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

Max Brehmer

2023-05-17

Abstract

It is well established that climate change has led to significant changes in the timing and intensity of phenological events around the world. One such event is the release of pollen by plants, which is a major cause of allergies in humans. This study aims to investigate the impact of climate change on the timing and intensity of pollen release in Sweden. We also look at whether or not the latitude of plants has a significant influence on seasonal pollen release.

Using data provided by the Swedish Museum of Natural History, we compare nonparametric quantile regression and linear regression on empirical quantiles to model the changes in pollen season. We find that empirical quantiles produce better results. Our findings show that the start and peak dates of various species of pollen advance at a similar rate over time in a warming climate while the ending dates move at a slower rate. The amount of days a year the pollen season is deemed to be active thus increases over time. Furthermore, the effects mentioned appear to be intensified at higher latitudes, apart from the extension of seasonal length for grass pollen, which was found to be reduced at more northerly latitudes.

To conclude, our findings suggest that climate change is having a significant impact on the timing and intensity of pollen release in Sweden, which is of concern for reasons of public health and potentially for wildlife and agriculture, thus it is important to monitor these changes in pollen release patterns and develop strategies to mitigate their impact.

Keywords: Pollen, phenology, climate change, quantile regression, linear regression, empirical quantiles.

Preface

I have conducted this research paper as my bachelor's thesis in mathematical statistics at Stockholm University. I want to thank my advisor Martin Sköld who has been of great value throughout this semester by providing me the guidance necessary to complete this thesis.

I must acknowledge that I have used artificial intelligence tools. More specifically OpenAI's ChatGPT has been a valuable tool for gaining a deeper understanding of certain concepts. However I have not used any AI tools to produce any of the contents in my texts, nor has it been used as a source for any in-text information.

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 (Hansen et al. 2006). Limiting our view to only Sweden, we also see significant shifts in this regions' otherwise stable climate. Several studies in recent decades draw the conclusion that plant phenology is impacted by this increase in temperature. (Van Vliet et al. 2002) discusses in a 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.

Results may differ for various species of pollen. (Lind et al. 2016) 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 (García-Mozo 2017). 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 (D'Amato et al. 2017).

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 (Natural History 2023), 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 (Hansen et al. 2006). An analysis will be conducted based on two separate frequentist quantile regression models, namely linear regression on empirical quantiles (EQ) and nonparametric quantile regression (QR). In a study of seasonal shifts of migratory birds, (Karlsson and Hössjer 2022) performs both these methods. This paper covers the majority of the theory in regards to the construction of the statistical models. We also make good use of (Takeuchi et al. 2006), a paper on nonparametric quantile estimations, for a more in depth description of the QR 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.

We begin our research with the section **1. Litterature review**, where we present 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 following section is **2. Data**, in which all aspects of the data used in the research is presented. We cover these aspects in four smaller sections. Firstly **2.1 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 **2.2 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. In **2.3 Data transformation** we follow with a description of how the data has been restructured. **2.4 Data selection** is where we conduct an explanation of how we deal with missing values and how the data sets are reduced. We conclude the **Data** section with **2.5 Annual distribution of pollen** in which we use density ridge plots to glance at what sort of results to expect.

3. Methods 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. The sections (**3.1 Nonparametric quantile regression**) and (**3.2 Linear regression on empirical quantiles**) explain the theoretical aspects of respective method, while simultaneously providing descriptions of how we have applied them in our models.

The most important findings of our models are presented in **4. Results**. This part of the paper is divided in to 3 sections for increased readability. The first, **4.1 Statistical model performance across quantiles** contains an analysis of how well each of our methods perform across quantiles of the data. it is in this section we determine which method to primarily base our conclusions from in the following sections. There are a lot of interesting results to share from this statistical analysis, the most important of which are presented as a table in **4.2 Estimations and predictions per species**. Here we analyze which species of pollen see the most significant seasonal shift and make estimations of which dates to expect the pollen season to be active at certain years. One of our assumptions prior to performing our analysis is that high latitudes correspond to larger shifts in the climate and would thus affect the pollen season to a greater extent than at lower latitudes. In the section **4.3 Geographic influence** we attempt to find out whether this assumption is correct or not in regards to predicting seasonal trends in pollen release in Sweden. Results that are not of central importance to our discussion and conclusions are nonetheless accessible in **Appendix C**.

Section **6. Discussion**, is divided in to (**6.1 Comparison with prior studies**) and (**6.2 Further improvements**). The former section covers a comparative analysis of our results with other, similar studies. For each of our questions at issue, we may be able obtain a definitive conclusion, in which case we present this in (**5. Conclusions**). If we are not able to obtain a definitive conclusion, a description of what we is 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 (**6.2 Further improvements**) on which parts of our work is inconclusive and how they can be improved upon.

This paper concludes with the 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 refered to by, and refers back to a result or explanation.

1. Literature review

In a paper from the journal "Proceedings of the National Academy of Sciences" (Hansen et al. 2006), James Hansen and his team were able to deduce that global surface temperature has increased $\sim 0.2°C$ per decade between the mid 1970's and mid 2000's. Assuming this trend also holds for the past 2 decades, we estimate the global average temperature to be almost $1°C$ warmer than half a century ago. Using these trends in global average temperature over time, we are able to form conclusions about how plant phenology, and more specifically pollen release, is affected by a warming climate.

A Dutch study analyzing a 31 year long collection of pollen in the Netherlands saw an advance in the pollen season by 3-22 days, depending on species. With species such as elm, oak and alder advancing between 15 and 18 days, while they recorded lesser advances in willow, birch and grass pollen at 12, 10 and 6 days respectively (Van Vliet et al. 2002).

In a more local study conducted by the Palynological laboratory in Stockholm (Lind et al. 2016), there was found to have been an advance in the seasonal starting dates for many arboreal species of pollen such as birch, oak, pine etc. Peak dates of pollen release were found to have largely followed the same trajectories as the starting dates, while the end dates for herbaceous pollen species like grass, were seen to have in fact moved to later dates. The length of pollen seasons had increased for some species, grass, mugwort and birch being among them, while most arboreal pollen types saw a decrease in overall seasonal length. A study of plant phenology in high latitude environments found that their advancement in annual phenology tends to be greater than equivalent plants at lower latitudes (Alecrim, Sargent, and Forrest 2023). This is consistent with findings that show a greater increase in average yearly temperature near the poles than closer to the equator (Hisano et al. 2021). Research has found that the ongoing increase in temperature extremes may be contributing to an extended seasonal duration of airborne allergenic pollen across the northern hemisphere (Ziska et al. 2019).

For those whom suffer from pollen allergies, the length and intensity of pollen release would preferably be as low as possible. Not all types of pollen are as allergenic as each other however. People may react differently to various species of pollen. (García-Mozo 2017) concludes that grass pollen are the most allergenic worldwide, whereas in an Italian study, it was revealed that birch pollen was the leading cause of allergic pollen reactions in northern europe (D'Amato et al. 2017), due to the fact that this type of pollen is more prevalent in this region. What we can conclude from these findings is that seasonal changes in grass and birch pollen affect the population to a greater extent than other types and thus may be of greater interest to follow.

Our primary source of material covering the theoretical background of our methods can be found in a study of migratory birds (Karlsson and Hössjer 2022). They came to the conclusion that quantile regression (QR) is more effective than linear regression on quantiles (EQ) at predicting seasonal shifts in migratory birds as a response to climate change. We compare our results in the **Discussion** section. Another piece of work we have used to build our models is (Takeuchi et al. 2006) in which we find more detailed definitions and properties of quantile regression. Many of the most central aspects of quantile regression, like the pinball loss function among other things, are thoroughly explained in (Koenker 2005).

2. Data

2.1 Data collection

As 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 (HIRST 1952) 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 (Natural History 2022) (Natural History 2017). However considering the structure of the data set and the consequent data analysis being relativistic, for which a descriptive presentation follows in the **Data** section, this phenomenon has been ignored.

2.2 Understanding the data

The data set 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 (Alecrim, Sargent, and Forrest 2023). This variable is however entirely dependent on `station`. A short description of the meaning of each variable is shown in table 1. Equipment for data collection is active only during the period of the predicted pollen season, estimated through using predictive models for pollen activity based on historic results.

Table 1: Description of the variables present in the data sets used in this research paper.

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 coordinates of said station.

Viewing the years of availability in table 2 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. For the English and Swedish names of each pollen genus, see **Appendix A**.

Table 2: Years for which data for different pollen species are available at the pollen monitoring stations used in our research, as well as their latidudal coordinates.

Station	Latitude	Pollen genus (since the year)
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)

2.3 Data transformation

In order to be able to perform methods like quantile regression and linear regression on quantiles, a transformation of the given data structure is necessary. To be more precise, we must turn each individual pollen, quantified by the `count` variable, in to its own data point. Using the function `uncount()` in `R` we are able to perform the described transformation of data. The daily counts of certain pollen types at any given station can be in the thousands. Each pollen season usually lasting around a month, and the data covering multiple species of pollen at a multitude of geographic locations over a number of years, means in return, that the size of the data set increases dramatically and as a consequence slows down computation time.

2.4 Data selection

After altering the structure of the data in to an individualistic format, we first remove any observations with missing values in any of the columns `date`, `station` or `species`, since these are fundamental parameters to perform the following data analysis. For the purposes of EQ regression, the time it takes to compute the models on the entire data set 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 data set (one data set 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 data sets 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 effect this has on our results the rows must be removed uniformly over the time of year. If the number of observations in a data set 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 mostly the same and thus the results of the quantile regression. Information about by how much each data set was reduced by can be found in **Appendix B**.

2.5 Annual 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 1 we present a ridge plot containing information about the distribution of observed pollen by date of year. The `geom_density_ridges()` function from the `R` library `ggridges` uses kernel density estimation (KDE) to fit a continuous density function to what is, in essence, a histogram of the frequency of pollen on each date. KDE is a non-parametric method used to estimate the probability density function of a random variable. This is achieved by placing a kernel at each data point. In our case the kernels are Gaussian distributed. We then summarise each of the kernels' distributions in order to obtain a smooth continuous approximation of the density. The width of the kernel, known as the bandwidth, determines by how much each data point contributes to the density estimate. Larger bandwidths result in smoother but less detailed density estimates, while smaller bandwidths result in more detailed but noisier density estimates (Duong 2001). A common way to find the optimal bandwidth is to use the Asymptotic Mean Integrated Squared Error (AMISE). The function `geom_density_ridges()` uses an alternative method based on the (Sheather and Jones 1991) plug-in bandwidth selection method.

By looking at the average distribution from the first and last 5 years in the data set 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 (Poaceae) 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 annual pollen release may suggest that the average pollen season now begins earlier by up to a month compared to 45 years ago.

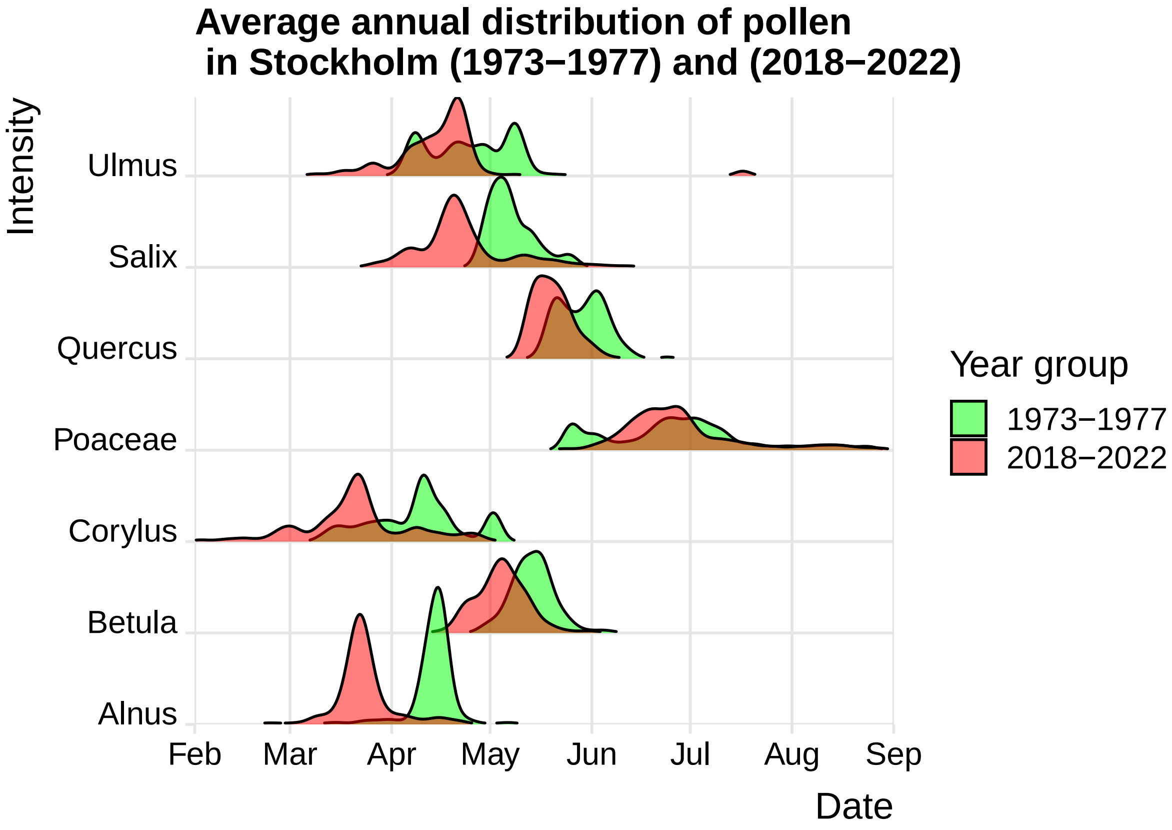

Figure 1: Ridgeline figures approximating the density functions of the distribution of annual pollen. Comparing the observed dates of pollen capture for the average of the years 1973-1978 (green) with the average for 2018-2022 (red).

3. Methods

In this section we explain the underlying theory behind the statistical approaches used for our data analysis. We will compare “linear regression on empirical quantiles” (EQ), as used in (Karlsson and Hössjer 2022) with the supposedly more powerful method “nonparametric quantile regression” which we refer to as QR.

In the QR approach, we estimate the response variable as the conditional median of the predictor variables, of which the median can be substituted to any other quantile of data. In the EQ method however, we use the method of ordinary least squares (OLS) to estimate the conditional mean of a subset of the predictor variables corresponding to the desired quantile level. Ordinary linear 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 they are less sensitive to 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 (Bastl, Kmenta, and Berger 2018). 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 annual pollen have been counted. By the same definition, the pollen season is said to end when 95% of all pollen has been released. A reason as to why the starting quantile level and the one at the end of the season are not symmetric (i.e. why $Q_{start} \neq 1 - Q_{end}$) may be similar to how infection transmission models tend to behave, namely that the beginning of the season tends to be very intense while the distribution decreases more slowly towards the end of the season.

When performing calculations of the arithmetic mean of a vector, we use the following formula:

$$\bar{x} = \arg\min_{\mu \in \mathbb{R}} \sum_{i=1}^{n} (y_i - \mu)^2, \tag{1}$$

where y corresponds to the vector of observed values and μ is a scalar value that minimizes the sum of squares.

If we instead want to calculate the arithmetic median, we use the absolute values between the vector of observations and the average instead of the square of these values. The median is thus provided by

$$\hat{x} = \arg\min_{\mu \in \mathbb{R}} \sum_{i=1}^{n} |y_i - \hat{x}|. \tag{2}$$

Furthermore, if one wants to compute any given quantile τ given a vector of observations, one may use the minimization algorithm

$$x_\tau = \arg\min_{\mu_\tau \in \mathbb{R}} \sum_{i=1}^{n} \mathcal{L}_\tau(y_i - \mu_\tau), \tag{3}$$

where $\mathcal{L}_\tau(\xi)$ represents the pinball loss function (Koenker 2005). This is a convex, piecewise linear function which computes the deviation between the predicted quantile and the actual value of the target variable. It is defined by

$$\mathcal{L}_\tau(\xi) = \begin{cases} \xi \cdot \tau & \text{if } (\xi) \geq 0 \\ \xi \cdot (\tau - 1) & \text{if } (\xi) < 0. \end{cases} \tag{4}$$

The name of the pinball loss function derives from the fact that it's shape somewhat resembles that of a pinball game as shown in figure 2. Rather trivially, the loss is minimized at $\xi = 0$, since that means the predicted quantile exactly attains the value of the observed value of the target variable. Overprediction is more heavily penalized than underprediciton for quantiles $\tau < 0$, while the opposite is true when $\tau > 0$. This is because the loss function places more weight on the residuals above the predicted value for overprediction and below the predicted value for underprediction (Takeuchi et al. 2006).

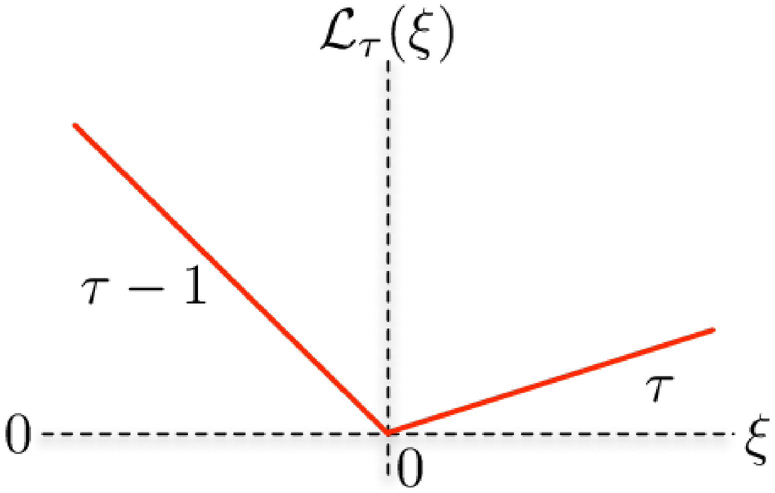

Figure 2: Illustration of the pinball loss function.

3.1 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 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

$$F_{Y_i|x_i}(y) = \mathbb{P}(Y_i \leq y \,|\, x_i), \quad -\infty < y < \infty \tag{5}$$

In order to find the τ-quantile of the conditional distribution of $Y_i \,|\, x_i$ 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

$$Q(\tau \,|\, x_i) = \inf\{y; F_{Y_i\,|\,x_i}(y) \geq \tau\} \tag{6}$$

of the conditional distribution function in (5) represents the conditional quantile function.

Moving forward, our model takes the form

$$Q(\tau \,|\, X) = X\beta(\tau) + \varepsilon(\tau) \tag{7}$$

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 (Karlsson and Hössjer 2022).

An optimization problem is then constructed as to minimize the objective function in (5). 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 however, we want to learn the parameters of a quantile regression model that can accurately predict the conditional quantiles of the target variable. To do this, we minimize the pinball loss function $\mathcal{L}_\tau(\xi)$, which we previously introduced in **3. Methods**, over a set of data X, y (Koenker 2005; Takeuchi et al. 2006).

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

$$\beta(\tau) = \underset{b \in \mathbb{R}^{\nvDash}}{\arg\min} \sum_{i=1}^{n} \mathcal{L}_\tau(y_i - x_i b). \tag{8}$$

Comparing equation 8 with equation 3, note that we replace the scalar μ_τ with $x_i b$ in the loss function and minimize over b leading to an estimation of the β-coefficient which coupled with a vector of data x grants a predictive linear model (Karlsson and Hössjer 2022).

Since the function (8) is non-differentiable at 0, no direct solution exists. Rather we use numerical estimation methods provided in the `R` library `quantreg`, such as the Frisch-Newton interior point method to find the optimal point along the ξ-axis (Koenker 2023).

3.2 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 year-wise empirical quantiles 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. Most source material and notations in this section are collected from (Karlsson and Hössjer 2022). 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.

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 year $t \in \mathcal{T}$ we extract a set of observations $\mathcal{Y}_t$. Let $\tau \in (0, 1)$ be a quantile. For each set $\mathcal{Y}_\sqcup$ we let $\hat{F}_{(t)}$ be the empirical distribution function formed by the elements of its set. The corresponding empirical quantile is defined as

$$\hat{Q}_{(t)}(\tau) = \inf\left\{ y \in \mathcal{Y}_t : \hat{F}_t(y) \geq \tau \mid \mathcal{Y}_t \neq \emptyset \right\}. \tag{9}$$

As in the previous method we construct the $(n \times 2)$ matrix X by stacking the intercept and vector of years $(1, t)$ on top of each other. For all $t \in \mathcal{T}$, we stack the quantiles into the vector of observations $Y(\tau)$. Next we formulate the linear model

$$Y(\tau) = X\beta(\tau) + \varepsilon(\tau). \tag{10}$$

As before $\beta = (\beta_0, \beta_t)$ defines the regression parameters for the intercept β_0 and year β_t. 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 of the model is given by

$$l(\beta(\tau), \sigma^2(\tau) \mid Y(\tau), X) = \sum_{t \in \mathcal{T}} \log f\left(\hat{Q}_t(\tau) \mid (X\beta(\tau))_t, \sigma^2(\tau) \right). \tag{11}$$

So far in this approach the model grants each year t the same weight, not each individual pollen. Since the amount of pollen observed each year does not remain constant. A reweighting of the log-likelihood may be conducted as to grant each pollen the same weight. To get even weighting of each pollen, we add a weight factor $w_t = |\mathcal{Y}_t|$ to each empirical quantile $\hat{Q}_t(\tau)$. The reweighed log-likelihood, takes the form

$$l_w(\beta(\tau), \sigma^2(\tau) \mid Y(\tau), X) = \sum_{t \in \mathcal{T}} w_t \log f\left(\hat{Q}_t(\tau) \mid (X\beta(\tau))_t, \sigma^2(\tau)\right), \tag{12}$$

This can easily be implemented with the `weights` argument of the `lm` function where we set the weight to be $\frac{1}{N_t}$ where N_t represents the number of observations at year t.

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

$$MLE(\beta(\tau), \sigma^2(\tau)) = \underset{\beta, \sigma^2 \in \mathbb{R}^{\nvDash}}{\arg\min} \left\{ l_w(\beta(\tau), \sigma^2(\tau) \mid Y(\tau), X) \right\} \tag{13}$$

for a given quantile τ. Since the function is continuous and twice differentiable, the solution can easily be found using conventional methods (Karlsson and Hössjer 2022).

4. Results

4.1 Statistical model performance across quantiles

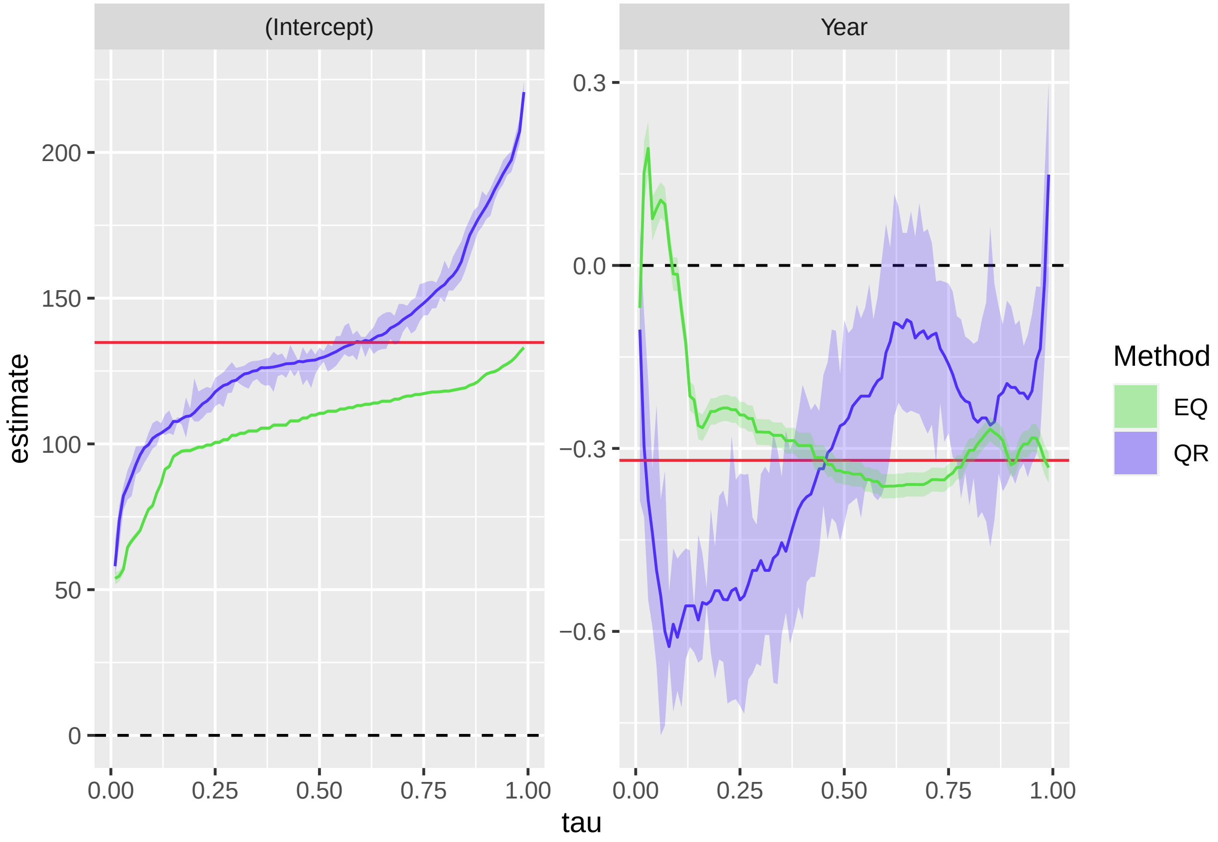

Figure 3: Line graph over the coefficient and intecept estimates for each quantile in Stockholm, all species. The thicker, opaque bounaries define the 95%-confidence intervals for each method. The red line shows the linear regression coefficient without quantile parameterization.

Putting both these methods to the test, we take a look at how each method behaves for any quantile of data. In figure 3 we limit the scope of our analysis to pollen monitored in Stockholm, similar patterns are however present for data found at other stations. We observe that the intercept begins at around day 50 for both models but increases more rapidly over the quantiles for QR than with EQ, leading to a significantly later estimation of when the pollen season ends (high quantiles) using the QR approach. Another phenomenon visible in figure 3 is that the direction of the slopes are different in the second graph. What we observe is that the QR estimates of the `year` variable move from strong negative coefficients, to lesser negative or even slightly positive coefficients across the quantiles. This means we get a significantly earlier start of the pollen season, but a less shifted end to the season. All in all this results in a longer and more spread out pollen season. In stark contrast to the QR method, when analyzing the EQ coefficient estimations, we find that the early parts of the pollen season actually moved to later dates, while most of the season beyond the first 20% of pollen or so, saw a shift of around 0.3 days per year to earlier dates.

Based on this information, one may ask oneself which of these approaches we deem to be the most fit to represent the behaviour of the pollen season. The fact that the size of the confidence intervals differ so greatly is, to us, sufficient material to determine the EQ model as more desirable for the purpose of this research paper. In order to understand this difference in confidence between the to methods, let us take a look at how the models look overlayed on the set of observations.

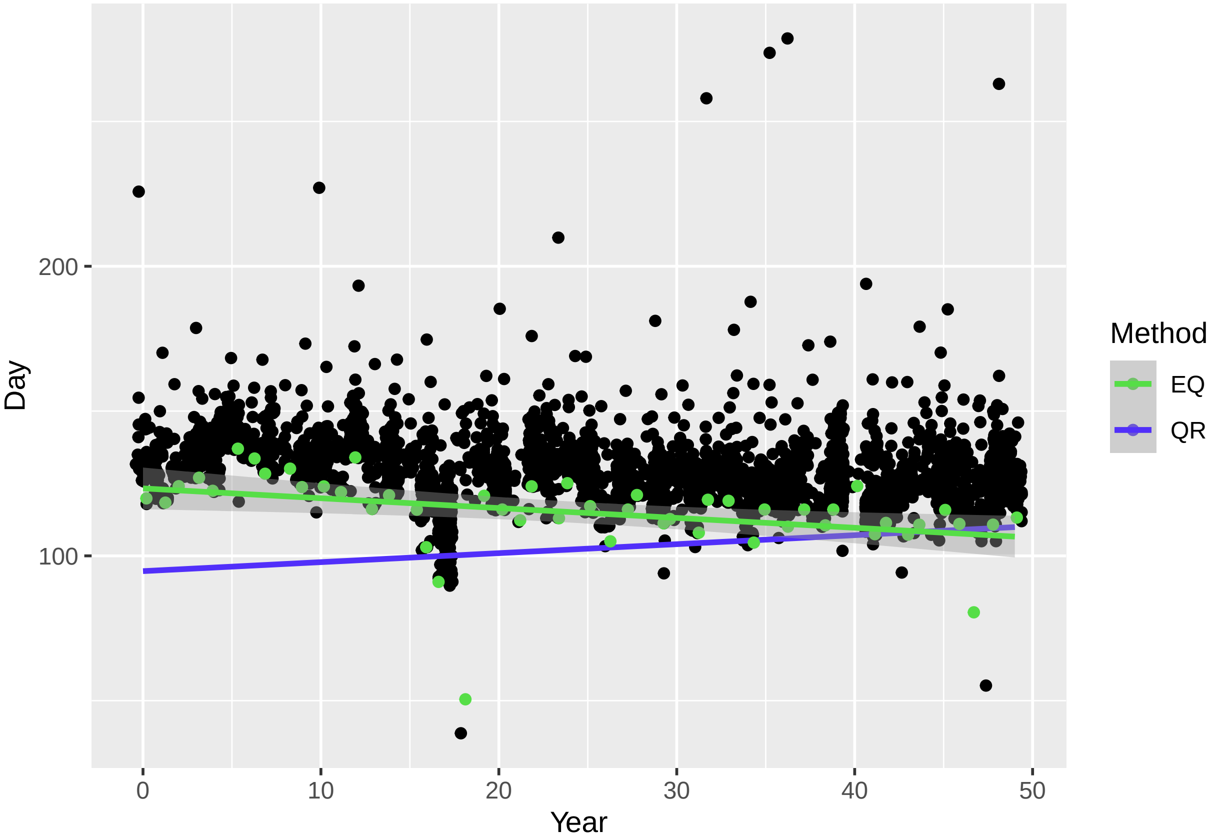

Figure 4: Jitter plot over the annual distribution of betula pollen in Stockholm comparing QR and EQ at the 1% quantile level.

Figure 4 tells us the difference in fitting of both our models. The quantile regression method fits a linear model to encompass a given quantile, in this case the first 1% of annual pollen release, below the regression line. We can observe that year 17 appears to be a heavy outlier in which the entire pollen season came significantly earlier than other years. Since the QR method tries to fit a linear trend with a specific amount of observations below it, and a large amount of these first 1% of observations are from the 17th year, the model is heavily skewed by the observations from this year and thus does not generate a good fit. EQ does not succumb to this issue since we in this approach aggregate the observations from each year to a specific date value (red points) and then perform ordinary linear regression on these aggregated observations. In this case the one outlier year corresponds to only 2% of the total observations while for QR this effect appears to be well over 50%. These results explain why the confidence intervals for QR are much larger than the equivalent measure of EQ. Thus we conclude that linear regression on empirical quantiles is a more fitting approach to use in the case of estimating the annual seasonal shift of pollen.

4.2 Estimations and predictions per species

Table 3: Modelled pollen season shift by species averaged over all monitored stations (p-values and R^2 are averages).

Species	Avg slope (1%)	P-value (1%)	Adj. R^2 (1%)	Avg slope (50%)	P-value (50%)	Adj. R^2 (50%)	Avg slope (95%)	P-value (95%)	Adj. R^2 (95%)	Slope (season length)
Corylus	-0.755	0.03	0.23	-0.539	0.09	0.13	-0.408	0.16	0.11	0.42
Alnus	-0.490	0.09	0.17	-0.418	0.09	0.17	-0.384	0.09	0.19	0.30
Salix	-0.432	0.07	0.19	-0.396	0.07	0.21	-0.178	0.09	0.10	0.30
Ulmus	-0.336	0.25	0.06	-0.184	0.15	0.09	-0.303	0.22	0.09	0.16
Quercus	-0.326	0.09	0.18	-0.217	0.19	0.11	-0.163	0.27	0.06	0.18
Betula	-0.188	0.33	0.12	-0.135	0.39	0.09	-0.144	0.35	0.07	0.00
Poaceae	-0.092	0.40	0.02	-0.201	0.13	0.11	0.132	0.15	0.09	0.28

Table 4: Estimated starting dates and seasonal length averaged over all monitored stations (p-values and R^2 are averages).

Species	Estimated start 1973	Predicted start 2023	Predicted start 2050	Season length 1973	Season length 2023	Season length 2050
Corylus	April 7	February 23	January 31	19	40	52
Alnus	April 4	March 1	February 11	21	36	44
Salix	April 20	April 1	March 22	35	50	58
Ulmus	April 25	April 1	March 19	17	25	30
Quercus	May 29	May 7	April 25	11	20	26
Betula	April 30	April 20	April 14	25	25	26
Poaceae	June 1	May 28	May 26	58	72	80

Table 5: Modelled pollen season shift by species in Stockholm.

Species	Slope (1%)	P-value (1%)	Adj. R^2 (1%)	Slope (50%)	P-value (50%)	Adj. R^2 (50%)	Slope (95%)	P-value (95%)	Adj. R^2 (95%)	Slope (season length)
Alnus	-0.777	0.00	0.38	-0.501	0.00	0.32	-0.351	0.00	0.16	0.70
Corylus	-0.648	0.00	0.18	-0.347	0.05	0.06	-0.332	0.01	0.11	0.36
Salix	-0.473	0.00	0.24	-0.432	0.00	0.27	-0.231	0.00	0.17	0.24
Ulmus	-0.420	0.02	0.09	-0.373	0.00	0.15	-0.335	0.01	0.11	0.18
Betula	-0.349	0.00	0.21	-0.298	0.00	0.28	-0.204	0.00	0.17	0.14
Quercus	-0.322	0.00	0.35	-0.266	0.00	0.27	-0.257	0.00	0.34	0.12
Poaceae	-0.042	0.50	-0.01	-0.128	0.06	0.05	0.205	0.00	0.14	0.30

Table 6: Estimated starting dates and seasonal length averaged in Stockholm (models with p-values larger than 0.05 are ignored).

Species	Estimated start 1973	Predicted start 2023	Predicted start 2050	Season length 1973	Season length 2023	Season length 2050
Alnus	March 31	February 26	February 8	16	51	70
Corylus	March 27	February 17	January 28	25	43	53
Salix	April 24	April 1	March 19	36	48	54
Ulmus	April 23	April 2	March 22	14	23	28
Betula	May 3	April 16	April 7	23	30	33
Quercus	May 26	May 10	May 2	11	17	20
Poaceae	June 2	May 29	May 27	66	81	88

Table 3 grants us information about how our EQ models predict the pollen seasons, for all monitored species, to respond to increasing temperatures in the atmosphere. The entries in the columns named `Slope` are the average of the β-coefficient values generated at different combinations of `species` and `station`, for each species of pollen. Although some pollen monitoring stations generated results with positive coefficients for certain species, such a phenomenon appears not to be present when we take the mean of our regression coefficients for each species from all stations. The exception to this however would be the 95%-quantile estimate for grass pollen (poaceae), meaning the grass pollen season is moving to later dates over time, while its start and peak shifts to earlier dates. This brings us to the conclusion that the pollen season unanimously is headed towards an earlier start, as well as peak, as the climate warms. The date at which 95% of all annual pollen gets monitored also seems to arrive earlier each year. The exception to this pattern is, as we point out, grass pollen. The slope estimates for the end of season arrival dates are however lower than the starting dates across all species. By using the linear regression coefficients shown in the table, we can predict the dates at which quantiles of the annual pollen release are expected to appear at.

Using the 1% quantile estimate, we observe in table 4 that the predicted dates for hazel and alder pollen moved from early April back in 1973 to late February as of 2023 and are expected to reach late January and early February respectively by 2050. The willow and elm seasons appear to shift at a slightly slower pace moving from late to early April over the same 50-year time span. Meanwhile birch pollen only accumulated a movement of 10 days over the past half century. Oak pollen now starts to appear in early May rather than towards the end of the month, which appears to have been the case in the 1970's. Just as we saw in the ridge plots in section **2.5 Annual distribution of pollen**, the starting dates of the grass pollen season are not seen to be as greatly affected by climate change as for arboreal pollen, although we find the peak to advance by 0.20 days annually, which is faster than at least one arboreal plant (birch).

Something we can observe across the board in regards to the monitored pollen species however, is that the season lengths (difference in dates between the 95%-quantile and 1%-quantile) become longer over time. This phenomenon is present in all species we have covered in this paper. The largest increase in pollen season length among the arboreal species is seen in hazel pollen (21 days in 50 years and 33 days in 77 years) while grass pollen, which saw negligible changes to its starting dates, still saw a significant increase in season length (14 days in 50 years and 22 days in 77 years), largely owing to the fact that the seasonal end moves to later dates by approximately 0.13 days a year. Birch pollen saw the smallest increase as the season length is extends by barely a day after 77 years of increasing temperatures.

Looking at the equivalent results from the Stockholm region in tables 5 and 6, we see similar slopes and predicted starting dates, although season lengths appear somewhat longer than average. In this region we find alder to be the species most heavily affected by climate change and a smaller advance in the start and peak of annual grass pollen distributions. While the slopes tell us a great deal about alterations in the dates of pollen release, it is important to consider that the average P-values for models of some species are above 5%, meaning they are not statistically significant. Looking at the average values for the adjusted R^2 of these models, we see that none are negative, meaning our models are a better fit that a horizontal line.

There could be many explanations for why the annual distribution of pollen behaves the way we have observed.

In this paper we are not aiming to try and answer this question. Some quick thoughts however are that with an ever warmer climate, the temperatures at which plants are able to pollinate are present more days of the year, meaning the window of opportunity for trees and grass to release their pollen gets longer over time. The early pollen is what, for most species, gets shifted most heavily to earlier dates. An idea to why this may be the case is that since the spring arrives earlier over time, plants are able to begin the process of pollination earlier and consequently release most of their pollen (which always occurs in the spring months for arboreal species) at earlier dates, leaving the later parts of the year with less relative amounts of pollen. Results for each individual station can be accessed in **Appendix C**.

4.3 Geographic influence

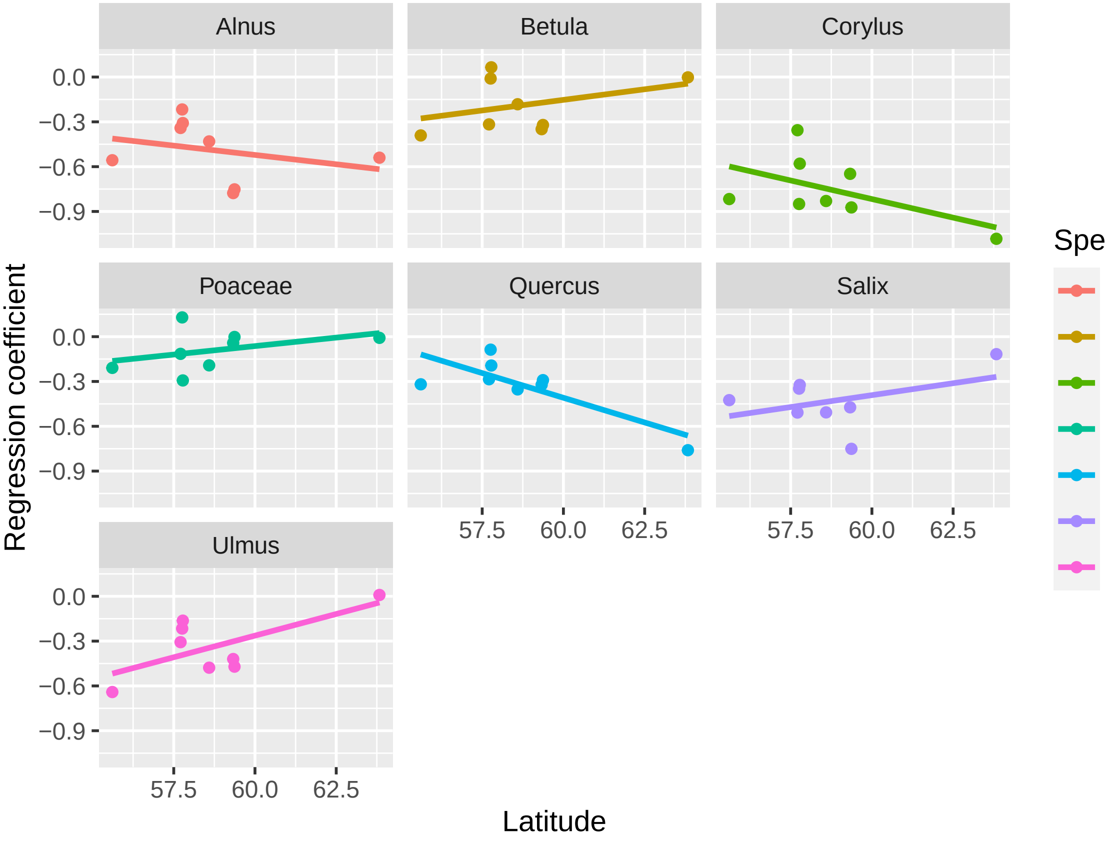

Figure 5: Visualization of latitudal impact on the EQ estimations (1% quantile)

In figures 5 through 7 we can view the regression coefficients for each species and how they differ based on the latitude of the station the pollen were collected at. Thus we can determine whether pollen in colder climates are affected by climate change by a greater extent than in warmer areas, or vice versa. Figure 5 tells us how the beginning of the pollen season shifts at the observed locations throughout Sweden. Alder, hazel and oak all appear to have more dramatic seasonal shifts at northerly latitudes, while grass, birch, willow and elm show the opposite effect. The median quantile, which we also refer to as the peak of the pollen season was not observed to have had a significant impact of the latitude of where it took place, as seen in figure 6. The steep regression slope of the peak of the elm pollen seasonal shift comes down to the fact that the coefficient at the highest latitude is vastly different to the others. By that metric, a conclusion can not be made about whether this is proof of climate change impacting elm pollen at higher latitudes simply an outlier. It is worth noting that the p-value of the aforementioned elm coefficient in Umeå is 0.981. By all accounts, this points to the data point being an outlier. By figure 7 we see how the coefficients for the 95%-quantiles are estimated. We now observe, in contrast to the 1%-quantiles, that grass pollen attributes to a quicker movement toward earlier dates, the further north one goes. To summarise, the length of the grass pollen season appears to

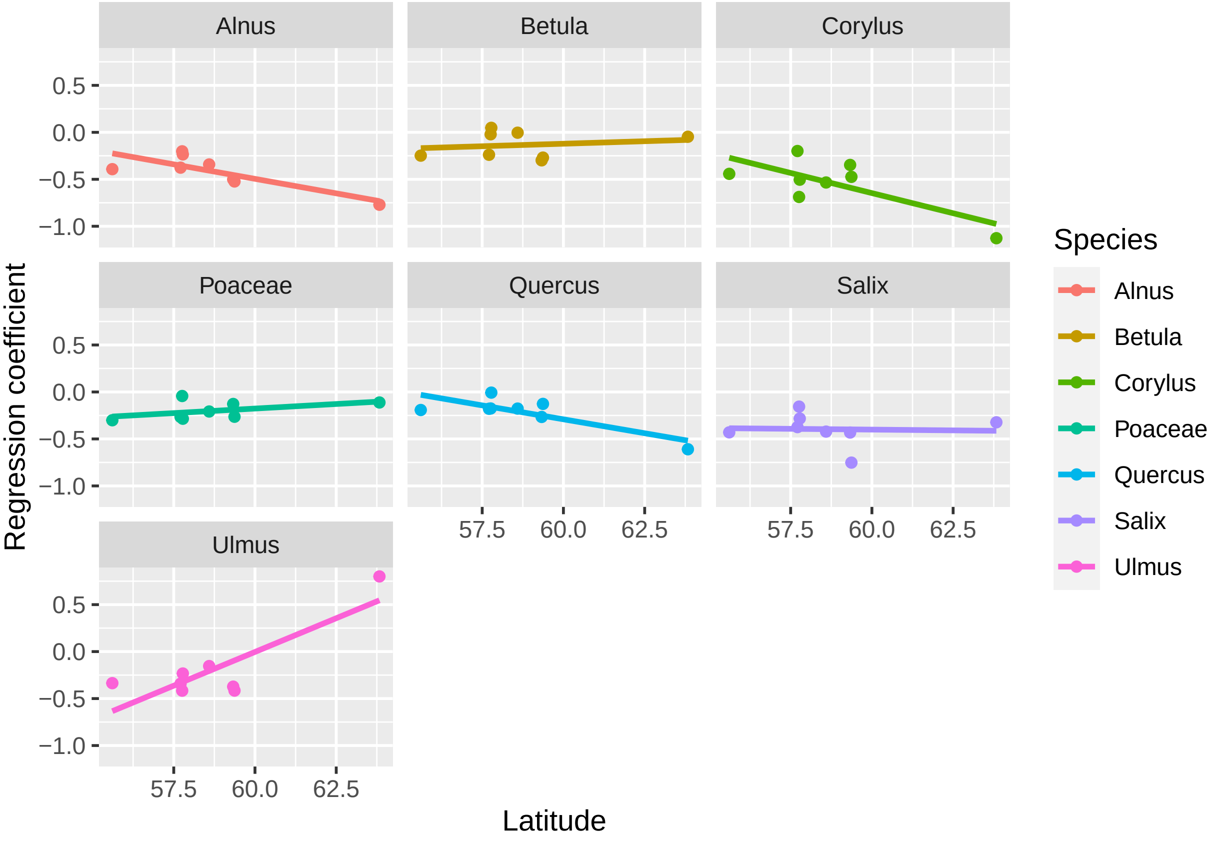

Figure 6: Visualization of latitudal impact on the EQ estimations (50% quantile)

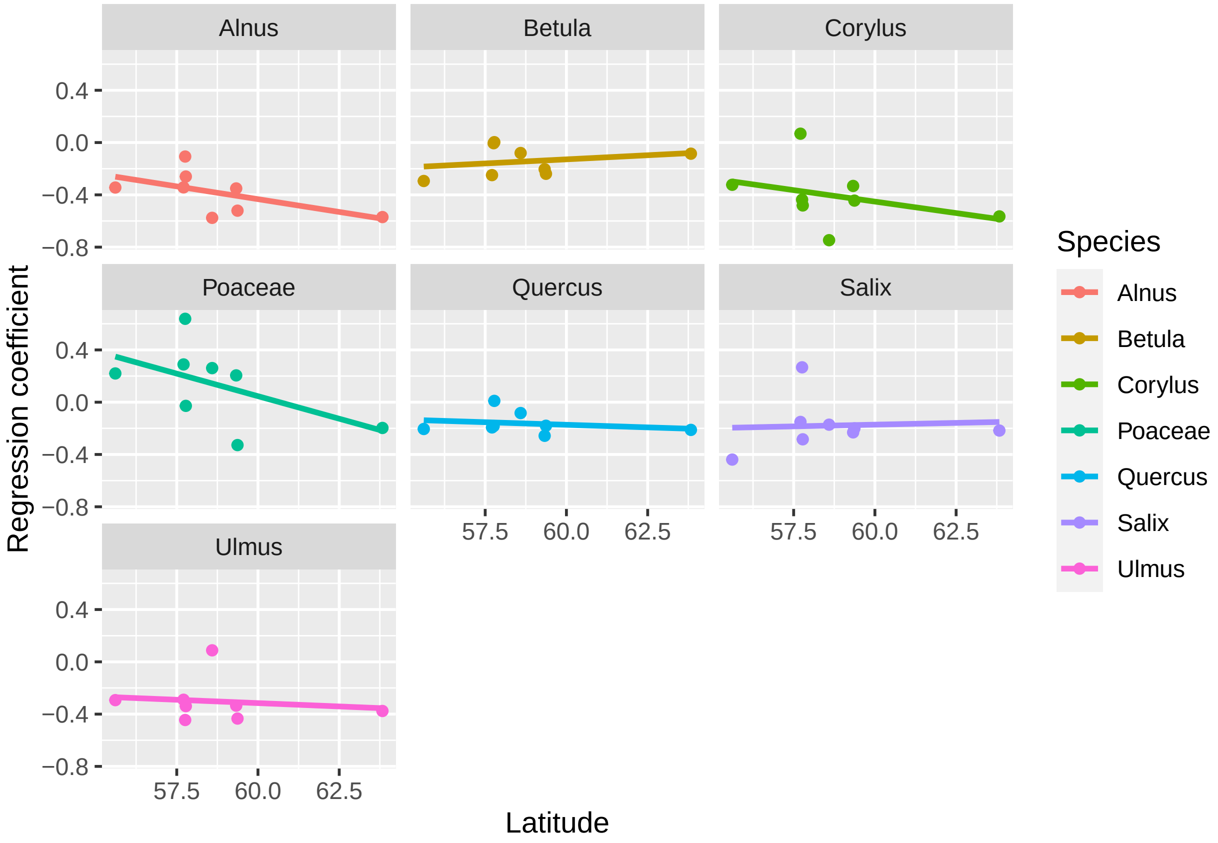

Figure 7: Visualization of latitudal impact on the EQ estimations (95% quantile)

be reduced at higher latitudes. The remaining pollen species' seasonal ends do not appear to be strongly affected by latitude. What we can conclude from this is that the seasonal window for grass pollen does not expand as much at high latitudes than what can be observed at lower latitudes. Alder pollen appears to, somewhat consistently, attain more strongly negative coefficients, for any quantile, at the higher latitudes. Other species were not observed to be affected by the latitude of their release by any greater extent.

5. Conclusions

The results in this study point towards EQ being a more fitting model than QR when predicting changes in seasonal pollen distributions. Largely, this has to do with the fact that the spread, or variance of dates is fairly low, leading to clustering of extreme observations in the case of abnormally hot or cold years. This can in turn be catastrophic when using nonparametric quantile regression to predict extreme quantiles, as discussed in **4.1 Statistical model performance across quantiles**.

Based on the slopes of our predictive linear models, we conclude that an advance in the seasonal start of pollen release among arboreal species occurs by between 0.19 and 0.76 calendar days per year. The seasonal peak in arboreal pollen appears to also be heavily affected by climate change, as an advance of 0.14 to 0.54 days per year is observed. The 95th percentile of annual observed pollen was found to not advance as dramatically, resulting in longer periods of seasonal activity. Since 1973, season lengths of arboreal pollen have increased by 8-21 days on average with the exception of birch, for which no noticeable change in season length was observed. Grass pollen saw an increase of 14 days over the same 50-year time frame. Grass pollen is the only species to attain a positive slope (pollination occurring later over time) as we see ending dates regress by 0.13 days annually.

A comparison of seasonal pollen distributions between various locations, as seen in **4.3 Geographic influence** reveals that the slopes can differ quite dramatically between latitudes. In particular we see that alder, hazel and oak pollen see more dramatic advances at higher latitudes, while herbaceous species appear less affected by climate change further north.

6. Discussion

6.1 Comparison with prior studies

The conclusion made in (Karlsson and Hössjer 2022) was that nonparametric quantile regression (QR) is the most well suited for single as well as multiple species analyses of migratory birds. What we can conclude from this study however is that linear regression on empirical quantiles (EQ) is the preferred method between the two when it comes to modelling the advance in seasonal distribution of pollen. As mentioned in **Statistical model performance across quantiles**, the confidence intervals are way smaller for the EQ model than QR. We observed how outlier years influence the slope of QR models more than EQ models. Perhaps this is where we can find an explanation as to why our findings differ in our respective studies. One reason could be that the variance in observed dates among pollen is lower than for migratory birds, leading to a more heavily skewed slope if a cluster of observations falls a lot earlier or later than normal, which tends to be the case for years with abnormal average temperatures.

As we recall the results observed in the **Estimations and predictions per species** section, our findings reveal that hazel pollen is subject to the largest advance in starting dates as well as overall dates of seasonal pollinosis. Grass pollen, which is not arboreal, we find to not advance its end of season date, rather the opposite effect is observed as there is a push back of 0.13 days per year on average. We find birch to be the species with the smallest change in season length and the shortest advance in starting dates among the arboreal pollen.

As mentioned in **1. Literature review**, a similar Stockholm based study on seasonal pollen shift in response to climate change (Lind et al. 2016) found that end-of-season dates for herbaceous species of pollen, like grass, had moved to later parts of the year but showed no significant change in starting or peak dates. This is consistent with our findings as we see a push back of 0.21 days per year on average in Stockholm. Just like in (Lind et al. 2016) we see a significant advance in starting, as well as peak dates for all arboreal species. Our averaged results agree on the fact that hazel was the species with the greatest advance in dates whereas if we look exclusively at the phenology in Stockholm, alder appears to have advanced more rapidly. We found oak to be the species with the smallest change in season length and the shortest advance in starting dates among the arboreal pollen in Stockholm, whereas the model provided in (Lind et al. 2016) found lower estimates for both elm and willow trees.

6.2 Further improvements

The QR approach struggles with clusters of outliers such as a year with a significantly earlier pollen season. One way to look at it is that there exists a random effect of year that we do not consider in this approach, thus deeming the assumption of independent observations incorrect in the case of our QR model. In order to improve upon these models, one may consider adding a random variable to deal with this yearly effect, resulting in a linear quantile mixed model, which may grant more significant slope estimations than these QR and EQ models.

This study only tells us about the distribution of pollen, we do not make any considerations about the volume of pollen, which may be a telling factor in determining why and how plants' pollination cycle reacts to a warming climate. Thus, further studies of this character may consider integrating pollen volume in their models.

A substantial amount of our models are not significant at 95% confidence level. The majority of these are for combinations of `species` and `station` that do not have extensive sets of data, leading to less accurate models. In particular, grass pollen and observations from the Umeå laboratory are overrepresented in poorly fitted models. To increase the accuracy of such models, larger data sets are necessary.

Bibliography and References

Alecrim, Evelyn F, Risa D Sargent, and Jessica RK Forrest. 2023. "Higher-Latitude Spring-Flowering Herbs Advance Their Phenology More Than Trees with Warming Temperatures." *Journal of Ecology* 111 (1): 156–69.

Bastl, Katharina, Maximilian Kmenta, and Uwe E Berger. 2018. "Defining Pollen Seasons: Background and Recommendations." *Current Allergy and Asthma Reports* 18: 1–10.

D'Amato, Gennaro, Carolina Vitale, Alessandro Sanduzzi, Antonio Molino, Alessandro Vatrella, and Maria D'Amato. 2017. "Allergenic Pollen and Pollen Allergy in Europe." *Allergy and Allergen Immunotherapy*, 287–306.

Duong, Tarn. 2001. "An Introduction to Kernel Density Estimation." 2001. https://www.mvstat.net/tduong/research/seminars/seminar-2001-05.pdf.

García-Mozo, H. 2017. "Poaceae Pollen as the Leading Aeroallergen Worldwide: A Review." *Allergy* 72 (12): 1849–58.

Hansen, James, Makiko Sato, Reto Ruedy, Ken Lo, David W Lea, and Martin Medina-Elizade. 2006. "Global Temperature Change." *Proceedings of the National Academy of Sciences* 103 (39): 14288–93.

HIRST, J. M. 1952. "AN AUTOMATIC VOLUMETRIC SPORE TRAP." *Annals of Applied Biology* 39 (2): 257–65. https://doi.org/https://doi.org/10.1111/j.1744-7348.1952.tb00904.x.

Hisano, Masumi, Masahiro Ryo, Xinli Chen, and Han YH Chen. 2021. "Rapid Functional Shifts Across High Latitude Forests over the Last 65 Years." *Global Change Biology* 27 (16): 3846–58.

Karlsson, Måns, and Ola Hössjer. 2022. "A Comparison Between Quantile Regression and Linear Regression on Empirical Quantiles for Phenological Analysis in Migratory Response to Climate Change." *arXiv Preprint arXiv:2202.02206*.

Koenker, Roger. 2005. *Quantile Regression*. Vol. 38. Cambridge university press.

———. 2023. "Package 'Quantreg'." 2023. https://cran.r-project.org/web/packages/quantreg/quantreg.pdf.

Lind, Tomas, Agneta Ekebom, Kerstin Alm Kübler, Pia Östensson, Tom Bellander, and Mare Lohmus. 2016. "Pollen Season Trends (1973-2013) in Stockholm Area, Sweden." *PloS One* 11 (11): e0166887.

Natural History, Swedish Museum of. 2017. "Så Gör Vi Prognoserna." 2017. https://pollenrapporten.se/ompollen/sagorviprognoserna.

———. 2022. "Hur Fungerar En Pollenfälla." 2022. https://www.youtube.com/watch?v=KgiFLbYWqDA&ab_channel=Naturhistoriskariksmuseet.

———. 2023. “Pollenlabb Och Mätstationer.” 2023. https://pollenrapporten.se/ompollen/pollenlabbochmatstationer.4.314e02dd13d69872ec0ab.html.

Sheather, S. J., and M. C. Jones. 1991. “A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation.” *Journal of the Royal Statistical Society. Series B (Methodological)* 53 (3): 683–90. http://www.jstor.org/stable/2345597.

Takeuchi, Ichiro, Quoc Le, Timothy Sears, Alexander Smola, et al. 2006. “Nonparametric Quantile Estimation.”

Van Vliet, Arnold JH, Aart Overeem, Rudolf S De Groot, Adrie FG Jacobs, and Frits TM Spieksma. 2002. “The Influence of Temperature and Climate Change on the Timing of Pollen Release in the Netherlands.” *International Journal of Climatology: A Journal of the Royal Meteorological Society* 22 (14): 1757–67.

Ziska, Lewis H, László Makra, Susan K Harry, Nicolas Bruffaerts, Marijke Hendrickx, Frances Coates, Annika Saarto, et al. 2019. “Temperature-Related Changes in Airborne Allergenic Pollen Abundance and Seasonality Across the Northern Hemisphere: A Retrospective Data Analysis.” *The Lancet Planetary Health* 3 (3): e124–31.

Appendix

Appendix A: Translation of the latin names of pollen species

Table 7: Pollen species translation in to English and Swedish.

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 B: Information about reduced data sets

In order to make the process of analyzing results a lot quicker, we have reduced the size of each data set to less than 5000 observations. Each data set is made up of observations of individual pollen for all combinations of species and monitoring station. Some data sets are already below this size limit, like all the hazel (corylus) data sets and a a few more at the Umeå station, while others need massive reductions in size. Table 8 denotes sizes of the data sets for combinations of species and station. The denominators of which we reduce the data sets with are represented in brackets. As mentioned in the section **Data selection**, a reduction in the size of each data set does not affect the shape of their distribution since we use the `slice()` function to remove observations uniformly across all dates.

Table 8: Data set sizes for combinations of species and location.
Brackets denote the factor of which each data set is reduced by.

Stations	Alnus	Betula	Corylus	Poaceae	Quercus	Salix	Ulmus
Eskilstuna	45 362 (10)	494 464 (99)	3 594 (1)	50 537 (18)	26 899 (6)	15 398 (4)	167 (1)
Göteborg	31 722 (7)	451 397 (91)	3 374 (1)	31 676 (7)	42 482 (9)	14 715 (3)	33 284 (7)
Jönköping	16 454 (4)	231 147 (47)	3 812 (1)	74 372 (15)	26 482 (6)	17 940 (4)	6 509 (2)
Malmö	55 346 (12)	212 232 (43)	4 254 (1)	48 265 (10)	54 957 (11)	13 929 (3)	35 560 (8)
Norrköping	24 089 (5)	372 668 (75)	113 (1)	37 613 (8)	57 914 (12)	19 285 (4)	4 884 (1)
Stockholm	31 596 (7)	286 630 (58)	5 848 (1)	22 639 (5)	40 942 (2)	31 227 (7)	11 645 (3)
Umeå	42 217 (9)	203 475 (41)	3 812 (1)	52 569 (11)	182 (1)	8 127 (2)	9 779 (2)

Stations	Alnus	Betula	Corylus	Poaceae	Quercus	Salix	Ulmus
Västervik	33 653 (7)	239 006 (48)	4 484 (1)	48 265 (11)	81 161 (17)	8 875 (2)	9 821 (2)

Appendix C: Results per species and location

Table 9: Start of pollen season for different species and locations

			P-	Adj.	Estimated start	Predicted start	Predicted start
Species	Location	Slope	value	R^2	1973	2023	2050
Corylus	Umeå	-1.083	0.00	0.35	May 22	March 15	February 6
Corylus	Eskilstuna	-0.873	0.00	0.42	April 12	February 21	January 25
Corylus	Västervik	-0.850	0.00	0.24	April 7	February 18	January 23
Corylus	Norrköping	-0.830	0.00	0.24	April 18	February 18	January 17
Corylus	Malmö	-0.817	0.00	0.28	March 27	February 14	January 24
Alnus	Stockholm	-0.777	0.00	0.38	March 31	February 26	February 9
Quercus	Umeå	-0.760	0.02	0.41	July 13	May 12	April 8
Alnus	Eskilstuna	-0.752	0.00	0.51	April 12	March 2	February 7
Salix	Eskilstuna	-0.751	0.00	0.38	April 24	April 2	March 21
Corylus	Stockholm	-0.648	0.00	0.18	March 27	February 18	January 29
Ulmus	Malmö	-0.641	0.05	0.08	April 12	March 19	March 6
Corylus	Jönköping	-0.580	0.03	0.11	March 18	February 21	February 8
Alnus	Malmö	-0.557	0.00	0.17	March 26	February 18	January 29
Alnus	Umeå	-0.540	0.01	0.13	May 2	March 13	February 14
Salix	Göteborg	-0.508	0.00	0.30	April 19	March 25	March 12
Salix	Norrköping	-0.507	0.00	0.31	April 19	March 30	March 19
Ulmus	Norrköping	-0.478	0.00	0.19	April 24	March 27	March 12
Salix	Stockholm	-0.473	0.00	0.24	April 24	April 1	March 20
Ulmus	Eskilstuna	-0.471	0.00	0.17	April 25	April 1	March 19
Alnus	Norrköping	-0.431	0.04	0.09	April 5	March 1	February 10
Salix	Malmö	-0.425	0.01	0.13	April 8	March 23	March 14
Ulmus	Stockholm	-0.420	0.02	0.09	April 23	April 3	March 23
Betula	Malmö	-0.391	0.00	0.30	April 28	April 10	March 31
Corylus	Göteborg	-0.356	0.22	0.01	March 20	March 6	February 26
Quercus	Norrköping	-0.353	0.01	0.17	May 26	May 6	April 25
Betula	Stockholm	-0.349	0.00	0.21	May 3	April 16	April 7
Salix	Västervik	-0.348	0.04	0.09	April 9	March 31	March 25
Alnus	Göteborg	-0.341	0.10	0.04	March 27	March 12	March 4
Salix	Jönköping	-0.323	0.09	0.06	April 15	March 31	March 23
Quercus	Stockholm	-0.322	0.00	0.35	May 26	May 11	May 2
Betula	Eskilstuna	-0.321	0.00	0.25	May 5	April 19	April 10
Quercus	Malmö	-0.319	0.00	0.21	May 18	April 28	April 17
Betula	Göteborg	-0.317	0.00	0.26	April 30	April 17	April 10
Alnus	Jönköping	-0.308	0.17	0.03	March 27	February 21	February 2
Ulmus	Göteborg	-0.307	0.10	0.04	April 18	March 25	March 12
Poaceae	Jönköping	-0.293	0.02	0.12	June 7	May 22	May 14
Quercus	Eskilstuna	-0.291	0.00	0.17	May 25	May 8	April 29
Quercus	Göteborg	-0.285	0.01	0.13	May 26	May 5	April 25
Alnus	Västervik	-0.217	0.38	-0.01	March 28	March 1	February 14
Ulmus	Västervik	-0.216	0.29	0.00	April 9	March 31	March 26
Poaceae	Malmö	-0.209	0.06	0.07	May 29	May 16	May 10
Quercus	Jönköping	-0.193	0.15	0.04	May 20	May 9	May 3
Poaceae	Norrköping	-0.192	0.12	0.04	June 2	May 29	May 26

Species	Location	Slope	P- value	Adj. R^2	Estimated start 1973	Predicted start 2023	Predicted start 2050
Betula	Norrköping	-0.182	0.10	0.05	April 29	April 18	April 13
Ulmus	Jönköping	-0.163	0.56	-0.02	April 15	March 29	March 20
Salix	Umeå	-0.117	0.44	-0.01	May 7	April 21	April 13
Poaceae	Göteborg	-0.115	0.20	0.02	June 1	May 23	May 18
Quercus	Västervik	-0.087	0.50	-0.02	May 17	May 9	May 5
Poaceae	Stockholm	-0.042	0.50	-0.01	June 2	May 29	May 27
Betula	Västervik	-0.010	0.94	-0.03	April 22	April 20	April 19
Poaceae	Umeå	-0.008	0.95	-0.02	June 15	June 14	June 13
Betula	Umeå	-0.002	0.99	-0.02	May 12	May 3	April 28
Poaceae	Eskilstuna	-0.002	0.99	-0.02	May 29	May 29	May 29
Ulmus	Umeå	0.009	0.98	-0.05	June 8	April 26	April 3
Betula	Jönköping	0.065	0.58	-0.02	April 21	April 24	April 25
Poaceae	Västervik	0.129	0.32	0.00	May 21	May 30	June 4

Table 10: Peak of pollen season for different species and locations

Species	Location	Slope	P- value	Adj. R^2	Estimated peak 1973	Predicted peak 2023	Predicted peak 2050
Corylus	Umeå	-1.127	0.00	0.41	May 27	March 19	February 9
Alnus	Umeå	-0.771	0.00	0.28	May 10	March 23	February 25
Salix	Eskilstuna	-0.752	0.00	0.47	May 19	April 20	April 5
Corylus	Västervik	-0.688	0.01	0.15	April 9	March 11	February 23
Quercus	Umeå	-0.610	0.03	0.38	July 11	May 20	April 22
Corylus	Norrköping	-0.534	0.05	0.08	April 23	March 11	February 16
Alnus	Eskilstuna	-0.523	0.00	0.43	April 18	March 19	March 2
Corylus	Jönköping	-0.504	0.12	0.05	March 26	March 15	March 10
Alnus	Stockholm	-0.501	0.00	0.32	April 9	March 21	March 10
Corylus	Eskilstuna	-0.474	0.00	0.19	April 14	March 15	February 28
Corylus	Malmö	-0.442	0.06	0.07	April 2	March 9	February 25
Salix	Stockholm	-0.432	0.00	0.27	May 10	April 16	April 4
Salix	Malmö	-0.430	0.00	0.26	May 14	April 26	April 16
Salix	Norrköping	-0.422	0.00	0.31	May 3	April 15	April 5
Ulmus	Västervik	-0.416	0.03	0.10	April 24	April 5	March 26
Ulmus	Eskilstuna	-0.415	0.00	0.26	May 3	April 13	April 2
Alnus	Malmö	-0.392	0.05	0.08	April 1	March 12	March 2
Alnus	Göteborg	-0.376	0.02	0.11	April 9	March 29	March 23
Salix	Göteborg	-0.374	0.00	0.25	May 3	April 15	April 6
Ulmus	Stockholm	-0.373	0.00	0.15	April 30	April 13	April 4
Corylus	Stockholm	-0.347	0.05	0.06	April 7	March 15	March 3
Alnus	Norrköping	-0.342	0.04	0.10	April 11	March 21	March 10
Ulmus	Göteborg	-0.342	0.02	0.11	April 28	April 6	March 25
Ulmus	Malmö	-0.336	0.23	0.01	April 20	April 8	April 2
Salix	Umeå	-0.323	0.10	0.04	May 27	May 21	May 18
Poaceae	Malmö	-0.301	0.00	0.22	July 2	June 17	June 9
Betula	Stockholm	-0.298	0.00	0.28	May 15	April 30	April 22
Poaceae	Jönköping	-0.285	0.01	0.17	July 10	June 25	June 17
Salix	Jönköping	-0.283	0.07	0.07	April 30	April 19	April 14
Betula	Eskilstuna	-0.269	0.00	0.19	May 13	April 29	April 21
Poaceae	Göteborg	-0.266	0.01	0.13	July 7	June 23	June 15
Quercus	Stockholm	-0.266	0.00	0.27	May 31	May 19	May 12

			P-	Adj.	Estimated peak	Predicted peak	Predicted peak
Species	Location	Slope	value	Rˆ2	1973	2023	2050
Poaceae	Eskilstuna	-0.264	0.00	0.23	July 8	June 25	June 18
Betula	Malmö	-0.248	0.00	0.18	May 9	April 23	April 15
Betula	Göteborg	-0.239	0.00	0.17	May 9	April 28	April 22
Alnus	Jönköping	-0.235	0.29	0.00	April 4	March 15	March 5
Ulmus	Jönköping	-0.233	0.27	0.01	April 29	April 11	April 2
Poaceae	Norrköping	-0.208	0.06	0.07	June 30	June 24	June 21
Alnus	Västervik	-0.202	0.32	0.00	April 2	March 20	March 13
Corylus	Göteborg	-0.199	0.41	-0.01	April 4	March 29	March 27
Quercus	Malmö	-0.192	0.08	0.06	May 26	May 14	May 7
Quercus	Göteborg	-0.179	0.07	0.05	May 29	May 18	May 12
Quercus	Norrköping	-0.177	0.11	0.04	May 27	May 19	May 15
Quercus	Västervik	-0.176	0.12	0.04	May 31	May 20	May 14
Salix	Västervik	-0.156	0.39	-0.01	April 25	April 22	April 21
Ulmus	Norrköping	-0.154	0.57	-0.02	April 21	April 25	April 27
Poaceae	Stockholm	-0.128	0.06	0.05	July 5	June 28	June 24
Quercus	Eskilstuna	-0.127	0.14	0.03	May 28	May 19	May 14
Poaceae	Umeå	-0.112	0.15	0.03	July 14	July 10	July 8
Betula	Umeå	-0.047	0.66	-0.02	May 27	May 18	May 12
Poaceae	Västervik	-0.043	0.74	-0.03	June 23	June 30	July 4
Betula	Västervik	-0.021	0.85	-0.03	May 8	May 1	April 27
Quercus	Jönköping	-0.007	0.95	-0.03	May 27	May 23	May 20
Betula	Norrköping	-0.004	0.97	-0.03	May 7	April 30	April 26
Betula	Jönköping	0.048	0.66	-0.02	May 8	May 2	April 29
Ulmus	Umeå	0.800	0.11	0.08	May 25	May 17	May 13

Table 11: End of pollen season for different species and locations

			P-	Adj.	Estimated end	Predicted end	Predicted end
Species	Location	Slope	value	Rˆ2	1973	2023	2050
Corylus	Norrköping	-0.747	0.00	0.35	May 6	March 27	March 5
Alnus	Norrköping	-0.576	0.00	0.29	April 28	April 5	March 23
Alnus	Umeå	-0.570	0.00	0.30	May 21	April 8	March 16
Corylus	Umeå	-0.565	0.28	0.01	June 7	March 29	February 19
Alnus	Eskilstuna	-0.521	0.00	0.55	May 6	March 30	March 11
Corylus	Jönköping	-0.481	0.10	0.05	April 9	April 8	April 8
Ulmus	Västervik	-0.445	0.01	0.17	May 9	April 18	April 7
Corylus	Eskilstuna	-0.444	0.00	0.22	April 26	March 31	March 16
Salix	Malmö	-0.439	0.00	0.26	June 7	May 12	April 28
Corylus	Västervik	-0.437	0.03	0.11	April 23	March 30	March 17
Ulmus	Eskilstuna	-0.434	0.00	0.33	May 13	April 21	April 9
Ulmus	Umeå	-0.376	0.48	-0.02	June 10	May 19	May 7
Alnus	Stockholm	-0.351	0.00	0.16	April 16	April 18	April 20
Alnus	Malmö	-0.344	0.03	0.09	April 16	April 1	March 23
Alnus	Göteborg	-0.343	0.02	0.11	April 22	April 10	April 4
Ulmus	Jönköping	-0.339	0.28	0.00	May 28	April 24	April 5
Ulmus	Stockholm	-0.335	0.01	0.11	May 7	April 26	April 20
Corylus	Stockholm	-0.332	0.01	0.11	April 21	April 2	March 23
Poaceae	Eskilstuna	-0.328	0.00	0.16	August 17	July 31	July 21
Corylus	Malmö	-0.323	0.10	0.05	April 14	April 4	March 29
Betula	Malmö	-0.294	0.01	0.16	May 21	May 9	May 2

Species	Location	Slope	P-value	Adj. R^2	Estimated end 1973	Predicted end 2023	Predicted end 2050
Ulmus	Malmö	-0.293	0.19	0.02	April 28	April 21	April 17
Ulmus	Göteborg	-0.290	0.01	0.14	May 3	April 17	April 8
Salix	Jönköping	-0.284	0.10	0.05	May 26	May 18	May 15
Alnus	Jönköping	-0.260	0.09	0.06	April 17	April 3	March 27
Quercus	Stockholm	-0.257	0.00	0.34	June 7	May 27	May 22
Betula	Göteborg	-0.249	0.01	0.14	May 20	May 11	May 6
Betula	Eskilstuna	-0.239	0.00	0.17	May 26	May 14	May 7
Salix	Stockholm	-0.231	0.00	0.17	May 30	May 19	May 13
Salix	Umeå	-0.217	0.04	0.08	June 9	June 10	June 11
Quercus	Umeå	-0.212	0.44	-0.04	July 1	May 31	May 15
Quercus	Malmö	-0.205	0.10	0.05	June 7	May 22	May 14
Betula	Stockholm	-0.204	0.00	0.17	May 27	May 16	May 10
Salix	Eskilstuna	-0.200	0.00	0.17	May 27	May 20	May 16
Poaceae	Umeå	-0.197	0.02	0.11	August 3	July 27	July 22
Quercus	Göteborg	-0.193	0.05	0.07	June 6	May 25	May 19
Quercus	Västervik	-0.183	0.10	0.05	June 8	May 27	May 21
Quercus	Eskilstuna	-0.180	0.07	0.05	June 7	May 27	May 20
Salix	Norrköping	-0.172	0.28	0.01	May 13	May 15	May 17
Salix	Göteborg	-0.151	0.15	0.03	May 20	May 15	May 11
Alnus	Västervik	-0.107	0.56	-0.02	April 14	April 6	April 2
Betula	Umeå	-0.085	0.41	-0.01	June 9	June 1	May 28
Quercus	Norrköping	-0.083	0.45	-0.01	May 31	May 29	May 28
Betula	Norrköping	-0.080	0.45	-0.01	May 21	May 13	May 9
Poaceae	Jönköping	-0.028	0.84	-0.03	August 8	August 1	July 29
Betula	Västervik	-0.005	0.97	-0.03	May 22	May 16	May 13
Betula	Jönköping	0.003	0.98	-0.03	May 22	May 13	May 9
Quercus	Jönköping	0.010	0.92	-0.03	June 2	May 31	May 30
Corylus	Göteborg	0.068	0.77	-0.02	April 13	April 29	May 7
Ulmus	Norrköping	0.088	0.76	-0.03	April 29	May 6	May 10
Poaceae	Stockholm	0.205	0.00	0.14	August 8	August 18	August 24
Poaceae	Malmö	0.220	0.16	0.03	July 27	August 9	August 17
Poaceae	Norrköping	0.261	0.15	0.03	July 17	August 9	August 22
Salix	Västervik	0.267	0.12	0.04	May 8	May 26	June 5
Poaceae	Göteborg	0.289	0.06	0.06	July 29	August 14	August 23
Poaceae	Västervik	0.638	0.00	0.18	June 29	August 19	September 16