), Corylus (1989)
Västervik	57.76	Alnus, Betula, Corylus, Poaceae, Quercus, Salix, Ulmus (1987)
Göteborg	57.71	Alnus, Betula, Corylus, Poaceae, Quercus, Salix, Ulmus (1979)
Malmö	55.60	Alnus, Betula, Corylus, Poaceae, Quercus, Salix, Ulmus (1979)

Data selection

First we remove any observations with missing values in any of the columns `date`, `station` or `species` since these are fundamental parameters to perform the data analysis. For the purposes of EQ regression, the time it takes to compute the models on the entire dataset in `R` is negligible. For nonparametric quantile regression however, the computational time appears to grow very quickly with the amount of observations. This has led us to the decision to reduce the amount of content in each dataset (one dataset containing all observations of any combination of `station` and `species`) to below a fixed limit. Without reducing any data, the largest data sets contain ≈ 500000 observations, while some combinations of data contain considerably fewer. By testing and optimizing the models for various sizes of the datasets we decided to set the limit to 5000 observations per combination of data. This reduces the running time considerably, making the process of analyzing our results much easier. In order to minimize the affect this has on our results the rows must be removed uniformly over the time of year. If the number of observations in a dataset is smaller than $n \cdot 5000$, we want to select every nth element and discard the rest. We can do this in `R` by using the `slice` and `bind_rows` functions from the `dplyr` package. By performing these modifications to our data, we are sure to keep the shape of the distributions the same and thus the results of the quantile regression.

Yearly distribution of pollen

Before constructing the models, we can take a glance at what sort of results to expect from the subsequent regression analysis. In figure ? we present a ridge plot containing information about the distribution of observed pollen by date of year. By looking at the average distribution from the first and last 5 years in the dataset respectively, we construct a somewhat consistent average in terms of the state of the pollen season at a given location, reducing the risk of falsely identifying outlier data as patterned occurrences. We can conclude that, in the Stockholm area, the pollen season seemed to begin earlier for all the arboreal pollen species while grass pollen did not appear to show those tendencies. In general the shape of the pollen distribution was not seen to have been altered over time. A visual analysis of the distribution of yearly pollen release may suggest that the average pollen season now begins earlier by approximately 10-20 days compared to 45 years ago.

Average yearly distribution of pollen in Stockholm (1973–1977) and (2018–2022)

Year group
■ 1973–1977
■ 2018–2022

Theory

In this section we explain the underlying theory behind the statistical approaches used for our data analysis. We will compare the simpler empirical quantile linear model (EQ), as used in [karlsson2022comparison] with the supposedly more powerful nonparametric quantile regression which we will refer to as QR. Both of these approaches estimate the response variable as the conditional median or other quantiles of the predictor variables as opposed to ordinary linear regression which uses the method of ordinary least squares (OLS) to estimate the conditional mean [Wikipedia].

OLS regression is the preferred method for many research purposes due to its inherent simplicity. In this research paper however, we are more interested in patterns for certain quantiles of data than the mean. To do this effectively we have conducted our research using quantile regression methods instead.

An aspect in which quantile regression performs better than linear regression is when data is not homoscedastic or normally distributed. Linear regression models perform poorly for data that is heteroscadastic and/or non-normally distributed. By estimating the conditional median however, as opposed to the mean, we get a model that yields better predictions for data with these properties. Another advantage of using quantile

based methods is that one can make more accurate predictions about outliers, since nonlinear tendencies may lead to abnormal behaviors for more extreme observations, which can be more accurately accounted for when looking at quantiles rather than the mean.

All models in this paper will use `date` as the response variable, or a yearly quantile of `date`. The release date of pollen grain $i \in \{1, ..., n\}$, where n is the total amount of pollen grains released, is predicted by the models as the response y_i. We form a covariate vector $x_i = (1, t_i)$ which includes an intercept and the predictor variable $t_i \in \mathcal{T} = \{1, ..., T\}$ which refers to the gregorian year of observation with T being the amount of years monitored.

In order to make an educated analysis of the behaviour of pollen release we need a strict definition for what constitutes a pollen season. There have been many attempts by various authors to find an optimal definition for the dates of which each pollen season covers. In this paper we classify the dates within which the pollen season is deemed active by referring to the EAN definition. [bastl2018defining] The EAN database contains a lot of data on pollen release in which they define the pollen season as the date at which at least 1% of the yearly pollen have been counted. By the same definition, the pollen season is said to end when 95% of all pollen has been released.

Nonparametric quantile regression

By nonparametric regression models we refer to models that do not make any parametric assumptions about the form of the conditional distribution function, the relationship between the response variable and predictor variable(s) are linear however. By applying the theory of nonparametric models to a linear predictor of the conditional median, we arrive at the basis of a nonparametric quantile regression model.

For a univariate nonparametric QR model, consider the observed response date y_i of grain i and the predictor variable year t_i with an intercept in the $(n \times 2)$-matrix X.

$$X = \begin{pmatrix} 1 & 1 & ... & 1 \\ t_1 & t_2 & ... & t_n \end{pmatrix}$$

Let Y_i be a stochastic variable corresponding to the observed value of y_i with the conditional distribution function

$$F_{Y_i|x_i}(y) = \mathbb{P}(Y_i \leq y \mid x_i), \quad -\infty < y < \infty \quad (?)$$

In order to find the τ-quantile of the conditional distribution of $Y_i \mid x_i$ we need to find the smallest value of y such that the conditional cumulative distribution function is greater or equal to τ.

So, for each quantile $0 < \tau < 1$ the inverse

$$Q(\tau \mid x_i) = \inf\{y; F_{Y_i \mid x_i}(y) \geq \tau\} \quad (?)$$

of the conditional distribution function in (?) represents the conditional quantile function.

Moving forward, our model takes the form

$$Q(\tau \mid X) = X\beta(\tau) + \varepsilon(\tau) \quad (?)$$

whereby the column vector $\beta(\tau) = (\beta_0(\tau), \beta_t(\tau))^\top$ contains the intercept $\beta_0(\tau)$ and year as the slope $\beta_t(\tau)$ of quantile τ. We define $\varepsilon(\tau)$ as a non-parameterised vector of the error terms.

An optimization problem is then constructed as to minimize the objective function in (?). In ordinary linear regression one typically uses the method of least squared errors (LSE) to find the model that leads to the lowest amount of loss of information. In our case we use the pinball loss function $l_\tau(\xi)$ introduced in [koenker2005qr] as a metric to determine the accuracy of a quantile estimate [takeuchi2006nonparametric]. The pinball loss function is given by

$$l_\tau(\xi) = \begin{cases} \tau\xi & \text{if } \xi \geq 0 \\ (\tau - 1)\xi & \text{if } \xi < 0 \end{cases} \quad (?)$$

hence the resulting regression parameter estimates of the τ-quantile is given by the solution to the minimization problem

$$\beta(\tau) = \arg\min_{b \in \mathbb{R}^2} \sum_{i=1}^{n} l_\tau(y_i - x_i b). \quad (?)$$

Since this function is non-differentiable at 0, we cannot use... Rather we use numerical estimation methods provided in the R library `quantreg` [koenker2021rpackage] to find the optimal point along the ξ-axis.

Linear regression on emprical quantiles

Another approach we used is to perform linear regression on empirical quantiles, a statistical method that combines linear regression with quantile regression to model the relationship between a response variable and one or more predictor variables. The basic gist of EQ regression is to fit a linear regression model to the empirical quantiles of the response variable, rather than to the mean. This method is called "empirical" quantile regression because it estimates quantiles of the response variable based on the empirical distribution of the data, rather than assuming a specific distribution for the response variable.

One way to perform empirical quantile linear regression is by using an indicator function $I(\cdot)$. The indicator function is a mathematical function that takes a value as input and returns 1 if the value satisfies a certain condition and 0 otherwise. In the context of empirical quantile linear regression, the indicator function is used to define the estimation problem in terms of a set of linear programming (LP) constraints.

Rather than using the raw gregorian date y_i as response variable, we predict an empirical quantile of the dates of a subset of observations. For each $t \in \mathcal{T}$ we extract the set of observations

$$\mathcal{Y}_t = \{y_i : I(t_i = t)\} \quad (?)$$

As in the previous method we construct the $(n \times 2)$ matrix X by stacking each of the covariates $(1, t)$ on top of each other. Let $\tau \in (0, 1)$ be a quantile. For each of our sets $\mathcal{Y}$ we let $\hat{F}_{(x,t)}$ be the empirical distribution function formed by the elements of its set. The corresponding empirical quantile is defined as

$$\hat{Q}_{(t)}(\tau) = \inf\left\{y \in \mathcal{Y}_t : \hat{F}_t(y) \geq \tau \mid \mathcal{Y}_t \neq \emptyset\right\}. \quad (?)$$

For all $t \in \mathcal{T}$, we stack these quantiles into the vector of observations $Y(\tau)$. Next we formulate the linear model

$$Y(\tau) = X\beta(\tau) + \varepsilon(\tau). \quad (?)$$

As before $\beta = (\beta_0, \beta_t)$ defines the regression parameters for the intercept β_0 and year β_1. Recall that in the nonparametric method, no assumptions were made about the error terms. In this approach however, we assume a normal distribution of the error terms $\varepsilon(\tau) \sim N(0, \sigma^2(\tau)I_T)$, where I_T is the identity matrix of rank T.

The log-likelihood (omitting τ) of the model is given by

$$l(\beta, \sigma^2 \mid Y, X) = \sum_{t \in \mathcal{T}} \log f\left(\hat{Q}_t(\tau) \mid (X\beta)_t, \sigma^2\right) \quad (?)$$

This model grants each year t the same weight, not each individual pollen. To get even weighting of each pollen, we simply add a weight factor $w_t = |\mathcal{Y}_t|$ to each empirical quantile $\hat{Q}_t(\tau)$. The reweighed log-likelihood (?), again omitting τ from the notation, takes the form

$$l_w(\beta, \sigma^2 \mid Y, X) = \sum_{t \in \mathcal{T}} w_t \log f\left(\hat{Q}_t(\tau) \mid (X\beta)_t, \sigma^2\right), \quad (?)$$

To proceed fitting the model, we attempt to find the maximum likelihood estimate (MLE) of $\beta(\tau)$ and $\sigma^2(\tau)$ by optimizing

$$MLE(\beta(\tau), \sigma^2(\tau)) = \arg\max l_w(\beta(\tau), \sigma^2(\tau) \mid Y(\tau), X) \quad (?)$$

for a given quantile τ. Since the function is continuous and twice differentiable, the solution can easily be found using conventional methods provided in the `quantreg` package.

Results

Results per species and location

Table 3: Start of pollen season for different species and locations, by most significant shift to earlier dates (EQ model)

Species	Location	Coefficient (EQ)	P value (EQ)	Predicted 1973 (EQ)	Predicted 2023 (EQ)
Corylus	Umeå	-1.362	0.009	141	73
Quercus	Umeå	-1.237	0.136	193	131
Corylus	Norrköping	-1.178	0.000	107	48
Alnus	Umeå	-1.001	0.000	121	71
Corylus	Eskilstuna	-0.995	0.000	102	52

Table 4: Start of pollen season for different species and locations, by most significant shift to later dates (EQ model)

Species	Location	Coefficient (EQ)	P value (EQ)	Predicted 1973 (EQ)	Predicted 2023 (EQ)
Poaceae	Västervik	0.184	0.158	141	150
Betula	Jönköping	0.055	0.748	110	113
Poaceae	Eskilstuna	0.007	0.943	148	148
Poaceae	Umeå	-0.024	0.848	165	164
Betula	Västervik	-0.035	0.785	111	110

Table 5: Peak of pollen season for different species and locations, by most significant shift to earlier dates (EQ model)

Species	Location	Coefficient (EQ)	P value (EQ)	Predicted 1973 (EQ)	Predicted 2023 (EQ)
Corylus	Umeå	-1.387	0.005	147	77
Quercus	Umeå	-1.034	0.207	191	140
Alnus	Umeå	-0.967	0.000	130	82
Corylus	Norrköping	-0.851	0.002	112	70
Alnus	Eskilstuna	-0.613	0.000	108	77

Table 6: Peak of pollen season for different species and locations, by most significant shift to later dates (EQ model)

Species	Location	Coefficient (EQ)	P value (EQ)	Predicted 1973 (EQ)	Predicted 2023 (EQ)
Poaceae	Västervik	0.135	0.362	174	181
Ulmus	Norrköping	0.067	0.859	111	114
Salix	Västervik	-0.052	0.780	114	111

Species	Location	Coefficient (EQ)	P value (EQ)	Predicted 1973 (EQ)	Predicted 2023 (EQ)
Poaceae	Umeå	-0.077	0.269	194	191
Quercus	Jönköping	-0.096	0.413	147	142

Table 7: End of pollen season for different species and locations, by
most significant shift to earlier dates (EQ model)

| Species	Location	Coefficient (EQ)	P value (EQ)	Predicted 1973 (EQ)	Predicted 2023 (EQ)
Corylus	Umeå	-1.398	0.011	157	87
Alnus	Umeå	-0.853	0.000	141	98
Corylus	Norrköping	-0.806	0.000	126	85
Alnus	Eskilstuna	-0.726	0.000	125	89
Ulmus	Jönköping	-0.690	0.064	148	113

Table 8: End of pollen season for different species and locations, by
most significant shift to later dates (EQ model)

Species	Location	Coefficient (EQ)	P value (EQ)	Predicted 1973 (EQ)	Predicted 2023 (EQ)
Poaceae	Västervik	1.028	0.000	180	231
Poaceae	Norrköping	0.472	0.031	197	221
Salix	Västervik	0.373	0.061	127	146
Poaceae	Göteborg	0.328	0.014	209	226
Corylus	Göteborg	0.312	0.518	103	118

Geographic influence
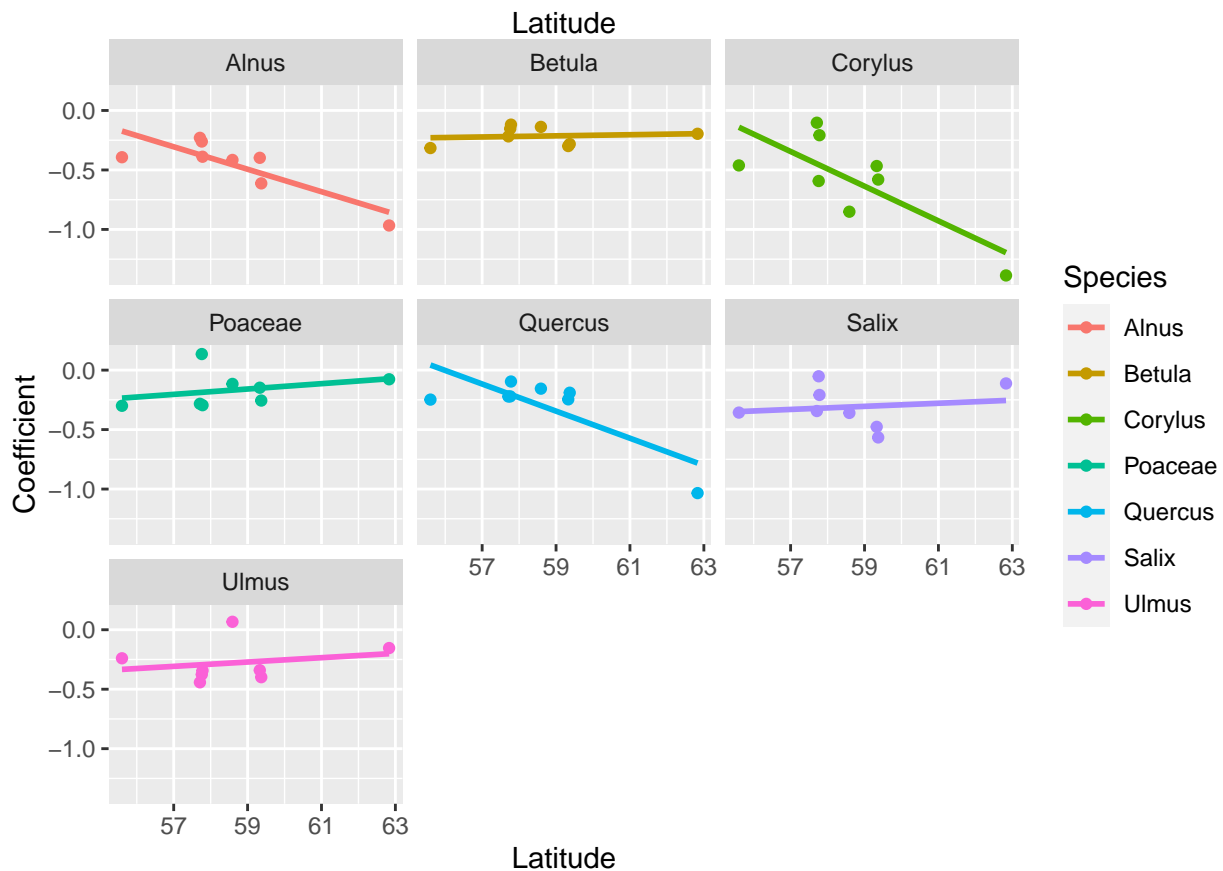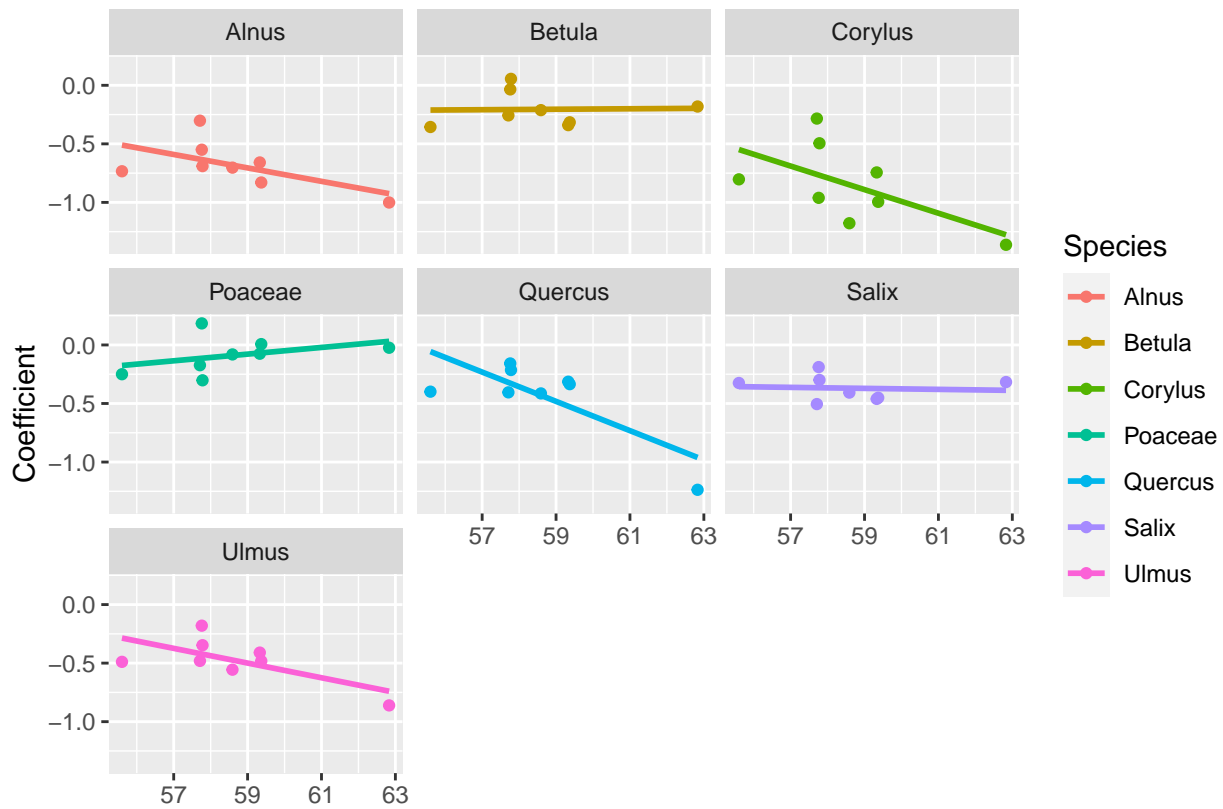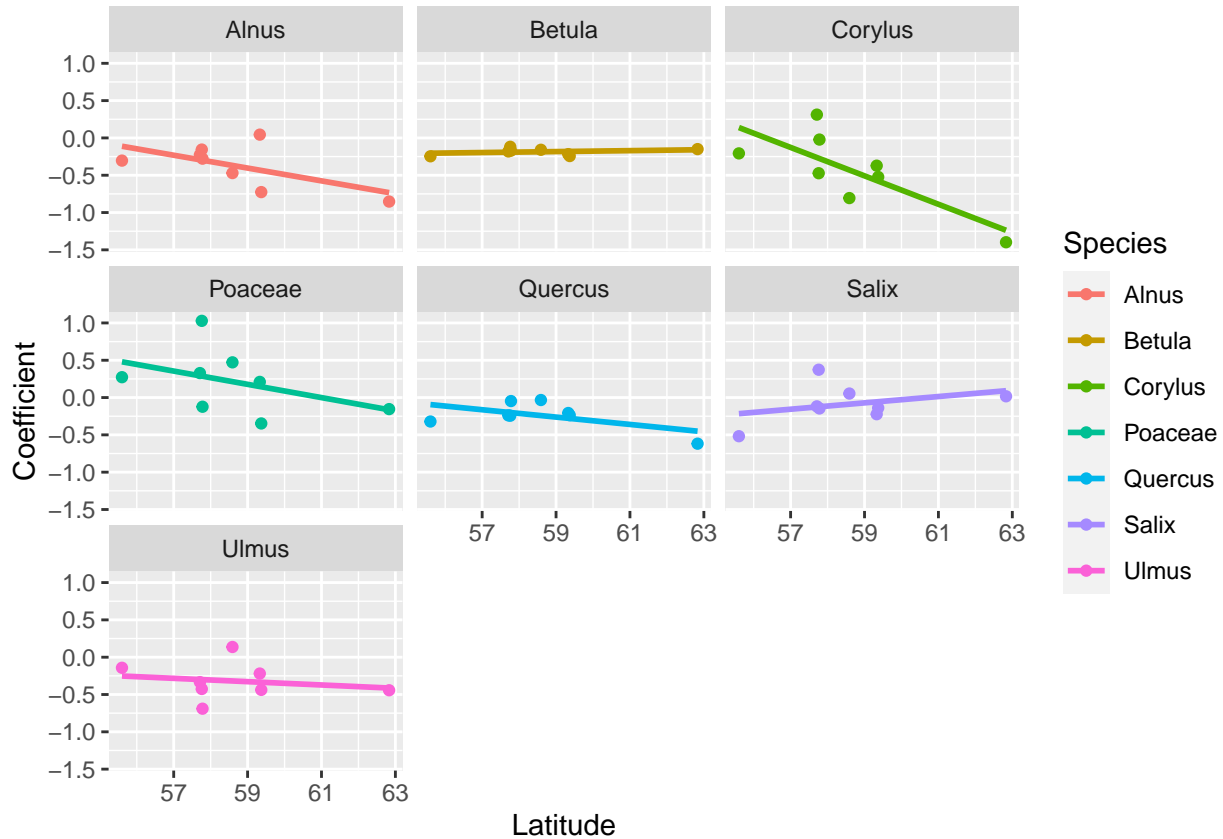
Discussion

Further improvements

Bibliography and References

Appendix

Appendix 1: Translation of the latin names of pollen species

Latin name	English name	Swedish name
Alnus	Alder	Al
Betula	Birch	Björk
Corylus	Hazel	Hassel
Poaceae	Grass	Gräs
Quercus	Oak	Ek
Salix	Willow	Viden
Ulmus	Elm	Alm

Appendix 2: Information about reduced data sets (ignore)

ESKILSTUNA: ALNUS: 45 362 / 10 < 5 000 (10) BETULA: 451 397 / 91 < 5 000 (91) CORYLUS: 3 812
POACEAE: 48 265 / 10 < 5 000 (10) QUERCUS: 57 914 / 12 < 5 000 (12) SALIX: 31 227 / 7 < 5 000 (7)
ULMUS: 9 779 / 2 < 5 000 (2)

GÖTEBORG: ALNUS: 33 653 / 7 < 5 000 (7) BETULA: 494 464 / 99 < 5 000 (99) CORYLUS: 3 374
POACEAE: 74 372 / 15 < 5 000 (15) QUERCUS: 54 957 / 11 < 5 000 (11) SALIX: 19 285 / 4 < 5 000 (4)
ULMUS: 11 645 / 3 < 5 000 (3)

JÖNKÖPING: ALNUS: 42 217 / 9 < 5 000 (9) BETULA: 239 006 / 48 < 10 000 (48) CORYLUS: 3 594 POACEAE: 31 676 / 7 < 5 000 (7) QUERCUS: 26 482 / 6 < 5 000 (6) SALIX: 13 929 / 3 < 5 000 (3) ULMUS: 4 884

MALMÖ: ALNUS: 31 596 / 7 < 5 000 (7) BETULA: 203 475 / 41 < 5 000 (41) CORYLUS: 4 562 POACEAE: 87 912 / 18 < 5 000 (18) QUERCUS: 42 482 / 9 < 5 000 (9) SALIX: 17 940 / 4 < 5 000 (4) ULMUS: 35 560 / 8 < 5 000 (8)

NORRKÖPING: ALNUS: 24 089 / 5 < 5 000 (5) BETULA: 286 630 / 58 < 5 000 (58) CORYLUS: 4 484 POACEAE: 50 537 / 11 < 5 000 (11) QUERCUS: 26 899 / 6 < 5 000 (6) SALIX: 14 715 / 3 < 5 000 (3) ULMUS: 6 509 / 2 < 5 000 (2)

STOCKHOLM: ALNUS: 55 346 / 12 < 5 000 (12) BETULA: 372 668 / 75 < 5 000 (75) CORYLUS: 5 848 / 2 < 5 000 (2) POACEAE: 52 569 / 11 < 5 000 (11) QUERCUS: 81 161 / 17 < 5 000 (17) SALIX: 15 398 / 4 < 5 000 (4) ULMUS: 33 284 / 7 < 5 000 (7)

UMEÅ: ALNUS: 16 454 / 4 < 5 000 (4) BETULA: 212 232 / 43 = 5 000 (43) CORYLUS: 113 POACEAE: 22 639 / 5 < 5 000 (5) QUERCUS: 182 SALIX: 8 875 / 2 < 5 000 (2) ULMUS: 167

VÄSTERVIK: ALNUS: 31 722 / 7 < 5 000 (7) BETULA: 231 147 / 24 < 10 000 (47) CORYLUS: 4 254 POACEAE: 37 613 / 8 < 5 000 (8) QUERCUS: 40 942 / 9 < 5 000 (9) SALIX: 8 127 / 2 < 5 000 (2) ULMUS: 9 821 / 2 < 5 000 (2)