Aerobiologia
DOI 10.1007/s10453-017-9488-0

ORIGINAL PAPER

Statistical modeling, forecasting and time series analysis of birch phenology in Montreal, Canada

Alain Robichaud · Paul Comtois

Received: 26 November 2016 / Accepted: 6 June 2017
© Springer Science+Business Media B.V. 2017

Abstract The aim of this study was to analyse birch pollen time series observed in Montreal (Canada) in order to understand the link between inter-annual variability of phenology and environmental factors and to build predictive models for the upcoming pollen season. Modeling phenology is challenging, especially in Canada, where phenological observations are rare. Nevertheless, understanding phenology is required for scientific applications (e.g. inputs to numerical models of pollen dispersion) but also to help allergy sufferers to better prepare their medication and avoidance strategies before the start of the pollen season. We used multivariate statistical regression to analyse and predict phenology. The predictors were drawn from a large basin (over 60) of potential environmental predictors including meteorological data and global climatic indices such NAO (*North Atlantic Oscillation index*) and ENSO/MEI (*Multivariate Enso Index).* Results of this paper are summarized as follows: (1) an accurate forecast for the upcoming season starting date of the birch pollen season was obtained (showing low bias and total forecast error of about 4 days in Montreal), (2) NAO and ENSO/MEI indices were found to be well correlated (i.e. 44% of the variance explained) with birch phenology, (3) a long-term trend of 2.6 days per decade ($p < 0.1$) towards longer season duration was found for the length of the birch pollen season in Montreal. Finally, perturbations of the quasi-biennial cycle of birch were observed in the pollen data during the pollen season following the Great Ice Storm of 1998 which affected south-eastern Canada.

Keywords Phenology · Airborne pollen · Birch (*Betula*) · NAO · ENSO/MEI · Statistical analysis and modeling

A. Robichaud (✉) · P. Comtois
Department of Geography, University of Montreal, 2900 Boulevard Edouard-Montpetit, Montreal QC H3T 1J4, Canada
e-mail: musgen13@gmail.com

1 Introduction

A significant proportion of Canadians and citizens throughout the world suffer from pollinosis (WHO 2003; CLA 2015). To improve the quality of life of sufferers (by avoidance strategies, adjust medication or desensitization programs), physicians need aeropalynological information and knowledge of phenology as accurate as possible before the start of the pollen season.[1] An increasing atmospheric variability is likely to occur in the context of the ongoing climate change (IPCC 2007a, b) therefore impacting phenology (Walther et al. 2002) making the forecast of phenology and spatio-temporal distribution of

[1] Asthmatic patients usually need to begin their anti-allergic treatment one or two weeks before pollination (Laaidi 2001).

Published online: 15 June 2017

airborne pollen more challenging in the future. Globally, for many plant species, the pollen season duration has already increase in about 10 days on average during the period from 1970 to 2000 in Europe (Tamburlini 2002). Since birch pollen is the most allergenic airborne pollen in Canada during springtime (Guérin 1993; Dales et al. 2008), phenological models of birch are therefore crucial to support pollen forecast today as well as in a changing climate of the future. However, phenology is challenging because: (1) it shows large inter-annual variations for most plant species (Andersen 1991; Laaidi 2001; Laaidi et al. 2002) and (2) there is at the current time no suitable phenological model available to represent the large scale (Siljamo et al. 2008a, b) and no consensus on which modeling approach is the best (Chuine et al. 1998; see also the review in Scheifinger et al. 2013). Two broad classes of phenological models emerge from the literature: (a) thermal forcing models based on the amount of spring warming and (b) more complex models including the effect of winter chilling (see a review in Sofiev and Bergmann, 2013). Due to its simplicity, thermal models have been mostly used to represent phenological models (Andersen, 1991; Laaidi 2001; Siljamo 2013). The first known model predicting the date of bud bursting for plants was introduced by de Réamur (1735) who deduces a general linear relation between thermal energy and the state of bud growth. More than two centuries later, Faust (1989) similarly reports that the spring blooming date for deciduous fruit trees is strongly dependent on temperatures in the weeks before pollination while Spieksma et al. (1995) noticed that air temperature during the period of 2 months preceding the start of the season is the dominant factor. Emberlin et al. (2002) found out that the birch phenology in Europe depends on the nonlinear balance between winter cooling required during the dormancy period and spring temperatures during quiescence period. For Rousi and Heinonen (2007), the blooming date in Scandinavia is regulated by the sum of degree-days greater than 5 °C (blooming occurs when the sum reaches 160°-days ± 10 accumulated since January 1st). On the other hand, Oikonen et al. (2005) found little correlation between the starting day of the birch pollen season and growing degree-days for the same region. In France (Burgundy region), Clot (2001) showed that an accurate forecast of the onset of the birch pollen season in Neuchâtel (Switzerland) can be made by using the cumulative temperature sum above zero from February 1st onward until 270 °C is reached. The trees are then ready to bloom as soon as the daily averaged temperature exceeds 10 °C. More complex models using local meteorological variables to define a state of forcing (warming) and state of chilling have also been proposed (Chuine et al. 1998; Rea and Eccel 2006) as well as mechanistic models (Garciá-Mozo et al. 2008). However, other authors found that phenology might also depend on other meteorological factors not only prior to the pollen season but also in the preceding year. Jones (1995) pointed out that the blooming is dependent on the amount of rainfall from several months prior to the start of the birch pollen season. Laaidi (2001) found that the length of the birch pollen season is given in France by the following two predictors: the minimum relative humidity of October of the previous year and the maximum temperature of February of the current year. Physically, this means that the duration of the season will be short if last October was humid and/or February of the current year was cold. Similarly, Rasmussen (2002) also noted a negative correlation between the total annual *Betula* airborne pollen concentration and precipitation in the previous growing season for *Betula* in Denmark. Similarly, in California, Fairley and Batchelder (1986) observed a strong correlation between the oak pollen abundance measured during the current season with the sum of precipitation of the previous year in California (USA). Andersen (1991) showed that the intensity of birch and alder pollen seasons in Denmark is related to the precipitation in April of the previous year whereas for oak and beech, the intensity would depend on the average temperature of the previous year. Other authors in Europe have confirmed the importance of the meteorological conditions during the previous years for tree pollen production to predict the upcoming season severity (Stanley and Linkens 1974; Faegri and Iversen 1989). The influence of the NAO (North Atlantic Oscillation index) on phenological processes is another feature which has emerged within the last decade or so in Europe (D'Odorico et al. 2002 and references therein; Avalo et al. 2008 and references therein). Stach et al. (2008) recognized that the severity of the birch pollen season in Poland and the UK is linked in some way to the different phases of the NAO. Winter NAO index is one of the most important predictor of the start of grass pollen season in Poznań (Poland) according to Stach et al. (2008). In

general, in Europe, it seems that NAO governs the temporal variability of the lower atmosphere and thus phenology (Scheifinger et al. 2002). In any cases, NAO is considered as one of the most important teleconnection patterns in all seasons (Barnston and Livezey 1987). Small changes of NAO could produce large impact in terrestrial ecosystems (Ottersen et al. 2001). The NAO index is based on the difference of normalized sea level pressure anomalies between Azores and Iceland. The NAO could be described as a north–south dipole anomaly which shows considerable intra-seasonal and inter-annual variability and produce very significant changes in temperature and precipitation patterns from eastern North America to western and central Europe. In the northern hemisphere, the NAO index shows a significant correlation with temperature and precipitation (Hurrell and van Loon 1997). In order to show the importance of the different phases of NAO, Table 1 summarizes its impact on the weather in the northern hemisphere.

ENSO (*El Niño/La Niña-Southern Oscillation*) is another prominent teleconnection pattern. ENSO is the most important coupled ocean–atmosphere phenomenon to cause global climate variability on inter-annual time scales. It is a quasi-periodic climate pattern that occurs across the tropical Pacific Ocean roughly every 5 years (Trenberth et al. 2007). The *Southern Oscillation* refers to variations in the temperature of the ocean surface (warming and cooling known as *El Niño* and *La Niña,* respectively) as well as in air surface pressure pattern. *El Niño* is characterized by high air surface pressure in the western Pacific, while the cold phase, *La Niña* is associated with low air surface pressure in the western Pacific. Precise mechanisms that cause the oscillation still remain unclear (Trenberth et al. 2007). ENSO in its acute phase causes extreme weather in many regions of the world in any season and any location particularly those bordering the Pacific Ocean which is the most affected. One of the specific forms of ENSO is the MEI (Multivariate ENSO Index) which is the atmospheric expression of ENSO and is based on the six main observed variables over the tropical Pacific (sea level pressure, zonal and meridional components of the surface wind, sea surface temperature, surface air temperature and total cloudiness fraction of the sky) (NOAA 2012a, b). ENSO was also found related to phenology. For example, Rozas and García-González (2012) showed a relationship between the growth of oak in the Northwest Iberian Peninsula and the ENSO of the previous year. In our study, NAO and ENSO indices of the previous months up to about 1 year are introduced as predictors for the phenological parameters of the upcoming season in addition to more traditional precursors (e.g. mostly based on temperature and precipitation). The aim of this research to study the time series of birch airborne pollen for Montreal (Canada) in order to: (1) build statistical forecast of phenology parameters for the upcoming season (starting date, length and severity of the pollen season), (2) analyse the link between phenology and the environmental factors described above, (3) evaluate current trends over the study period of the phenological parameters and (4) perform a spectral analysis of the phenological observations during the period 1996–2012.

2 Data and methodology

2.1 Study area

The region of study comprises the Montreal metropolitan area which lies in the extreme south of

Table 1 Impact of NAO (adapted from NOAA, 2012 and references therein)

Weather pattern and link with NAO	Positive NAO	Negative NAO
Atmospheric pressure patterns across the high latitudes of the North Atlantic	Below	Above
Pressure over Central North Atlantic, Eastern US and Canada and Western Europe (increase in the zonal flow)	Above	Below
Temperatures in Eastern North America and across Western Europe	Above	Below
Temperatures in Greenland, Iceland and Southern Europe	Below	Above
Precipitation over Northern Europe, Scandinavia in winter	Above	Below
Precipitation over Southern/Central Europe	Below	Above

Below: weaker than average, *Above:* stronger than average

the province of Quebec (Canada) near the USA border (see location of Montreal identified by the letter M in Fig. 1). Montreal is an island which covers about 365 km^2 with a population of the metro region of about 3.8 million inhabitants.[2] The latitude of Montreal is 45°30′ North and the longitude is 73°33′ West with an altitude of approximately 30–45 m above the sea level. This region is interesting because airborne birch (*Betula*) pollen concentration could be at times quite high (Bapikee 2005) then affecting a large human population basin while the abundance of local birch tree population is low (density is less than 0.05, i.e. 5%) over the island of Montreal and immediate surroundings (see distribution of birch tree density in Fig. 1). In fact, hardwoods, pine and hemlock are the most characteristic tree species in Montreal (Braun 1950). Birch typically flower at the end of April or beginning of May in Southern Quebec. Montreal is characterized by a temperate climate with short and warm summers but long and cold winters with yearly precipitation around 1000 mm annually.

2.2 Airborne pollen data

In the province of Quebec, phenological records are rare and in any cases not available for this study (no existing network of phenological measurements). This could be seen as a handicap but according to Scheifinger et al. (2002) and García-Mozo et al. (2008), phenological models fitted with airborne pollen and meteorological data measured locally yields the best results. All pollen data used in this study were collected at Université de Montréal (thereafter UdeM) and concerns birch (*Betula*) tree. Birch pollen is studied here since it is considered the most allergenic tree species in eastern Canada (Guérin, 1993). Beside UdeM site, no other location in Quebec has recorded a long series of pollen data necessary to build and test statistical models such as presented in this paper and so this is also why Montreal is convenient for our study. The airborne pollen concentration in Montreal has been measured for more than 25 years at this site. A Lanzoni VPPS sampler (two stage vacuum pump) based on the Hirst method (1952) was used. This is a seven days recording volumetric pollen and spore trap installed on the rooftop of the geography department at UdeM roughly 20 m above the street level which has been providing routine pollen monitoring for the whole study period. The sampler at the site is calibrated to handle a flow of 10 litres of air per minute which roughly corresponds to human breathing. Pollen grains impact a cylindrical drum covered by a melinex-coated film. A 7-day clockwork allows an hourly resolution (Mandrioli et al. 1998). The last step is the pollen counting which is done under a light microscope. Limitations of this type of measurements are numerous: electric power outage may occur, malfunctioning of the clock, obstruction of the inlet and the risk of uncertainty may be potentially important regarding human manipulation. For example, laboratory methodology, microscope manipulation and other errors may introduce biases (INSPQ 2013).

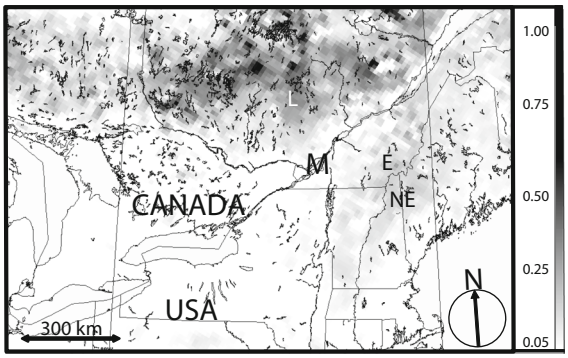

Fig. 1 Distribution of birch (*Betula*) vegetation fraction (density from 0 to 1) around the city of Montreal (*M*) and in the southern part of Canada and New England states in USA. The major birch pollen area sources are indicated on the figure: Laurentians to the north (L), Eastern Townships (E) to the east and New England (NE) to the south-east. The figure was processed from raw data obtained from US/EPA (BELD3 database, https://www.epa.gov/air-emissions-modeling/biogenic-emissions-landuse-database-version-3-beld3)

In the Montreal area, since there is no phenological recording site, the correspondence between local pollination and measured pollen is not available. However, Fig. 1 suggests that the distribution of birch in the Montreal area is rather weak (less than 5% density, i.e. <0.05) which implies that most of the birch pollen measured at the site comes from regional transport of scale of up to several hundreds of kilometres from different source regions: Eastern Townships and the Appalachian mountains to the east, New England in USA to the south-east or the Laurentian region to the north and to the west.

[2] http://ville.montreal.qc.ca/portal/page?_pageid=6897,67887840&_dad=portal&_schema=PORTAL.

Therefore, we assume here for Montreal that the vast majority of birch pollen is of foreign origin, i.e. from outside the Greater Montreal area (as suggested in Bapikee 2005).

2.3 Meteorological data and global weather indices

Temperature and precipitation were available from the Meteorological Service of Canada at Dorval-Trudeau airport (latitude 45.47°N, longitude −73.75°W) located approximately 13.5 km to the south-west from the pollen monitoring station (latitude 45.50°N, longitude −73.62°W). Since the pollen data (measured at UdeM) is representative of an area of 250 km^2 (Comtois and Gagnon 1988), we then consider that the meteorological measurements were taken within the same atmospheric environment as the pollen site (i.e. urban conditions). Only under an east and south-east winds, the effect of topography (from nearby Mount Royal, about 250 m height) will be felt at the pollen site and not at the airport. However, these wind directions are less frequent so on a statistical basis, meteorology conditions are not expected to be significantly different between the two sites.

NAO index is taken from the following NOAA website as a form of bi-monthly average index (see https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/NAO/ for further information). ENSO index used is the MEI normalized bi-monthly values (multivariate ENSO index), and average values were also taken from NOAA website. Note that we limit ourselves to the period after 1995 since ENSO index time series entered a different regime compared to the period before. Moreover, the methodology changed after the mid 90s on how to compute MEI index which limits the use of data before that period (see https://www.esrl.noaa.gov/psd/enso/mei/index.html). Table 2 summarizes the data used in our study.

2.4 Definition for the phenological parameters

In this paper, the start of the pollen season is defined as when 2.5% of the total pollen annual catch is reached and the end of the season as when 97.5% of the total pollen of the season has been recorded (following Goldberg et al. 1988). Bapikee (2005) found that, for the airborne pollen measured at the UdeM site, the latter definition was the most robust for the purposes of phenological studies. In the rest of this study, we will then refer as to the D2 method following Bapikee (2005). The definition for the duration of the season (in unit of days) is simply the difference between the Julian day corresponding to the end and the start of the season, respectively. Similarly, the total airborne pollen concentration or seasonal pollen index (SPI) is calculated as the cumulated sum of daily pollen concentration during the same period, i.e. when the total annual pollen is greater or equal than 2.5% and less or equal than 97.5% of the total annual catch (see Andersen, 1991 or Goldberg et al. 1988). Finally, the peak value is defined as the maximum daily concentration for a given pollen season.

Table 2 Description of data used in this study and their source

Data	Institution	Method/instrument/variable measured	Curve fittings (Years used)	Cross-validation (Years used)
Pollen (*Betula*)	UdeM	Lanzoni VPPS volumetric pollen trap (pollen count)	1996–2009	2010
			1996–2010	2011
			1996–2011	2012
NAO index[a]	NOAA	Bi-monthly mean NAO indices	idem	idem
ENSO index[b]	NOAA	Monthly mean ENSO/MEI indices	idem	idem
Meteorology[c]	Environment Canada	Monthly temperature and precipitation at Trudeau-Dorval Airport (YUL)	idem	idem

[a] https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/NAO

[b] https://www.esrl.noaa.gov/psd/enso/mei/index.html

[c] http://climate.weather.gc.ca

2.5 Statistical method

In order to model phenology, we need to establish a relation between a phenological stage (the predicted or response variables P) and several weather or aeropalynological parameters (explanatory independent predictors Y). Formally, the relation between a predictand P_j (endogenous variables) and n predictors (exogenous variables) could be written as (Weisberg 2013):

$$P_j = a_{oj} + a_{1j}Y_{1j} + a_{2j}Y_{2j} + a_{3j}Y_{3j} + \ldots a_{nj}Y_{nj} + \varepsilon \quad (1)$$

where a_{nj} are the coefficient of regression, Y_{nj} the values of the predictors and ε a term which accounts for the random error. In this study, more than 60 different predictors were initially considered and selected according to information found in the literature (see a review in Introduction section). These predictors include both monthly and seasonal temperature (minimum, average and maximum), monthly and seasonal precipitation (autumn or winter rain amount) of both the current and the previous year, different combination of sum of degree-days, various combination of past NAO and ENSO indices of the current and previous year, various combinations of past phenological parameters (e.g. previous year starting date, length of the season and SPI). The complete list of potential predictors initially considered in this study is given in Appendix 1. The four phenological statistical predictands P are mstart (model for the start of the birch pollen season expressed in Julian day), mlength (model for the length duration of the pollen season expressed in number of days), mspi (model for the total pollen seasonal index, which is unitless) and mpeak (model for the highest peak of the season given in grains per cubic metre).

Multivariate linear regression methods have been used to solve Eq. 1. These methods include regression models in which the final selection of predictive variables is carried out by using a automatic robust statistical procedure (stepwise regression, SAS®, 1989) with the input and output carefully scrutinized by the researcher (see description of the methodology in Appendix 2). SAS-STAT® (Statistical analysis software package) version 9.2 (for Windows) has been used for all statistical treatment (using different stepwise procedures) except for the spectral analysis of time series where SAS-ETS® was used. First, simple correlations between the potential predictors and the predictand were calculated according to the Pearson's coefficient of correlation to screen out undesirable variables and obtained preselected predictors (i.e. those having a p value <0.15). Then the methodology described in Appendix 2 was used for each predictand in a sequential way with this subset of the initially preselected variables in order to identify the best non-colinear predictors. Avoidance of over-fitting the model to the noise in the data is addressed by the C_p statistics which is used as a stopping rule in the process of stepwise regression. C_p is given by the following (Mallows 1973):

$$C_p = \mathrm{SSE}_p / S^2 - n + 2\,\mathrm{m} \quad (2)$$

where SSE_p is the error sum of squares for the model with m regressors, S^2 is the residual mean square after regression on the complete set of predictors and n is the sample size. Whenever C_p is roughly equals to the number of predictors m, the model is not considered to be over-fitted. As C_p drops below m, adding up predictors would cause model over-fitting (Mallows 1973; Daniel and Wood 1980). Note that if the data set is too small (i.e. n not larger than m in Eq. 2), inflated regression coefficient in multiple regression could appear (Scheifinger et al. 2013). Note also that multicollinearity between predictors is avoided by checking the correlation of the final predictors and carefully removing those predictors being correlated with others which do not bring new information. Note that another multiple regression method has also been used as an alternative to check the robustness and the validity of the main predictors found with the stepwise multiple regression procedure. This alternative method is based on the minimization of the information content called the Akaike's information criteria (thereafter AIC) for objectively selecting the predictors (see Beal 2005 for more details). This latter method was found to give slightly better coefficient of determination (few per cent higher) and slightly better curve fit but was giving poor performance with independent tests and statistical global F-test and t-test.

The statistical predictors tested in our study (Appendix 1) could be classified into four groups: (1) local meteorological predictors from the current year (e.g. March temperature (tmarch), degree-days over 5 °C accumulated since January 1st at a specific Julian day

before the pollen season (e.g. sum05_115)), (2) local predictors from the previous year (e.g. autumn rainfall (lrainf), mean maximum temperature of the previous summer (ltsummax)), (3) global index such as NAO, ENSO of the current (prior to blooming) and from the previous year, and, finally, (4) predictors based on persistence in order to take advantage of the time autocorrelation of the predictands, e.g. duration of the season (llength), end date registered last year (lendD2) and SPI for the previous year (lspi). Note that the letter *L* means 1 year lag for the predictors. Many other studies have noticed similar correlation with several of these predictors mentioned above (see Introduction). But it is the first time, as far as the authors are aware that these correlations are made with Canadian or North America data and with a very large basin of potential predictors (about 60 here). Finally, it is important to be aware that it is not proven that explanatory variables (right-hand side of Eq. 1) describe any cause and effect direct link with the predictands. Since our primary goal is not to find cause and effect but to construct a phenological *forecast* model for the upcoming season, we will take advantage of these correlations for mostly prediction purpose of the phenological parameters. Nevertheless, the physical meaning and the possibility of causal behaviour of predictors will be discussed.

2.6 Periodogram

Periodograms are plots of spectral energy whose peaks identify return periods or phenological cycles. We have used the procedure SPECTRA of SAS/ETS (SAS® 1989) to plot the periodograms of the four phenological parameters under study. The SPECTRA procedure computes the spectral densities of a time series.

2.7 Cross-validation

A way to validate phenological models created by multiple regression is to assess the curve fittings against a set of data that was not used to create them (Mayers and Forgy 1963; Mark and Goldberg 2001). This is referred to cross-validation or "leave-one-out" procedure. Three periods, 1996–2009, 1996–2010 and 1996–2011 were chosen to build curve fittings and the corresponding verification years 2010, 2011 and 2012, respectively, are used as independent validation data (but were not utilized in the data set used to build the main curve fittings). Regression equations were calculated for the three periods above mentioned and applied sequentially in a predictive mode to the year following the end of the corresponding retrofit period. These three periods are called training periods thereafter. For example, to produce a forecast for 2010, we use the period 1996–2009 curve fitting, for the forecast of 2011 we use the period 1996–2010 and for the forecast of 2012, the fitting 1996–2011. We then have a sample of 3 years to validate the statistical model in the forecast mode in a total independent way. Comparisons with a model based on average aeropalynological data (average values of the predictand over a long period, i.e. calendar's method) and persistence (last year predictand value) are also presented for comparison. We then are able to evaluate the true forecast skill of the models for the upcoming pollen season while testing them using the following year after the end of the fitting period. Finally, we compare the results with a forecast based on LTAA (calendar's method) or persistence. Accuracy is measured by the standard deviation and mean of the error residuals (Observed minus Predicted values, thereafter OmP) in both cases: for independent (external) data validation (2010–2012) and as well as for internal validation (period 1996–2009). The choice of this set-up for validation is justified by the desire to forecast phenology for the upcoming season as explained above. Note that a similar validation procedure has been used by both Laaidi (2001) and Adams-Groom et al. (2002) for predicting birch phenology for the upcoming year.

3 Result

Basic descriptive statistics of the phenological parameters observed in Montreal (UdeM site) for the period 1996–2012 are given in Table 3. A lot of inter-annual variability is noticeable for each of the phenological parameter as can be seen from the standard deviation and the range (max–min) which are very large. This is especially true for the observed seasonal pollen index (SPI) as well as for the peak value. As an example, SPI ranges from 173 (year 2000) to 11,570 (year 1998). For the rest of this chapter, the results presented in

Table 3 Basic statistics for the phenological parameters observed for birch pollen for the period 1996–2012 in Montreal

Variable	*N*	Mean	Median	SD	Min	Max	Q1 (25%)	Q3 (75%)
Start date (Julian day)	17	124.2 (May 4th)	122.5 (May 2nd)	9.68	111 (April 21st)	157 (June 6th)	121 (May 1st)	125 (May 5th)
Length (nb days)	17	30.0	29.0	7.77	12	67	25	37
SPI (unitless)	17	3629.7	2520	3135.4	173	11,570	1566	5166
Peak (grains/m^3)	17	742	561	859.3	24	3774	289	816

Length is the duration of the season, SPI stands for the seasonal pollen index and peak refers to the highest peak of the pollen season. Q1 and Q3 are the 25 and 75% quantiles, respectively

Table 3 will be used to represent the LTAA (long-term aeropalynological average). Note that an LTAA-based forecast is equivalent to the so-called *calendar's* method. We now examine predictive models for the four phenological parameters.

3.1 Predictive model for the day of the start of the birch pollen season (mstart)

One of the most important phenodate is the start of the pollen season (Latalowa et al. 2002 and references therein). Among the list of potential predictors (Appendix 1), the preselected predictors are those having significant coefficient of correlation with the observation d2start (i.e. p value <0.15; see Table 4). Equations (3)–(5) give the linear relationships for the final model selection for three training periods obtained from the stepwise procedure for the phenodate (predictand mstart) among a very large possibility of models.[3] Appendix 3 (Table 11A) shows that the strongest three predictors are lrainf (autumn total rainfall of the previous year), ltsummax (average maximum temperature during the summer of the previous year) and tmarch (average temperature of the month of March of the current year). Note that these three main predictors remain the same in Eqs. 3 through 5 indicating some stability of the equations against removing one particular year (as required by the cross-validation procedure). According to Table 11A (final stepwise summary), the secondary independent predictors which succeeded the numerous selection tests are linked with NAO and ENSO phenomena of the previous year (e.g. lnao1, lenso6 and lenso1). For the period 1996–2011 (Eq. 5), the final predictors retained from the STEPWISE procedure are lrainf (partial $R^2 = 0.4041$), ltsummax (partial $R^2 = 0.2437$), tmarch (partial $R^2 = 0.1846$), lnao1 (partial $R^2 = 0.0434$), lenso6 (partial $R^2 = 0.0405$) and lenso1 (partial $R^2 = 0.0235$). For other retrofit periods (1996–2009 and 1996–2010)

Table 4 Preselected predictors with significant Pearson's coefficient of correlation (i.e. $p \leq 0.15$) for the d2start (day of the start of the birch pollen season)

Predictor	Coefficient of correlation	*p* value
Lrainf	−0.6357	0.0081
Ltsummax	0.6147	0.0233
Rainm	**−0.5990**	**0.0142**
Ensowin	**−0.5661**	**0.0223**
Tmarch	**−0.5097**	**0.0437**
Lenso12	−0.5080	0.0445
Ltsum	0.5077	0.0450
Lensof	−0.5000	0.0484
Nao1	**0.4956**	**0.0510**
Lenso10	−0.4936	0.0520
Lenso11	−0.4924	0.0530
Lenso6	−0.4757	0.0625
Lensosum	−0.4735	0.0640
Lenso9	−0.4687	0.0671
Lenso8	−0.4549	0.0766
Lnao1	0.4532	0.0780
Tfev	**−0.4502**	**0.0802**
Twinmin	**−0.4410**	**0.0872**
Twinavg	**−0.4140**	**0.1113**
Nao2	**0.3885**	**0.1370**
Lenso1	0.3882	0.1370
Lnao5	−0.3820	0.1440

The description of the variables used is given in Appendix 1. Current year predictors are indicated by bold character and last year predictor by a letter starting by L (for lag)

[3] The number of possible model sis 2^p-1. With p = 22 preselected predictors (see Table 4), this gives over 4 million possibility of models which requires the use of automatic selection procedures.

similar results were found (corresponding to Eqs. 3 and 4, respectively). Although the predictors have not necessary a causative link with the predictand, this nevertheless suggests that a dry (wet) autumn, combined with a warm (chilly) summer during the previous year would be associated with a late (early) date of birch pollination for the current year upcoming season in the Montreal region. Similarly, a cold (warm) March temperatures of the current year is associated with a late (early) start of the season. Note that the main selected predictor of the statistical model (lrainf) shows colinearity with the mean autumn ENSO/MEI index (i.e. $R = 0.54$ between ensof and lrainf, p value <0.05) and consequently according to the procedure of Appendix 2, ensof does not appear in the final list of predictors. Nevertheless, this colinearity suggests a control of ENSO/MEI on lrainf.

The fact that NAO/ENSO seems to play a minor role in the final selection of predictors (see Table 11A, Appendix 3) even if their correlation coefficient were high with the predictand (ensowin, lenso12, lensof, nao1, etc. see Table 4) could be explained the following way: since their variability was already explained by local predictors (such as lrainf, tsummax and tmarch) some NAO/ENSO predictors will tend to be removed by the statistical procedure mostly because they show less predictive power and lower coefficient of correlation (see Table 4) than the three main predictors but also to avoid multicollinearity between predictors. Nevertheless, these predictors are considered to have an indirect effect on phenology. Figure 2 shows the retrofit curve fitting for the statistical model compared with observations for the period 1996–2011. For other training periods (1996–2009 and 1996–2010), similar results are obtained and therefore not shown. Note that the observed value for year 2000 (d2start occurred on Julian day 157 for that year) is considered here as an outlier (since the observed d2start is outside a range of 2.5 standard deviation, according to the statistics of Table 3 and also identified as having an abnormal Cook's distance: Cook 1977) as computed by the SAS procedure and therefore not used in building the retrofit models, nor used in the verification. The fitting of the model value for the start of the birch pollen season (mstart) against observation (d2start) are remarkable with $R^2 = 0.9079$ (period 1996–2009), 0.911 (period 1996–2010) and 0.9397 (period 1996–2011), respectively. The complete equation of

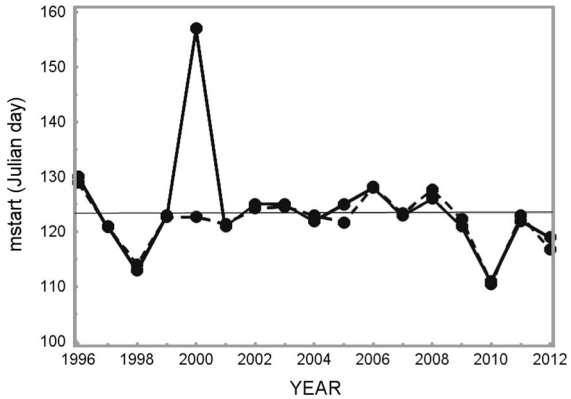

Fig. 2 Statistical model (mstart: *dotted lines*) versus observed start of the season (d2start: *solid lines*). The curve fittings hold for the period 1996–2011 and the forecast is for 2012. *Note* the horizontal line is the long-term average (LTAA). Year 2000 is considered as an outlier here

the fitting obtained for the three periods for the phenodate are given below:

Period 1996–2009 ($R^2 = 0.9079$)

$$\begin{aligned}\text{mstart} = {} & 38.81395 - 1.18656 * \text{tmarch} + 4.60 \\ & * \text{lenso9} - 6.74 * \text{lensosum} + 3.3462 \\ & * \text{ltsummax} - 0.11281 * \text{lrainf}\end{aligned} \tag{3}$$

Period 1996–2010 ($R^2 = 0.911$)

$$\begin{aligned}\text{mstart} = {} & 47.6025 - 0.9415 * \text{tmarch} - 1.84058 \\ & * \text{lenso6} + 2.9961 * \text{ltsummax} - 0.09949 \\ & * \text{lrainf}\end{aligned} \tag{4}$$

Period 1996–2011 ($R^2 = 0.9397$)

$$\begin{aligned}\text{mstart} = {} & 66.2478 - 0.9545 * \text{tmarch} + 1.8356 * \text{lnao1} \\ & + 0.8856 * \text{lenso1} - 1.86 * \text{lenso6} + 2.2258 \\ & * \text{ltsummax} - 0.05878 * \text{lrainf}\end{aligned} \tag{5}$$

The fact that the coefficient of determination R^2 does not change much from one equation to the other indicate robustness of the statistical procedure. Moreover, the main predictors (lrainf, ltsummax, tmarch and various forms of ENSO/NAO indices) remain present in all equations indicating consistency between the different training periods. Some variation of the coefficients and intercepts from one equation to another is due to the small number of data used ($N \sim 15$ years). Equations 3–5 provide three training periods which allow for 3 independent validations for 2010, 2011 and 2012, respectively (see details in Sect. 3.7.2).

3.2 Predictive model for the length of the season (mlength)

Using the same statistical methodology as above for the date of start of the season, we now examine the case of another phenological parameter: the duration of the pollen season (predictand mlength). The preselected predictors in this case are given in Table 5. With 18 preselected predictors, there are up to about 260,000 possible combinations of statistical models. The best model for the fitting period 1996–2011 and the final selected predictors appear in Table 11B (Appendix 3). For this case, the variance of the predictand length is best explained by the following predictors: the ENSO/MEI index of the month of May of the previous year (lenso5, partial $R^2 = 0.5059$), the NAO index of March of the previous year (lnao3, partial $R^2 = 0.1545$), the Julian date corresponding to the end of the season of the previous year (lendD2: partial $R^2 = 0.0775$), the pollen season duration of the previous year (llength: partial $R^2 = 0.1040$), the average annual ENSO/MEI index of the previous year (lensoa, partial $R^2 = 0.0377$), the February average temperature of the previous year (ltfev: partial $R^2 = 0.0316$), the NAO index of August of the previous year (lnao8, partial $R^2 = 0.0345$), the March temperature of the current year (tmarch, partial $R^2 = 0.0189$) and, finally, the ENSO/MEI index of April of the previous year (lenso4, partial $R^2 = 0.014$). The link established here between mlength (pollen season duration) and its predictors suggests that a strong positive value of ENSO/MEI (multivariate ENSO index) of May of the previous year (i.e. lenso5) tend to be associated with a short duration of the birch pollen season in the current year (due to the negative correlation, see Table 5). Note that a long duration of the previous year pollen season (llength) is negatively correlated with the current year duration suggesting some memory of the previous year phenology. In fact, most of the best predictors are related with variables during the previous year. Finally, according to our results, an above normal March temperature (tmarch) will tend to be associated with a longer than normal duration of the current year pollen season which reproduces known results found elsewhere (see a review in Scheifinger et al. 2002 and also the Discussion section below). Figure 3 provides model and observed curves for the period 1996–2011 (other training periods 1996–2009 and 1996–2010 show similar results and therefore not shown).

The overall fitting between model and observation is remarkable ($R^2 = 0.9892$ or 99% of the variance explained for the period 1996–2009, see Eqs. 6). The complete equation of the fittings obtained for all training periods are presented below:

Table 5 Preselected predictors with significant Pearson's coefficient of correlation (i.e. $p \leq 0.15$) for the duration of the season

Predictor	Coefficient of correlation	p value
Lenso5	−0.66,630	0.0035
Lnao6	0.59670	0.0115
Lensospr	−0.58221	0.0142
Lnao3	−0.56360	0.0287
Llength	−0.50651	0.038
Lenso4	0.49988	0.0410
Lensoa	0.44287	0.0750
Lenso1	−0.41007	0.1021
Lenso2	−0.40496	0.1069
Lensow	−0.40265	0.1091
Lenso6	−0.39478	0.1168
Ltfev	0.38065	0.1198
Lnso3	−0.37924	0.1333
tmarch	**0.37561**	**0.1390**
Lnao7	0.37330	0.1403
Lnao8	0.36587	0.1467
Lnao5	0.36498	0.1503
LendD2	−0.36151	0.1509

The description of the variables used is given in Appendix 1. Current year predictors are indicated by bold character and last year predictor by a letter starting by L (for lag)

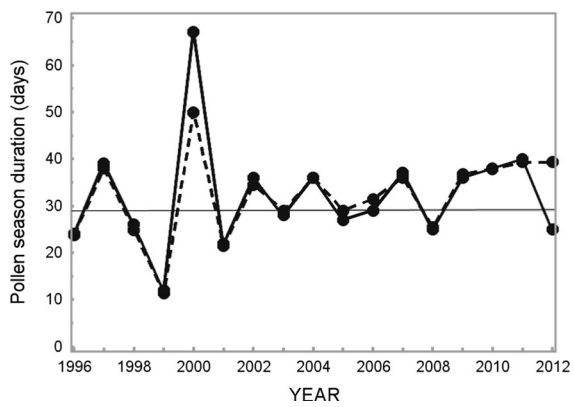

Fig. 3 Statistical model (mlength: *dotted lines*) versus observed season duration (*solid lines*) for the period 1996–2011. Year 1999 is considered as an outlier here

Period 1996–2009 ($R^2 = 0.9892$)

$$\text{mlength} = 65.95095 - 13.02086 * \text{lenso5} + 1.9640 * \text{tmarch} - 0.40245 * \text{llength} + 1.5490 * \text{ltfev} + 3.61 * \text{lnao5} + 2.75312 * \text{lnao7} \quad (6)$$

Period 1996–2010 ($R^2 = 0.9732$)

$$\text{mlength} = 61.11932 - 13.41878 * \text{lenso5} + 1.8412 * \text{tmarch} - 0.299984 * \text{llength} + 1.417 * \text{ltfev} + 2.1738 * \text{lnao5} + 2.99488 * \text{lnao7} + 1.33674 * \text{lnao8} \quad (7)$$

Period 1996–2011 ($R^2 = 0.9787$)

$$\text{mlength} = 109.37 - 24.87 * \text{lenso5} + 0.63087 * \text{tmarch} - 0.72614 * \text{llength} + 1.765 * \text{ltfev} - 1.864 * \text{lnao3} - 0.498 * \text{lendD2} + 4.661 * \text{lensoa} + 2.1704 * \text{lnao8} + 4.163 * \text{lenso4} \quad (8)$$

3.3 Predictive model for the birch seasonal pollen index (mspi)

The preselected predictors for input to the stepwise procedure for this predictand are given in Table 6 and the final predictors selected among a large numbers of possible models by the same procedure as used above are provided in Table 11C (period 1996–2011). For the seasonal pollen index curve fitting, results from the stepwise procedure provide the following results for the best independent predictors: average maximum temperature of the previous summer (ltsummax, partial $R^2 = 0.3881$), duration of the season of the previous year (llength: partial $R^2 = 0.1919$), temperature sum above 5 °C cumulated up to Julian day 115 of the current year (sum05_115, partial $R^2 = 0.1851$), NAO index of May of the previous year (lnao5, partial $R^2 = 0.0790$) and NAO index of August of the previous year (lnao8, partial $R^2 = 0.0740$). As previously seen for mstart and mlength, Fig. 4 shows that the fitting between model and observation is quite remarkable (for example, $R^2 = 0.9181$ for the period 1996–2011). Note that the results for other training periods (1996–2009 and 1996–2010) show similar results and are not shown. The link between the model seasonal pollen index (mspi) and its predictors

Table 6 Preselected predictors with significant Pearson's coefficient of correlation (i.e. $p \leq 0.15$) for SPI (seasonal pollen index)

Predictor	Coefficient of correlation	p value
Precipw	**0.66286**	**0.0098**
Sum03_110	**0.57579**	**0.0312**
Sum05_110	**0.54703**	**0.0429**
Sum03_115	**0.51000**	**0.0624**
Nao2	**−0.48312**	**0.0801**
Lnao5	−0.48297	0.0802
Nao1	**0.48281**	**0.0803**
Sum05_115	**0.45790**	**0.0997**
Ltsummax	0.45578	0.1014
Llength	0.45356	0.1015
Nao3	**−0.45248**	**0.1043**
Ltsum	0.40822	0.1473
Nnao8	0.37981	0.1455
Rainj	**0.28298**	**0.1445**

The description of the variables used is given in Appendix 1. Current year predictors are indicated by bold character and last year predictor by a letter starting by L (for lag)

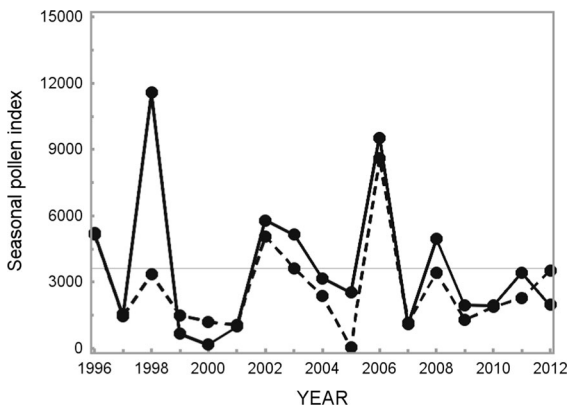

Fig. 4 Statistical model for seasonal pollen index (*dotted lines*) versus observed pollen seasonal index (*solid lines*) during the period 1996–2011. The year 1998 is considered as an outlier here

suggests again an influence of NAO index but no dependency with ENSO/MEI in this case. However, the dependency is stronger with local variables (ltsummax, llength and sum05_115) for this phenological parameter. Note that in Table 6, various form of the sum of temperature prior to the start of the season are well correlated with mspi but only one of these predictors (sum05_115) is selected to avoid

redundancy (i.e. multicollinearity). Note also that the peak of year 1998 is considered here as an outlier (according to Table 3, i.e. greater than 2.5 standard deviations and also because it shows an abnormal Cook's distance). Therefore, this particular year was not included in the computation of the statistical model for this predictand. However, the peak of 2006 was not considered as an outlier (since it lies within 2.5 standard deviations) and, in fact, is very well reproduced by the statistical model (Fig. 4). Equations 9–11 (see below) present the results of the stepwise procedure summary associated with mspi for the three training periods:

Period 1996–2009 ($R^2 = 0.9316$)

$$mspi = -51844 + 1857.3 * \text{ltsummax} + 60.073 * \text{sum05}_{115} + 134.6 * \text{llength} - 1127.92 * \text{lnao5} + 1009.05 * \text{lnao8} \quad (9)$$

Period 1996–2010 ($R^2 = 0.9195$)

$$mspi = -59402 + 2153.36 * \text{ltsummax} + 51.56498 * \text{sum05}_{115} + 147.85 * \text{llength} - 1207.82 * \text{lnao5} + 896.33 * \text{lnao8} \quad (10)$$

Period 1996–2011 ($R^2 = 0.9181$)

$$mspi = -60129 + 2180.25 * \text{ltsummax} + 51.0377 * \text{sum05}_{115} + 150.61 * \text{llength} - 1229.07 * \text{lnao5} + 851.93 * \text{lnao8} \quad (11)$$

3.4 Predictive model for the highest seasonal peak value (mpeak)

The link between severity of the pollen season and environmental variables as observed for model SPI or the peak value here is well known in the literature (Stanley and Linkens 1974; Faegri and Iversen 1989; Andersen, 1991). During the pollen season, different daily airborne pollen peak values may occur at different dates. Here, we build a predictive model for the strongest daily peak value of the season in Montreal using the same procedure as before. Note that the peak does not necessarily corresponds to the first chronological peak in the season but rather to the absolute maximum peak of the season. For mpeak, the preselected predictors are given in Table 7. The final selection of predictors obtained from the multiple regression procedure (period 1996–2011) among a large possible number of model combination are, respectively, according to Table 11D (Appendix 3), ENSO/MEI index of March of the current year (enso3, partial $R^2 = 0.3162$), ENSO/MEI index of January of the previous year (lenso1, partial $R^2 = 0.1615$), the pollen seasonal index of the previous year (lspi, partial $R^2 = 0.1727$), NAO index of May of the previous year (lnao5, partial $R^2 = 0.2225$) and, finally, the NAO index of April of the previous year (lnao4, partial $R^2 = 0.0545$). The results suggest that: (1) the strongest predictors are linked with ENSO/NAO global indices, (2) the previous year environmental variables play again a strong role. The link of ENSO/NAO indices seems unequivocal in the case of the peak value. Current and previous year ENSO/NAO combine with the variable lspi (seasonal pollen index of the previous year) to bring the variance explained of the model to R^2 beyond 0.90 for all periods (see Eqs. 12–14; Table 11D and Fig. 5). The curve fittings

Table 7 Preselected predictors with significant Pearson's coefficient of correlation (i.e. $p \leq 0.15$) for peak1 (first peak of the season)

Predictor	Coefficient of correlation	p value
Lenso1	0.71232	0.0043
Lenso8	0.69417	0.0059
Lisp	−0.65690	0.0073
Lnao5	−0.64364	0.0087
Lenso9	0.62347	0.0172
Precipw	**0.57085**	**0.0209**
Lnao4	0.57003	0.0335
Lenso6	0.56929	0.0336
Enso3	**0.54229**	**0.0451**
Enso1	**0.53806**	**0.0472**
Lensoa	0.53346	0.0333
Lenso10	0.52845	0.0521
Lnao12	−0.49884	0.0697
Lenso12	0.48289	0.0803
Lenso11	0.47012	0.0898
Rainj	**0.45526**	**0.0764**
Lnao9	−0.42180	0.1330
Ltsummax	0.41222	0.1126

The description of the variables used is given in Appendix 1. Current year predictors are indicated by bold character and last year predictor by a letter starting by L (for lag)

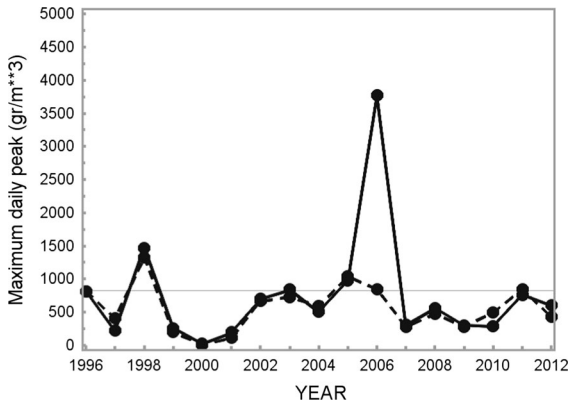

Fig. 5 Statistical model for mpeak (*dotted lines*) versus observed yearly first highest peak values (*solid lines*) for the period 1996–2011. The year 2006 is considered as an outlier here

for different periods are presented below for this parameter:

Period 1996–2009 ($R^2 = 0.982$)

$$\text{mpeak} = 972.723 + 235.49 * \text{lenso1} + 348.40 * \text{enso3} - 0.09707 * \text{lisp} - 229.8 * \text{lnao5} + 179.97 * \text{lnao4} - 174.78 * \text{lnao7} \quad (12)$$

Period 1996–2010 ($R^2 = 0.9312$)

$$\text{mpeak} = 999.53 + 230.18 * \text{lenso1} + 257.86 * \text{enso3} - 0.11364 * \text{lisp} - 277.7 * \text{lnao5} + 121.86 * \text{lnao4} \quad (13)$$

Period 1996–2011 ($R^2 = 0.9274$)

$$\text{mpeak} = 970.34 + 220.18 * \text{lenso1} + 263.0 * \text{enso3} - 0.10713 * \text{lisp} - 261.3 * \text{lnao5} + 118.38 * \text{lnao4} \quad (14)$$

Note that the year 2006 is now considered as an outlier for the computation of mpeak model (observed values of peak greater than 2.5 standard deviations, see Table 3 for the value of standard deviation and abnormal Cook's distance). That year is therefore not considered in the calculation whereas the year 1998 (not an outlier in this case) is very well reproduced by the curve fittings (Fig. 5). Note that for all statistical models (i.e. Eqs. 3–14), the normality of residus and homoscedasticity have both been checked and found reasonable.

The model final selections contain 4 to 9 predictors depending on the given phenological parameter and the given training period. As a summary of the results, Table 8 shows the explained variance of different types of predictors obtained for each predictand. As an average, local predictors (LOCAL (CY) and LOCAL (PY)) overall together explain about 50% of the variance while NAO/ENSO global indices about 44% of the variance of phenology predictands (see last line of the table). Of that amount, local predictors of the current year (CY) accounts as an average for about 10% of the variance while local predictors of the previous year (PY) about 40%. ENSO of the current year contributes as an average to nearly 8% of the total variance explained and finally, NAO/ENSO of the previous year for about 36% (i.e. 19.6% for ENSO(PY) and 16.6% for NAO(PY)). According to

Table 8 Summary of results for the explained variance obtained by the statistical models for the four predictands (mstart : day of the start of the season, mlength : duration of the season, SPI : seasonal pollen index and peak: daily maximum of the season) in terms of the predictor type

Predictand	Local (CY)	Local (PY)	NAO (PY)	ENSO (CY)	ENSO (PY)	Total variance (0–1)
Mstart	0.1846	0.6478	0.0434	0	0.064	0.9397
Mlength	0.0189	0.2131	0.1890	0	0.5576	0.9787
SPI	0.1851	0.580	0.1530	0	0.0	0.9181
Peak	0.0	0.1727	0.277	0.3162	0.1615	0.9274
Average	0.097	0.40365	0.1656	0.0791	0.1958	0.9412

CY indicates current year and *PY* previous year. The total variance is compiled from Appendix 3 for each predicand. The last line (average) is a composite mean of the four predictands

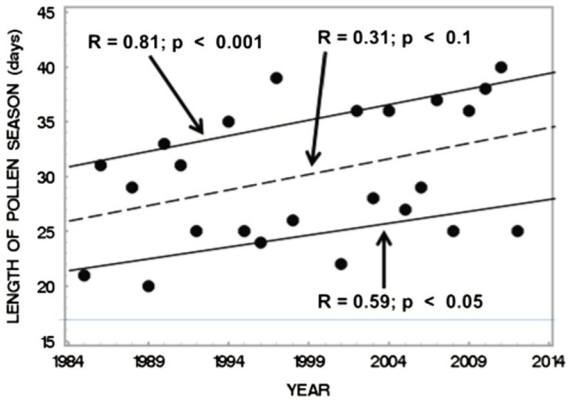

Fig. 6 Temporal trend for the length of birch (*Betula*) pollen season against year for two cases: long duration seasons only, $R = 0.81$ (*top curve*) and short duration seasons only, $R = 0.59$ (*bottom curve*). The *middle curve* represents the linear regression curve for all data lumped together (with no stratification between short and long season). Note that outliers were removed based on the Cook's distance method (see Cook 1977; Weisberg 2013)

the results found, it is then suggested that statistical modeling of phenology should consider ENSO/MEI indices of the previous year and not only heat sum or conditions of temperature/precipitation of the current year as done in many studies in the literature (see Introduction). More specifically, predictors based on NAO/ENSO of the current and previous year increase the information content about spatio-temporal variability. These results improve understanding of spring phenology and suggest more complex interactions among environmental factors. There is no doubt that local variables contain a great deal of information as well (lrainf, ltsummax, tmarch, etc.) but it is possible that this information be redundant with that provided by NOA/ENSO indices (i.e. NAO and ENSO together would then control the phenology to a degree more important than shown in this study).

3.5 Long-term trend of observed phenological parameters

Temporal trends of the four phenological parameters studied above have also been evaluated from all available observed time series (1996–2012). For the length of the birch pollen season, a significant trend towards longer season duration has been found (Fig. 6). The average trend for the length of the season (dashed middle curve) has a slope of 0.26 (i.e. increase of 0.26 day per year for the season duration). The coefficient of correlation found was 0.31 (p value <0.10). This value for the trend (0.26 day per decade) of the season's length is fairly consistent with other values found in the literature (see Discussion Section for

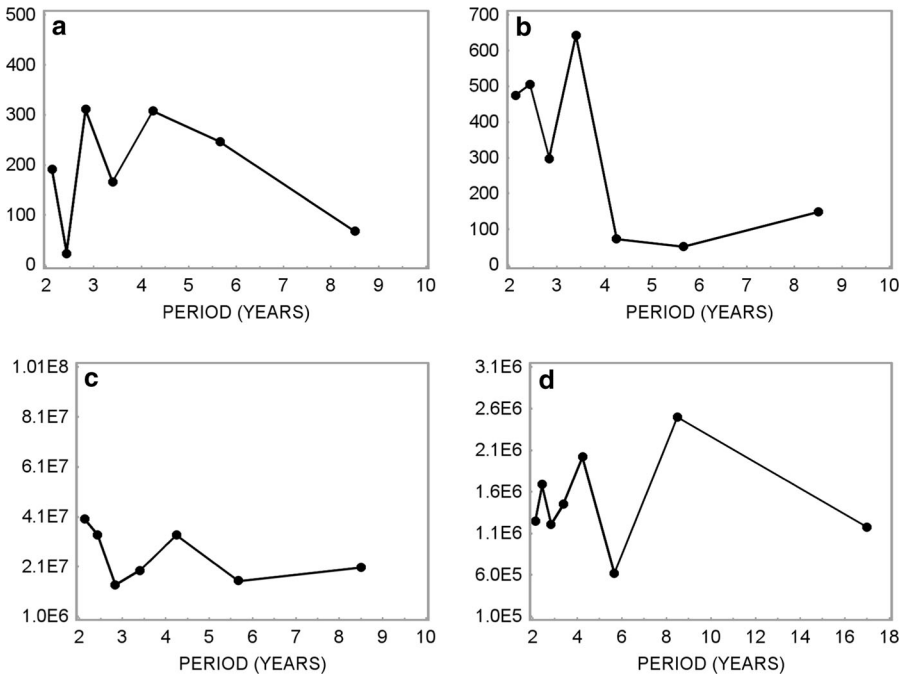

Fig. 7 Periodograms for observed **a** start date of the pollen season, **b** length of the season, **c** seasonal pollen index and **d** yearly peak value

more details). Note that it is interesting to divide the data into two groups: long duration seasons (upper curve in Fig. 6; $R = 0.81, p < 0.001$) or short duration seasons (bottom curve; $R = 0.59, p < 0.05$). Partitioning into these two regimes showed more robust and more significant correlations. In any cases, increasing trends for the duration of the season is expected since it is in line with climate change which translates into an increasing health risk of atopic patients in the future. Note that a higher temperature (as part of climate change scenarios, IPCC 2007a,b) is consistent with an increase in pollen season duration as shown in Eqs. (6)–(8) above since predictors based on temperature (i.e. ltfev, tmarch) have positive correlation with the duration of the season (i.e. mlength). No significant trends were observed for other phenological parameters (i.e. the seasonal pollen index or peak pollen concentration nor for the date of start of the pollination).

3.6 Periodogram of phenological parameters and link with an extreme weather event

For the phenological parameters, cycles between 2 and 4 years (first peaks to the left on Fig. 7A through D) are noticeable. However, this is not exactly consistent with values found in the literature for birch natural cycle which is known to be quasi-biennial (e.g. Emberlin et al. 1993; Laaidi et al. 1997; Latalowa et al. 2002). We believe that the discrepancy with the two-year cycle could be explained by perturbation due to abnormal or extreme weather events. For example, it is likely that the strong *El Niño* event of 1997–1998 and the following Great Ice Storm of January 1998 affecting Montreal, the Eastern Townships and the New England states in USA (see Abley 1998) have caused strong perturbations of the tree cycle and therefore abnormal behaviour for several years after the event possibly disrupting the quasi-biennial cycle (high pollen year alternating with a low pollen year). Strong anomalies of the phenological parameters were indeed observed from the aeropalynological data recorded in Montreal in the following few years after the storm and are obvious on Figs. 2, 3, 4 and 5. For example, the year 1999 (year following the ice storm in Southern Quebec) has the shortest length of the pollen season (duration of about 10 days only, see Fig. 3) whereas the following year (2000) has the longest duration (about 67 days). The transition years (1999–2000) were accompanied with a noticeable amplification of the cycle in the years following the ice storm (which happened in January 1998). Note also that the 1998 spring (few months after the ice storm) had a very strong anomaly for the seasonal pollen index (Fig. 4). In fact, a large variability of all phenological parameters appeared during the period 1998–2000. Therefore, we suggest that the ice storm which affected Southern Quebec and New England in January 1998 (and associated with the strong *El Niño* event 1997–1998) had a deep impact on the phenology producing possible perturbations in the birch cycles for many subsequent years. A study of tree survival after the ice storm revealed high mortality in the following 3 years (Shortle and Smith 2003) in some major source regions affecting Montreal (New England in USA and Eastern Townships as well as the southern part of the Laurentides in Southern Quebec, see geographical location in Fig. 1). According to Andersen (1991) damage to plants due to abnormal weather events affects both the amount of pollen dispersal and the pollen germination and likely, as seen here, the quasi-biennial cycle. Hormonal regulation mechanisms can explain biennial tendency (high and low pollination in succession) but abnormal weather can impact these cycles as well (Andersen 1991). In the Northeast US, a decline of birch tree has been noted and the ice storm which also affected New England (USA) may have served as a catalyst (Halman et al. 2009) for perturbing these natural cycles.

3.7 Validation of statistical models

We distinguish and present here the results of two types of validation for the statistical models: internal and external. Internal validation (same observations used for verification as that used for the curve fittings) is necessary to ensure consistency and to verify absence of gross errors. External validation or cross-validation (see description on Sect. 2.7) is also required to estimate the true predictive power of the different statistical models (Eqs. 3–14).

3.7.1 Internal validation

Mean residuals (that is the mean bias: i.e. OmP) and standard deviation of the residuals (i.e. SD of OmP thereafter) turn out to be small as can be seen from inspection of Figs. 2, 3, 4 and 5 for the phenological parameters (start of the season, length of the season,

Table 9 Results of internal validation for (A) statistical model, (B) aeropalynological average (i.e. LTAA or calendar's method) (C) persistence (1996–2009, $N = 14$) (NA: non-applicable, NS: non statistically significant, i.e. $p > 0.15$)

Predictand (1996–2009)	R^2 (model vs obs)	Mean OmP (bias)	SD of OmP
A			
Start of season (days)	0.9079 ($p < 0.15$)	−0.00325 (<0.003%)	1.26 (~1%)
Length of the season	0.9413 ($p < 0.05$)	1.954 (~7%)	7.26 (~25%)
SPI (total nb. grains)	0.9361 ($p < 0.1$)	526.66 (~13%)	2083.5 (50%)
Peak value (nb. grains)	0.9820 ($p < 0.05$)	0.013546 (<0.003%)	52.74 (<10%)
Predictand (1996–2009)	R^2 (clim vs obs)	Mean OmP (bias)	SD of OmP
B			
Start of season (days)	N/A	−2.392 (<0.02%)	4.151 (~3%)
Length of the season	N/A	2.714 (~9%)	12.542 (~40%)
SPI (total nb. grains)	N/A	0.857 (~0.02%)	3405.8 (~88%)
Peak value (nb. grains)	N/A	0.615 (~0.1%)	400.9 (~75%)
Predictand (1996–2009)	R^2 (persis vs obs)	Mean OmP (bias)	SD of OmP
C			
Start of season (days)	N/S ($p > 0.15$)	−0.6364 (~0.5%)	5.853 (~5%)
Length of the season	0.303 ($p < 0.05$)	0.7857 (~3%)	22.14 (~75%)
SPI (total nb. grains)	0.135 ($p \sim 0.20$)	251.385 (~6.5%)	5803.8 (~150%)
Peak value (nb. grains)	N/S ($p > 0.15$)	14.636 (~3%)	648.77 (~120%)

Note that the % values are obtained with respect to the long-term mean values (the latter are computed using Table 3). OmP means Observed minus Predicted and SD stands for standard deviation

seasonal pollen index and the peak value, respectively). For example, for the curve fitting error of the day of the start of the season, the arithmetic mean of the residuals (mean OmP) and the standard deviation of residuals are very small (about 1% or less, see first line of Table 9A) and in any cases much more accurate than the calendar's method (i.e. LTAA: about 3% or less, see first line of Table 9B) or persistence-based prediction (about 5% ore less, see first line of Table 9C). For the other phenological parameters (length of the season, seasonal pollen index and peak value), mean and standard deviation of residuals are also relatively small and inferior to the values when compared to LTAA (Table 9B) or persistence (Table 9C).

3.7.2 *External validation (cross-validation)*

The predictive model for the start of the season is excellent with a mean arithmetic bias of 0.5 day, and a total error of about 4 days (Table 10A). This is much lower than LTAA or persistence-based model. Note that the total error is computed by taking the square root of the sum of squares of mean and standard deviation of residuals (OmP). The result of Table 10A is considered remarkable when compared with results obtained elsewhere (see Discussion Section for more details). Note that in Figs. 2, 3, 4 and 5, the last year plotted (2010, 2011 and 2012, respectively) is the one which was obtained in the true forecast mode and used for the external validation. Tables 10B–D show the results of the cross-validation for other phenological parameters. Results show that only for the predictand "start of the pollen season", the statistical model has a superior true predictive power compared to LTAA or persistence. For other phenological parameters, it seems that rather a combined average value of LTAA and persistence (persis. & LTAA)-based forecasts give the best results (Tables 10B–D). We suggest that the latter predictands have great inter-annual variability (as can be seen in Table 3) and therefore are very challenging to forecast even for the upcoming season as far as birch pollen is concerned in our study area.

Table 10 Results of cross-validation (period 2010–2012, $N = 3$) for prediction based on the statistical model, LTAA, persistence and a combination of LTAA and persistence for the following predictand: (A) start of the season (mstart), (B) pollen season duration (mlength), (C) seasonal pollen index (spi) and (D) seasonal peak value (mpeak)

Period (2010–2012) start of the season	Mean OmP (bias)	SD of OmP	Total error
A			
Statistical model	0.5008	3.9680	3.9995
LTAA	−6.0667	5.6862	8.3149
Persistence	−0.6667	10.693	10.713
Persis & LTAA	−3.3667	8.0052	8.6795
Period (2010–2012) length of the season	Mean OmP (bias)	SD of OmP	Total error
B			
Statistical model	−2.600	12.70	12.960
LTAA	4.787	8.503	9.7579
Persistence	−3.667	9.815	10.478
Persis & LTAA	0.560	9.152	9.1690
Period (2010–2012) seasonal pollen index	Mean OmP (bias)	SD of OmP	Total error
C			
Statistical model	53.33	1702.1	1702.9
LTAA	−1176.9	850.03	1451.7
Persistence	9.20	1469.9	1469.9
Persis & LTAA	−583.9	1121.5	1264.0
Period (2010–2012) daily max peak value	Mean OmP (bias)	SD of OmP	Total error
D			
Statistical model	−172.94	490.12	519.74
LTAA	−270.3	240.82	362.0
Persistence	98.667	333.37	347.7
Persis & LTAA	−85.83	257.71	271.6

OmP means Observed minus Predicted and SD stands for standard deviation

4 Discussion

4.1 Importance of ENSO/NAO in phenology

Predictors based on NAO/ENSO often appear in the multiple regression models (i.e. Equations 3–14), so it deserves more attention here. Table 8 suggests that NAO/ENSO together could explain 44% of the total variance (composite average for all predictands). On the other hand, traditional predictors for the start of the season relies on thermal parameters of the environment such as growing degree-days few weeks to few months before bud development (see review in the Introduction section). The fact that ENSO and NAO indices of the current and previous year come out as significant non-collinear predictors in Eqs. 3 through 14 suggests that: (1) phenology is also significantly linked to the variability of global weather patterns not only local predictors, (2) trees keep memory of the previous year global weather pattern dynamics since both NAO and ENSO of the current year or the previous year have significant correlation with phenology (Tables 4, 5, 6, 7 and 8). Moreover, global indices improve substantially the curve fittings as compared to a situation where those were not used (in such case the coefficient of determination drops substantially and the total error becomes much larger: results not shown).

We believe that several variables selected here as predictors have a causal relationship with the phenology rather than just a statistical coincidence. For example, it has been observed that in eastern Canada, an *El Niño* event produces above normal temperatures during the following winter months (Shabbar and Khandekar 1996) and this is known to influence phenology of the following spring. ENSO control is revealed in our study but also NAO.

4.2 Importance of previous year conditions

The importance of past year environmental conditions has been clearly demonstrated in this study and recognized as well in the literature. Fairley and Batchelder (1986) observed a strong correlation between the oak pollen season duration and the rainfall during the previous year in Northern California (USA). More recently, these links between tree pollen season duration and environmental variables during the previous year has been noted by some researchers across European countries for birch (Latalowa et al. 2002 for Poland; Méndez et al. 2005 for Spain; Ranta et al. 2011 in Scandinavia). Laaidi et al. (1997) noted that despite the fact that trees keep memory of current and previous year environmental changes, traditional models of phenology found in the literature do not sufficiently account for it. This is still true nowadays and this is another major point that our study has addressed. Table 8 clearly reveals that most of the explained variability by the statistical models presented is linked with predictors of the previous year (indicated by PY). Such predictors explain up to 77% of the total variance (as a composite average for all predictands).

For the start of the birch pollen season, the fact that the predictors lrainf (rainfall accumulated during the fall season of the previous year) and ltsummax (summer average maximum temperature of the previous year) are the best two predictors may, at first glance, seem counter-intuitive. However, several independent observations point towards the validity and the great importance of these predictors and their link with phenology. Jones (1995) found that the blooming date is dependent on the amount of rainfall from several months prior to the start of the birch pollen season in UK. Adams-Groom et al. (2002) also noticed that rainfall prior to the birch pollen season turns out to be a good predictor in UK. Moreover, recently, Xie et al. (2015) have established that wet conditions in the fall (among other environmental stress) is a factor inducing earlier dormancy in deciduous forest of New England (one of the source region of Montreal's airborne birch pollen, see Fig. 1). Other studies (mentioned in Xie et al.'s paper) reported a positive correlation between spring and fall phenology. Rozas and García-González (2012) suggested that *El Niño-Southern Oscillation* (ENSO) is controlling the regional hydrological regime in Northwest Spain implying a link between ENSO, water availability of the previous year and tree growth of the current year. Moreover, according to Antépara et al. (1995), water availability plays a role and the amount of rainfall prior to the pollen season could be an important predictor since tree roots can penetrate the phreatic mantle. The latter would be consistent with a strong negative correlation between lrainf (autumn rainfall of the previous year) and the start date of the pollen season as found in our study. Similarly, Xie et al. (2015) also reported that a heat-stress (directly related to the predictor ltsummax in our study) has also an impact on the next fall dormancy. All these independent observations seem to support our results that is the ENSO dynamics and local predictors of the previous months and even through the previous year which act together and have a control over the start date of the upcoming pollen season. In fact, Table 4 shows that various form of ENSO/MEI (ensowin, lenso12, lensof) have coefficient of correlation significant (R greater than 0.5, p value <0.05) with d2start (Julian date for the start of the season). The third best non-collinear predictor for the starting date is tmarch (mean temperature of March of the current year; see Appendix 3, Table 11A). This finding for our study area is in agreement with Stach et al. (2008) for the case of birch (*Betula*) in Europe. Similarly, according to D'Odorico et al. (2002), late-winter temperature has a strong influence on spring phenology at the mid and high latitudes so that higher spring temperatures are associated with earlier occurrences of the phenodates. Finally, Spieksma et al. (1995) found that air temperature 2 months before the beginning of the birch season was strongly related with the start date of pollination. This result has been successfully reproduced in our study. In the case of the duration of the pollen season for birch, all the non-collinear predictors (except tmarch) are linked with the previous year (Appendix 3, Table 11B). Similarly, for the seasonal pollen index (SPI), and peak (Appendix 3, Table 11C and D, respectively), most of the predictors belong to the previous year. The strong relation found between the

environmental variables of the previous year and SPI has been noted elsewhere in several studies (mostly in Europe, see Ranta and Satri 2007; Stach et al. 2008) and then give support again to our study. Note that predictors used for the previous year correspond to the idea that the memory of birch trees is roughly 2 years (quasi-biennial cycle, as discussed above). In ecology, a lagged correlation between NAO/ENSO and the biological response has been frequently observed (Ottersen et al. 2001) but rarely included in phenology models. Finally, we believe that it is important to consider a large amount of predictors in order to identify confounders[4] and non-traditional predictors and not simply use classical simple predictors (temperature, heat sum and precipitation few weeks or months before the upcoming season). We believe that previous studies of phenology might have missed useful predictors such as NOA/ENSO and the lag effect of the previous year.

4.3 Numerical comparison of results with other studies in the literature

The current approach in the literature of using phenology predictors only based on thermal units of chilling and warming of the current year or to consider finer spatio-temporal scale to study phenology (down to growing degree-hour resolution for example, see Rea and Eccel 2006) was not adopted in our study. The choice of a statistical multivariate model such as presented here to forecast phenology was made because: (1) simple models such as the sum of heat (popular in Europe) did not correlate well with the start of the pollen season in Montreal for many predictands including the start date of the season, (2) the level of noise in the observational data used to construct phenological models is usually high (Siljamo et al. 2008a, b). On the other hand, introducing too much complexity in phenological models (i.e. process-based and computational intelligence models.) has not shown to bring any gain in Europe (see review in Sofiev and Bergmann 2013). We believe that multivariate statistical model such as presented here is an acceptable alternative to other kinds of approach for *Betula* phenology and we have found that the performance of the forecast for the start of the pollen season seems superior to other statistical method shown in the literature for similar problems.

In fact, the methodology shown in this paper was proven efficient with R^2 superior to 0.90 for all models of phenological parameters. Such high coefficient of determination has rarely been obtained in phenological studies. For example, in a study published by Adams-Groom et al. (2002) (AG02 thereafter), the prediction of the start of the birch pollen season in the UK was attempted and R^2 (coefficient of determination) obtained was about 84% (average of 3 sites in UK) for the explained variance whereas in our study, R^2 is superior to 90% for all cases (see Eqs. 3 to 5), i.e. our actual and predicted time series are closer to observations than that in AG02. Moreover, in AG02, the mean number of day difference between predicted and actual start dates for their study period was 5 days for London, 3 days for Cardiff and 1.5 day for Derby, respectively. For the test years, the mean difference was 1, 4.5 and 7.5 days for the 3 sites in UK in AG02 (average of 4.33 days when all sites combined) whereas our study shows only 0.5 day difference as an average (i.e. mean OmP) using independent data with a total error of about 4 days. Our model is also superior to the multiple regression model developed by Laaidi (2001) to predict the start of the birch pollination. Galán et al. (2005) (cited in Sofiev and Bergmann 2013, Chap. 4) obtained a mean absolute bias of 4.8 days (with an RMSE in the range 6.2–7.8 days) for the start date for olive in Spain whereas White et al. (1997) obtained mean absolute errors in the range 5.3–7.1 days for deciduous forests and grasslands in the temperate zone. Our predictive model for the starting date of birch shows a random error of about 4 days, a mean bias of 0.5 day and a mean absolute bias of about 3 days. These figures are largely inferior to errors found in the studies described above. Note that an accuracy of 4 days for the random error is an excellent result since it was established that irreducible uncertainties in the timing of the season are of the order of the meteorological turnover time which is about 3 days (Siljamo et al. 2008a,b). However, for the remaining phenological parameters (length of the season, SPI and peak values) the statistical models developed here show more error (when compared to independent data) than that obtained with long-term average aeropalynological data (calendar's method) or

[4] A confounding variable has the property of correlating (positively or negatively) with both the dependent and independent variable (see Montgomery 2005 for further details).

persistence-based forecast. This is likely due to a very high variability of phenology and the lack of normal (Gaussian) behaviour (as it can be seen in Table 3) for these phenological parameters. In any cases, more independent data are needed in the future to confirm these results.

As far as trend is concerning, in our study, we noted a significant increase in the pollen season duration of about 0.26 day per year (2.6 days per decade; see Fig. 6) for birch as seen above. In Europe, Frey and Gassner (2008) found an increasing trend of 0.4 day per year for the length of the season also for birch. According to WHO (2003), during the last 30 years, the length of the growing season has increased about 10–11 days for most species (i.e. about 0.3 day per year). We speculate that different trends in different areas of the world could be explained by spatially non-homogeneous upward temperature trends related to global warming and plant distribution. Note that these results contrast with Bapikee (2005) who found no significant trend in any phenological parameters for birch (*Betula* spp.) for the period 1985–1998 using data collected in Montreal. This could be explained by the fact that the more recent period of data used here (1996–2012) shows more evidence of climate variability than the period prior to the study period.

5 Summary and conclusions

This study is the first analysis of pollen time series and phenology forecasting for birch in Canada. The intent of this paper was:

1. to study the time series of birch pollen in Montreal (Canada) and to build statistical models for predicting phenology for the upcoming season (few weeks in advance) for the following predictands: starting date of the pollen season, length of the season, seasonal pollen index and peak value for birch tree pollen in the Montreal region and surrounding areas,
2. to evaluate the role of global weather indices (NAO, and ENSO/MEI) and the previous year environmental factors as statistical predictors of phenology in our area,
3. to perform a trend and a spectral analysis of the pollen time series.

Since numerical pollen forecast requires phenological data, a precise prediction of the phenological parameters few weeks in advance is of paramount interest. As a matter of fact, the latter information can be used as a guide to start medicating for asthma, allergic rhinitis and other allergenic disorders. More importantly for aerobiologists, the prediction of phenological parameters of the upcoming season is of relevance as input to numerical models of pollen dispersal (e.g. Helbig et al. 2004; Sofiev et al. 2006a, b). Many uncertainties affect the phenology and a great deal of inter-annual variability could explain the lack of predictive power of some phenological parameters. Moreover, the fact that some predictands show non-Gaussian behaviour (such as SPI and peak, figures not shown) could be a drawback for the use of linear statistical model to predict for the upcoming season. Nevertheless, the role of global weather indices and of the previous year has been shown to be important for all predictands. It is of interest from a scientific point of view to notice the fact that NAO and ENSO have both significant correlations with phenological parameters for birch in eastern North America. These weather teleconnections indices present the advantages that they provide robust correlation with phenology and they could be easily and freely obtained from NOAA websites. We suggest that NAO/ENSO global indices both of the current and previous year should be included into phenology models in the future since these global indices have a much higher spatial representativeness than local variables and seem to explain a significant amount of the variability. Moreover, global indices will have a growing importance in the future in relation with climate change and its impact on ecology (Ottersen et al., 2001). Finally, damage to trees due to extreme weather should also be considered since they obviously perturb phenology as suggested in this paper. We believe that this study is innovative since it is the first time as far as we know that phenology is investigated with a statistical analysis encompassing a large basin of potential predictors (over 60) including global indices such as NAO and ENSO (for the current and the previous year) in Canada or elsewhere. It is likely that the implications and the methodology adopted in this paper are quite general and could be applicable to other geographical areas sensitive to NAO/ENSO variations and to other plant species. The conclusions of the study can be summarized here:

1. among the best statistical predictors for phenology in our study area (Montreal, Canada) are NAO and ENSO together explaining, as an overall average, 44% of the variance, whereas local environmental predictors could explain up to 50% in average (see last line of Table 8).
2. the previous year predictors were found to explain an overwhelming amount of variability (77%) as compared to predictors of the current year (18%).
3. independent validation shows that the forecasting for the upcoming season was found successful for the start of the day (mean bias of 0.5 day and total error less than 4 days) while for other predictand, the inter-annual variability was found too large for accurate prediction enough to beat the forecast based on persistence or calendar's method.
4. pollen time series observed in Montreal show a trend towards longer birch pollen season (increase of 2.6 days per decade) in agreement with similar European studies.

A link is suggested between damage caused by an extreme weather event (such as the Great Ice Storm of 1998 in Eastern Canada) and the impact on the airborne pollen inter-annual variability during the following years. Future work will combine satellite data with regional scale statistical models as developed in this study to eventually represent better phenology on a larger scale.

Compliance with ethical standards

Conflict of interest None.

Appendix 1: List of potential predictors for phenological models

Predictors of the current year apply for the period January 1st to March 31st just prior to the upcoming pollen season. Predictors for the previous year start with the letter l (for lag) and apply for the previous calendar year.

Current year predictors:

tjan average January temperature
tfev average February temperature
tmarch average March temperature
twinmin mean winter minimum temperature (Jan-March)
twinavg mean winter temperature (Jan-March)
sum05_115 sum of temperature above 5 °C since January 1st cumulated at Julian day 115
sum03_115 sum of temperature above 3 °C since January 1st cumulated at Julian day 115
sum05_110 sum of temperature above 5 °C since January 1st cumulated at Julian day 110
sum03_110 sum of temperature above 3 °C since January 1st cumulated at Julian day 110
rainj total January rain amount
rainf total February rain amount
rainm total March rain amount
precipw total rain precipitation for the period January–March
nao1 NAO index for January
nao2 NAO index for February
nao3 NAO index for March
enso1 bi-monthly MEI index for January
enso2 bi-monthly MEI index for February
enso3 bi-monthly MEI index for March
ensowin mean bi-monthly MEI index for the period January–March

Previous year predictors:

ltjan last year average January temperature
ltfev: last year average February temperature
ltmar last year average March temperature
ltsum last year average summer temperature (June–July–August)
ltsummax last year average summer maximum temperature (June–July–August)
lrainf last year total amount of rain during fall (October–December)
lnao1 last year NAO index for January
lnao2 last year NAO index for February
lnao3 last year NAO index for March
lnao4 last year NAO index for April
lnao5 last year NAO index for May
lnao6 last year NAO index for June
lnao7 last year NAO index for July
lnao8 last year NAO index for August
lnao9 last year NAO index for September
lnao10 last year NAO index for October
lnao11 last year NAO index for November
lnao12 last year NAO index for December

lnaospr mean values of lnao4, lnao5 and lnao6 (spring months of the previous year)
lnaosum mean values of lnao7, lnao8 and lnao9 (summer months of the previous year)
lnaofal mean values of lnao10, lnao11 and lnao12 (fall months of the previous year)
lnaowin mean values of lnao1, lnao2, lnao3 (winter months of the previous year)
lensow same as ensowin but for the previous year
lenso1 last year bi-monthly ENSO/MEI index for January
lenso2 last year bi-monthly ENSO/MEI index for February
lenso3 last year bi-monthly ENSO/MEI index for March
lenso4 last year bi-monthly ENSO/MEI index for April
lenso5 last year bi-monthly ENSO/MEI index for May
lenso6 last year bi-monthly ENSO/MEI index for June
lenso7 last year bi-monthly ENSO/MEI index for July
lenso8 last year bi-monthly ENSO/MEI index for August
lenso9 last year bi-monthly ENSO/MEI index for September
lenso10 last year bi-monthly ENSO/MEI index for October
lenso11 last year bi-monthly ENSO/MEI index for November
lenso12 last year bi-monthly ENSO/MEI index for December
lensoa last year annual average of bi-monthly ENSO/MEI
lensospr last year spring average of bi-monthly ENSO/MEI
lensosum last year summer average of bi-monthly ENSO/MEI
lensof last year fall average of bi-monthly ENSO/MEI
ensowin winter average of bi-monthly ENSO/MEI of the current year
llength last year length of the season according to the D2 method
lendD2 last year Julian date for the end of birch pollen season based on the D2 method
lspi last year birch seasonal pollen index

Appendix 2: Steps followed for multiple regression

1. List of predictors established according to a literature survey (see a list in Appendix 1 above).
2. Scatter plots to evaluate linearity, correlation, outliers, etc.
3. Use of the procedure REG sequentially (options selection C_p, *Rsq* in SAS®) to examine and select the most appropriate model among $2^p - 1$ model (chunks of 7 predictors are sequentially input giving 2^7–1 = 128 combinations of models for each chunk).
4. Step 3 is applied to the whole list of preselected predictors and model building is performed using forward, backward and stepwise procedure. The results are checked against another similar procedure (proc REG with STEPWISE option). Check with *t*-Test, *F*-Test, C_p (Mallows, 1973), R^2 adjusted, etc. for regression coefficients of the model (see similar procedure, Hoang Diem Ngo 2012).
5. Model adequacy test (global *F*-test, optimum adjusted R^2, RMSE).
6. Check for model assumptions (model error are random and all pairs of random errors are independent), residual plots, normal probability plot and checking for outliers (Cook's distance, Cook 1977), Durbin-Watson test for autocorrelation of residuals.
7. Check for multicollinearity of predictors and over-fitting.
8. Repeat steps 3–7 for three periods (1996–2009, 1996–2010 and 1996–2011) and for all predictands.
9. Model validation with independent data (the independent data is the year following a given training period).

Note that the entry value for a given potential predictor is set to $p \leq 0.15$ and the elimination criteria to $p > 0.15$ (standard values in the community of multiple regression modeling and are often the default values in SAS multiple regression procedures, see Beal 2005; Hoang Diem Ngo 2012). As an alternative method, multiple regression using AIC criteria for minimization was also used but found having less accuracy against independent data (see text for details).

Appendix 3: Results of STEPWISE procedure

See Table 11.

Table 11 Final results from computer outputs for stepwise selection summary for the period 1996–2011, (A) start of the birch pollen season (mstart), (B) duration of the season (mlength), (C) seasonal pollen index (mspi) and (D) mpeak

Step	Variable entered	R^2 partial	R^2 of model	C(p)	Pr > F
A. Stepwise procedure summary for mstart					
1	lrainf	0.4041	0.4041	77.0	0.0081
2	ltsummax	0.2437	0.6478	42.6	0.0103
3	tmarch	0.1846	0.8324	17.0	0.0034
4	lnao1	0.0434	0.8758	12.6	0.0759
5	lenso6	0.0405	0.9163	8.5	0.0525
6	lenso1	0.0235	0.9397	7.0	0.0937
B. Stepwise procedure summary for mlength					
1	lenso5	0.5059	0.5059	104.8	0.0029
2	lnao3	0.1545	0.6604	70.6	0.0377
3	lendD2	0.0775	0.7379	54.4	0.0987
4	llength	0.1040	0.8419	32.0	0.0281
5	lensoa	0.0377	0.8796	25.2	0.1275
6	ltfev	0.0316	0.9113	19.8	0.1298
7	lnao8	0.0345	0.9458	13.7	0.0726
8	tmarch	0.0189	0.9647	11.3	0.1233
9	lenso4	0.0140	0.9787	10.0	0.1302
C. Stepwise procedure summary for mspi					
1	ltsummax	0.3881	0.3881	56.227	0.0131
2	llength	0.1919	0.5800	37.146	0.0373
3	sum05_115	0.1851	0.7651	18.808	0.0133
4	lnao5	0.0790	0.8441	12.125	0.0480
5	lnao8	0.0740	0.9181	6.00	0.0191
D. Stepwise procedure summary for mpeak					
1	enso3	0.3162	0.3162	82.18	0.0234
2	lenso1	0.1615	0.4776	61.94	0.0663
3	lspi	0.1727	0.6504	40.15	0.0315
4	lnao5	0.2225	0.8729	11.51	0.0011
5	lnao4	0.0545	0.9274	6.00	0.0208

Note that for other periods (1996–2010 or 1996–2012), the results are similar and not shown (see Appendix 1 for the meaning of the variables). The stopping criteria to avoid over-fitting is when C(p) becomes roughly equal to the number of selected predictors. For each selected predictor, Pr > F (global Fischer test) must be <0.15

References

Abley, M. (1998). *The ice storm: An historic record in photographs of January 1998*. In J. Robinson, McClelland and Stewart Inc., ISBN 0-7710-6100-5.

Adams-Groom, B., Emberlin, J., Corden, J., Millington, W., & Mullins, J. (2002). Predicting the start of the birch pollen season at London Derby and Cardiff, United Kingdom, using a multiple regression model, based on data from 1987 to 1997. *Aerobiologia, 18,* 117–123.

Andersen, T. B. (1991). A model to predict the beginning of the pollen season. *Grana, 30,* 269–275.

Antépara, I., Fernández, J. C., Gamboa, P., Jauregui, I., & Miguel, F. (1995). Pollen allergy in the Bilboa area (European Atlantic seabord climate): pollination forecasting methods. *Clinical and Experimental Allergy, 25,* 133–140.

Avalo, E., Pasqualoni, L., Federico, S., et al. (2008). Correlation between large-scale atmospheric fields and the olive pollen season in Central Italy. *International Journal of Biometeorology, 52,* 787–796.

Bapikee, C. (2005). *Contenu pollinique atmosphérique à Montréal (Québec) et les variations climatiques interannuelles de 1985 à 1988*. Mémoire de maîtrise: Département de géographie de l'Université de Montréal.

Beal, D. J. (2005). SAS Code to Select the Best Multiple Linear Regression Model for Multivariate Data Using Information Criteria. http://analytics.ncsu.edu/sesug/2005/SA01_05.PDF. Last Accessed January 20 2017.

Braun, E. L. (1950). *Deciduous forest of Eastern North America*. New York: The Free Press.

Chuine, I., Cour, P., & Rousseau, D. D. (1998). Fitting models predicting dates of flowering of temperate-zone trees using simulated annealing. *Plant, Cell and Environment, 21,* 455–466.

CLA (Canadian Lung Association) (2015). http://www.lung.ca/home-accueil_e.php. Accessed January 20 2017.

Clot, B. (2001). Airborne birch pollen in Neuchatel (Switzerland): onset, peak and daily patterns. *Aerobiologia, 17,* 25–29.

Comtois, P., & Gagnon, L. (1988). Concentration pollinique et fréquence des symptômes de pollinose: une méthode pour déterminer les seuils cliniques. *Revue Française d'Allergologie, 28*(4), 279.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics (American Statistical Association), 19*(1), 15–18.

D'Odorico, P., Yoo, J. C., & Jager, S. (2002). Changing seasons: An effect of the North Atlantic oscillation? *Journal of Climate, 15,* 435–445.

Dales, R. E., Cakmak, S., Judek, S., & Coates, F. (2008). Tree pollen and hospitalization for asthma in urban Canada. *International Archives of Allergy and Immunology, 146,* 241–247. doi:10.1159/000116360.

Daniel, C., & Wood, F. (1980). *Fitting equations to data* (Revised ed.). New-York: WileyInc.

de Réamur, M. (1735). Observations du thermomètre, faites à Paris pendant l'année 1735, comparées avec celles qui ont été faites sous la ligne, à l'isle de France, à Alger et quelques unes de nos isles de l'Amérique. *Académie des Sciences de Paris*, 545.

Emberlin, J., Detandt, M., & Gehrig, R. (2002). Responses in the start of *Betula* (birch) pollen seasons to recent changes in spring temperatures across Europe. *International Journal of Biometeorology, 46,* 159–170. doi:10.1007/s00484-006-0038-7.

Emberlin, J., Savage, M., & Woodwood, R. (1993). Annual variations in the concentrations of *Betula* pollen in the London area, 1961–1990. *Grana, 32,* 359–363.

Faegri, K., & Iversen, J. (1989). *Textbook of pollen analysis*. Chichester: Wiley.

Fairley, D., & Batchelder, G. L. (1986). A study of oak-pollen production and phenology in northern California: Prediction of annual variation in pollen counts based on geographic and meteorological factors. *Journal of Allergy and Clinical Immunology, 78,* 300–307.

Faust, M. (1989). *Physiology of temperate zone fruit trees* (pp. 169–234). London: Wiley.

Frey, T., & Gassner, E. (2008). Climate change and its impact on birch pollen quantities and the start of the pollen season: an example from Switzerland for the period 1969–2006. *International Journal of Biometeorology, 52,* 667–674. doi:10.1007/s00484-008-0159-2.

Galán, C., García-Mozo, H., Vazquez, L., Ruiz, L., Diaz de la Guardia, C., & Trigo, M. M. (2005). Heat requirement for the onset of the *Olea europaea* L. pollen season in several sites in Andalusia and the effect of the expected future climate change. *International Journal of Biometeorology, 49,* 184–188.

García-Mozo, H., Chine, I., & Aira, M. J. (2008). Regional phenological models for forecasting the start and peak of the Quercus pollen season in Spain. *Agricultural and Forest Meteorology, 148,* 372–380.

Goldberg, C., Bugh, H., Moseholm, L., & Weeke, E. R. (1988). Airborne pollen record in Denmark, 1977–86. *Grana, 27,* 209–217.

Guérin, B. (1993). *Pollen et allergies*. Éd. Allerbio, Varennes-en-Argonne, 279 pages.

Halman, J. M., Schaberg, P. G., Hawley, G. J., Hansen, G. J., & Hansen, C.F. (2009). *Potential role of Ca in the response of paper birch (Betula papyrifera) to ice storm induced decline in Vermont*, USA. In *94th ESA annual meeting, Albuquerque*, New Mexico, August 2–7.

Helbig, N., Vogel, B., Vogel, H., & Fiedler, F. (2004). Numerical modelling of pollen dispersion on the regional scale. *Aerobiologia, 20,* 3–19.

Hirst, J. M. (1952). An automatic volumetric spore trap. *The Annals of Applied Biology, 39*(2), 257–265.

Hoang Diem Ngo, T. (2012). *The steps to follow in a multiple regression analysis*. Paper 333-2012. SAS Global Forum 2012. April 22-25 Orlando, Florida. Available at: support.sas.com/resources/papers/proceedings12/333-2012.pdf.

Hurrell, J. W., & van Loon, H. (1997). Decadal variations in climate associated with the North Atlantic oscillation. *Climate Change, 36,* 301–332.

Institut national de santé publique du Québec (INSPQ) (2013). *États des connaissances sur le pollen et les allergies*. *Les assises pour une gestion efficace*, ISBN 978-2-550-68403-9 (available at http://www.inspq.qc.ca).

IPCC (Inter-governmental Panel for Climate Change). (2007a). *The physical science basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press.

IPCC (Inter-governmental Panel for Climate Change). (2007b). Climate change 2007: Impacts, adaptation and vulnerability. *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change,* In Solomon, S., D. Qin, M. Manning, & Z. Chen et al. (eds). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp. available at http://www.ipcc.ch.

Jones, S. (1995). *Allergenic pollen concentrations in the United Kingdom*. PhD Thesis. London : University of North London.

Laaidi, M. (2001). Regional variations in the pollen season of *Betula* in Burgundy: Two models for predicting the start of the pollination. *Aerobiologia, 17,* 247–254.

Laaidi, K., Laaidi, M., & Besancenot, J.-P. (1997). Pollens, pollinoses et météorologie. *La météorologie, 20*(8), 41–55.

Laaidi, K., Laaidi, M., & Besancenot, J.-P. (2002). Synergie entre pollens et polluants chimiques de l'air: les risques croisés. *Environnement, Risques et santé, 1*(1), 42–49.

Latalowa, M., Miętus, M., & Uruska, A. (2002). Seasonal variations in the atmospheric *Betula* pollen in Gdańsk (southern Baltic coast) in relation to meteorological parameters. *Aerobiologia, 18,* 33–43. doi:10.1023/A1014905611834.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics, 15*(4), 661–675.

Mandrioli, P., Comtois, P., & Levizzani, V. (1998). *Methods in aerobiology*. Pitagora Editrice Bologna ISBN 8-371-1043-X.

Mark, J., & Goldberg, M. A. (2001). Multiple regression analysis and mass assessment: A Review of the issues. *The Appraisal Journal, 56,* 89–109.

Mayers, J. H., & Forgy, E. W. (1963). The development of numerical credit evaluation systems. *Journal of the American Statistical association, 58*(303), 799–806.

Méndez, J., Comtois, P., & Iglesias, I. (2005). Betula pollen: One of the most important aeroallergens in Ourense, Spain. Aereobiological studies from 1993 to 2000. *Aerobiologia, 21,* 115–123.

Montgomery, D. C. (2005). *Design and analysis of experiments*. 6th edn, Wiley. New York, Section 7.3.

NOAA (2012a). *NAO bi-monthly index*. (https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/NAO/). Accessed 29 January 2017.

NOAA (2012b) *Multivariate ENSO index (MEI)* (https://www.esrl.noaa.gov/psd/enso/mei/index.html). Accessed 29 January 2017.

Ottersen, G., Planque, B., Belgrano, A., Post, E., Reid, P. C., & Stenseth, N. C. (2001). Ecological effects of the North Atlantic oscillation. *Oecologia, 12,* 1–14. doi:10.1007/s004420100655.

Ranta, H., & Satri, P. (2007). Synchronised inter-annual fluctuations of flowering intensity affects the exposure to allergenic tree pollen in North Europe. *Grana, 46*(4), 274–284.

Ranta, H., Siljamo, P., Oksanen, A., Sofiev, M., Linkosalo T., Bergmann, K.-C., Bucher, E., Ekebom, A., Emberlin, J., Gehrig, R., Hallsdottir, M., Jato, V., Jäger, S., Myszkowska, D., Paldy, A., Ramfjord, H., Severova, E., & Thibaudon, M. (2011). *Aerial and annual variation of birch pollen loads and a modelling system for simulating and forecasting pollen emissions and transport at an European scale*. In B. Clot, P. Comtois, & B. Escamilla-Garciá (eds), *Aerobiological monographs. Towards a comprehensive vision* (Vol.1, pp. 115–131). MeteoSwiss (CH) and University of Montreal(CA), Montréal, Canada,. ISBN 978-2-8399-0466-7.

Rasmussen, A. (2002). The effects of climate change on the birch pollen season in Denmark. *Aerobiologia, 18,* 253–265.

Rea, R., & Eccel, E. (2006). Phenological models for blooming of apple in a mountainous region. *International Journal of Biometeorology, 51,* 1–16.

Rousi, M., & Heinonen, J. (2007). Temperature sum accumulation effects on within-population variation and long-term trends in date of bud burst of European white birch (*Betula pendula*). *Tree Physiology, 27,* 1019–1025.

Rozas, V., & García-González, I. (2012). *Non-stationary influence of El Nino-Southern Oscillation and winter temperature on oak latewood growth in NW Iberian Peninsula*. Int: J. Biometeorol. doi:10.1007/s00484-011-0479-5.

SAS Institute Inc. (1989). *SAS/STAT and SAS/ETS User's Guide, Version 6, vol. 2* (4th ed.). Cary: SAS Institute.

Scheifinger, H., Belmonte, J., Buters, J., Celenk, S., Damialis, A., Dechamp, C., García-Mozo, H., Gehrig, R., Grewling, H., Halley, J. M. et al. (2013). Monitoring, modelling and forecasting of the pollen season. Chap 4. In *Allergenic pollen. A review of the production, release, distribution and health impacts*. Chap. 4. Sofiev, Mikhail, Bergmann, Karl-Christian (Eds.), doi 10.1007/978-94-007-4881-1.

Scheifinger, H., Menzel, A., Koch, E., & Peter, C. (2002). Atmospheric mechanisms governing the spatial and temporal variability of phenological phases in central Europe. *International Journal of Climatology, 22,* 1739–1755.

Shabbar, A., & Khandekar, M. (1996). The impact of El Niño—Southern Oscillation on the temperature field over Canada. *Atmosphere-Ocean, 34,* 401–416.

Shortle, W. C., Smith, K. T., Dudzik, K. R. et al. (2003). *Tree survival and growth following ice storm injury*. USDA Forest Service, Northeastern Research. Research paper NE-723.

Siljamo, P. (2013). *Numerical modelling of birch pollen emissions and dispersion on regional and continental scales*. Finnish Meteorological Institute Contributions No. 99, FMI-CONT-99. Helsinki, Finland. ISBN 978-951-697-795-2.

Siljamo, P., Sofiev, M., Ranta, H., Linkosalo, T., Kubin, E., Ahas, R., et al. (2008a). Representativeness of point-wise phonological *Betula* data collected in different parts of Europe. *Global Ecology and Biogeogeography, 17*(4), 489–502.

Siljamo, P., Sofiev, M., Severova, E., et al. (2008b). Sources, impact and exchange of early-spring birch pollen in the Moscow region and Finland. *Aerobiologia, 24,* 211–230. doi:10.1007/s10453-008-9100-8.

Sofiev, M., Belmonte, J., Gehrig, R., Izquierdo, R., Smith, M., Dahl, A., Siljamo, P. (2013b). Airborne pollen transport. Chap 5. In M. Sofiev & K. C. Bergmann (eds.), *Allergenic pollen. A review of the production, release, distribution and health impacts*. Springer Science + Business Media, Dordrecht, 256 pp. ISBN 978-94-007-4880-4e-ISBN 978-94-94-007-4881-1.

Sofiev, M., & Bergmann, K. C. (Eds.). (2013). *Allergenic pollen. A review of the production, release, distribution and health impacts*. New York : Springer. doi:10.1007/978-94-007-4881-1.

Sofiev, M., Siljamo, P., Ranta, H., Linkosalo, T., Jaeger, S., Rasmussen, A., et al. (2013a). A numerical model of birch pollen emission and dispersion in the atmosphere. Description of the emission module. *International Journal of Biometeorology, 57,* 45–58. doi:10.1007/s00484-012-0532-z.

Sofiev, M., Siljamo, P., Ranta, H., & Rantio-Lehtimäki, A. (2006a). Towards numerical forecasting of long-range air transport of birch pollen: Theoretical considerations and a feasibility study. *International Journal of Biometeorology, 50,* 392–402. doi:10.1007/s00484-006-0027-X.

Sofiev, M., Siljamo, P., Valkama, I., Ilvonen, M., & Kukkonen, J. (2006b). A dispersion modelling system SILAM and its evaluation against ETEX data. *Atmospheric Environment, 40,* 674–685.

Spieksma, F Th M, Emberlin, J. C., Hjelmroos, M., et al. (1995). Atmospheric birch (*Betula*) pollen in Europe: Trends and fluctuations in annual quantities and the starting dates of the seasons. *Grana, 34,* 51–57.

Stach, A., Emberlin, J., Smoth, B., Adams-Groom, B., & Myszkowska, D. (2008). Factors that determine the severity of *Betula* spp. Pollen season in Poland (Poznań and Krakow) and the United Kingdom (Worcester and London). *International Journal of Biometeorology, 52,* 311–321. doi:10.1007/s00484-0127-2.

Stanley, R. G., & Linskens, H. F. (1974). *Pollen, biology, biochemistry, management*. Berlin: Springer.

Tamburlini, G., et al. (2002). Children's health and environment: a review of evidence, Environmental European Agency. Report no 29, 44-57. Available at http://www.euro.who.int/__data/assets/pdf_file/0007/98251/E75518.pdf. Accessed January 29 2017.

Trenberth, K. E., Jones, P. D., Ambenje, P., Bojariu, R., Easterling, D., Klein, T., Parker, D., Rahimzadeh, F., Renwick, J.A., Rusticucci, M., Soden, B., & Zhai, P. (2007). Observations: Surface and atmospheric climate change. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor & H.L. Miller (eds.), *Climate change 2007: The physical science basis. contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change.* Cambridge: Cambridge University Press. pp. 235–336. https://www.ipcc.ch/publications_and_data/ar4/wg1/en/ch3.html. Accessed January 29 2017.

Walther, G. R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., et al. (2002). Ecological responses to recent climate change. *Nature, 416,* 389–395. doi:10.1038/416389a.

Weisberg, S. (2013). *Applied linear regression*, 4th edn. ISBN: 978-1-118-38608-8. Wiley Publishers.

White, M. A., Thornton, P. E., & Running, S. W. (1997). A continental phenology model for monitoring vegetation responses to interannual climatic vriability. *Global Biogeochemical Cycles, 11*(2), 217–234.

WHO (World Health Organization). (2003). *Phenology and human health: Allergic disorders*. Copenhagen; WHO regional office for Europe, 55p. Available at http://apps.who.int/iris/bitstream/10665/107479/1/e79129.pdf. Accessed January 29 2017.

Xie, Y., Wang, X., & Silander, J. A. (2015). Deciduous forest responses to temperature, precipitation and drought imply complex climate change impacts. *PNAS, 112*(44), 13585–13590. doi:10.1073/pnas.1509991112.