NIH Public Access
Author Manuscript
Atmos Environ (1994). Author manuscript; available in PMC 2016 February 01.

Published in final edited form as:
Atmos Environ (1994). 2015 February ; 103: 297–306. doi:10.1016/j.atmosenv.2014.12.019.

Predicting Onset and Duration of Airborne Allergenic Pollen Season in the United States

Yong Zhang[a,b,c], **Leonard Bielory**[a,d,e], **Ting Cai**[a,b,d], **Zhongyuan Mi**[a,b], and **Panos Georgopoulos**[a,b,c,d,*]

[a]Environmental and Occupational Health Sciences Institute (EOHSI), Rutgers University, 170 Frelinghuysen Rd., Piscataway, NJ 08854, USA

[b]Department of Environmental and Occupational Medicine, Rutgers University – Robert Wood Johnson Medical School, Piscataway, NJ 08854, USA

[c]Department of Chemical and Biochemical Engineering, Rutgers University, 98 Brett Rd., Piscataway, NJ 08854, USA

[d]Department of Environmental Sciences, Rutgers University, 14 College Farm Rd., New Brunswick, NJ 08901, USA

[e]Department of Medicine, Section of Allergy and Immunology, Robert Wood Johnson University Hospital, New Brunswick, NJ 08901, USA

Abstract

Allergenic pollen is one of the main triggers of Allergic Airway Disease (AAD) affecting 5% to 30% of the population in industrialized countries. A modeling framework has been developed using correlation and collinearity analyses, simulated annealing, and stepwise regression based on nationwide observations of airborne pollen counts and climatic factors to predict the onsets and durations of allergenic pollen seasons of representative trees, weeds and grass in the contiguous United States. Main factors considered are monthly, seasonal and annual mean temperatures and accumulative precipitations, latitude, elevation, Growing Degree Day (GDD), Frost Free Day (FFD), Start Date (SD) and Season Length (SL) in the previous year. The estimated mean SD and SL for birch (*Betula*), oak (*Quercus*), ragweed (*Ambrosia*), mugwort (*Artemisia*) and grass (*Poaceae*) pollen season in 1994–2010 are mostly within 0 to 6 days of the corresponding observations for the majority of the National Allergy Bureau (NAB) monitoring stations across the contiguous US. The simulated spatially resolved maps for onset and duration of allergenic pollen season in the contiguous US are consistent with the long term observations.

Keywords

Pollen season; Allergy; Start date; Season length; Model

*Corresponding author: Panos Georgopoulos, Tel: 848-445-0159; Fax: 732-445-0915; panosg@ccl.rutgers.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

1. Introduction

Allergenic pollen from birch (*Betula*), oak (*Quercus*), ragweed (*Ambrosia*), mugwort (*Artemisia*) and grass (*Poaceae*) is one of the main triggers of Allergic Airway Disease (AAD) affecting 5% to 30% of the population in industrialized countries (Cakmak et al., 2002). Synergism of allergenic pollen with outdoor air pollutants like ozone and particulate matter, and indoor house dust mites has been reported and may exacerbate the AAD of allergy sufferers (Cakmak et al., 2012; Kim et al., 2013). Prediction of the onset and duration of allergenic pollen season benefits allergy sufferers, particularly to provide crucial information for operational models forecasting pollen timing and levels at multiple spatiotemporal scales.

Observation-based models and process-based models have been developed to predict the onset and duration of allergenic pollen seasons from trees, weeds and grass. Observation-based models developed the statistical relationships between observed pollen data, such as Start Date (SD) and Season Length (SL), and observed climatic or meteorological factors using either regression (Brighetti et al., 2014; Emberlin et al., 1993; Garcia-Mozo et al., 2006), time series analyses (Damialis et al., 2007), or computational intelligence (Voukantsis et al., 2010; Zhang et al., 2013b). They have been used to predict allergenic pollen seasons of birch, oak, ragweed, mugwort and grass in most European countries and the US (Garcia-Mozo et al., 2006; Voukantsis et al., 2010; Zhang et al., 2013b). The deviations in predicted SD and SL from observations ranged from a few days to around two weeks.

Process-based models estimate SD and SL using onset date of each phenological phase of allergenic trees, weeds and grass. The phenological phase can be predicted through functions of environmental factors including accumulation of forcing and/or chilling temperature, photoperiod and water availability (García-Mozo et al., 2009; Sofiev and Bergmann, 2013). Growing Degree Day (GDD) is the commonly used phenological model to describe the development of phenology phases. It implicitly assumes that the chilling requirement for releasing dormancy has been met before the initial date of forcing temperature accumulation (Sofiev and Bergmann, 2013). The Unified model by Chuine (Chuine, 2000) and Promoter-Inhibitor model by Schaber and Badeck (Schaber and Badeck, 2003) are two generalized phenological models meant to determine the developmental phases of plants. Because some of the mechanisms governing flowering are still unknown and due to the difficulty in precisely incorporating some factors (e.g., photoperiod and soil humidity) into the model (Sofiev and Bergmann, 2013), the accuracy of estimated SD and SL from process-based models is generally lower than that obtained from observation-based models.

Most of the observation-based and process-based models were developed and parameterized on the basis of locally observed pollen and climate data (Sofiev and Bergmann, 2013). They were usually evaluated using individual or a few pollen species from a single or a small set of monitoring stations within a limited territory. Even for the same species, model parameters in different regions varied widely (Zhang et al., 2013a). Therefore it is hard to apply locally parameterized modeling approaches to a large geographical area (Sofiev and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Bergmann, 2013), which is usually required by operational pollen forecast models. The challenges of simulating allergenic pollen season onset and durations in a large geographical area have been emphasized by recent process-based pollen forecast models studying pollen production, emission and dispersion in the air (Sofiev et al., 2013; Zhang et al., 2014a; Zink et al., 2012).

In the current study, two observation-based models and one process-based model were developed to predict the SD and SL of allergenic pollen seasons of birch, oak, ragweed, mugwort and grass. The models were parameterized and cross validated using nationwide observations of airborne pollen data, and climate and/or meteorology data during the period of 1994–2010 in the contiguous United States (hereafter referred to as US). The study addressed the above mentioned challenges through: (1) providing a new scheme to handle the variable selection and parameter optimization in the field of predicting pollen season onset and duration; and (2) identifying modeling approaches applicable to large geographic areas (e.g. continent) to simulate the SD and SL for multiple allergenic species at multiple spatiotemporal scales.

2. Methods

2.1. Data source

Observed daily airborne pollen counts were obtained from all available monitoring stations of the National Allergy Bureau (NAB) at the American Academy of Allergy, Asthma and Immunology (AAAAI) during the period of 1994–2010 across the US. To retrieve more available data, airborne pollen data from two neighboring Canadian monitoring stations were also incorporated into the current study. The reported pollen data were only classified at the level of genus. Fifty-eight NAB stations were selected because they recorded valid data for at least four years for performing further analyses and modeling parameterization. Main climate characteristics and geographical locations of the studied stations are listed in Table S1 (Appendix).

Observed daily temperatures and precipitation from 1994 to 2010 were obtained from the National Oceanic and Atmospheric Administration (NOAA) meteorology stations nearest to the corresponding NAB pollen stations. Spatially resolved hourly surface temperatures for 2001 were obtained from established Weather Research and Forecasting (WRF) model simulations archived by the North American Regional Climate Change Assessment Program (NARCCAP) (Mearns et al., 2009). Configurations of the WRF simulations including model domain and physics parameterizations followed those described by Leung and Qian (2009) for the NARCCAP Phase I experiment.

2.2. Observed start date and season length

The long-range-transport pollen grains from surrounding regions may affect pollen counts at the beginning and ending of the local pollen season (Siljamo et al., 2013; Smith et al., 2008). In order to exclude the long-range-transport pollen grains from the local source, the Start Date (SD) and Season Length (SL) were defined based on the percentages of cumulative daily pollen count (Rasmussen, 2002; Sofiev and Bergmann, 2013). Start and end dates are defined as the Julian days. With day 1 being January 1st, the SD (days from January 1st) of a

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

pollen season is defined as the day when the cumulative pollen count reaches 5% and the end date is the day when it reaches 95% of annual total count. This definition is graphically demonstrated in Fig. S1 in the Appendix for pollen season observed at Springfield, NJ. Alternative visualization of the defined SD and SL together with the daily pollen counts can be found in the supplementary document of the previous study (Zhang et al., 2014b).

Furthermore, start date of spring-flowering birch, oak and grass should not be later than June 21st, because spring-flowering trees and grass are generally long-day species, thus shortening photoperiod after June 21st will trigger growth cessation, cold acclimation and dormancy development for these plants (Li et al., 2003). Weeds are generally short-day species which require accumulation of a consolidated period of darkness to flower (Deen et al., 2001; Ziska et al., 2011), start date of summer-flowering ragweed and mugwort should not be earlier than June 21st. Pollen data with SL of less than 7 or greater than 80 days are assumed unreasonable and excluded for further analyses (Zhang et al., 2013a); the exception is made for grass, which usually includes many different species and can have a pollen season length longer than 80 days.

2.3. Observation-based model (M1)

Allergenic pollen season onset and duration are associated with multiple climatic, meteorological and geographical factors. The associations can be generally described as $Y = f(T,P_r,Lat,H,GDD,Y_p,FFD,\Delta GDD,\Delta Y_p,\Delta FFD,\Delta T,\Delta P_r)$. Y is the vector of pollen season timing index ((SD, SL), days). T is a set which consists of the monthly, seasonal and annual mean temperatures (°C), representing the temperature effect on pollen season onset and duration. P_r is a set which consists of the monthly, seasonal or annual accumulative precipitations (mm), representing the precipitation effect on pollen season onset and duration. Monthly mean temperature and accumulative precipitation from September of the previous year to August of the current year were incorporated into the preliminary correlation analyses. Lat is the latitude (°N), H is the elevation above sea level (m), GDD is the growing degree days in a fixed period (Degree days), Y_p is the corresponding SD and SL in the previous year (days), and FFD is the Frost Free Days (days). ΔGDD, ΔY_p and ΔFFD, ΔT and ΔP_r are the deviations in GDD, Y_p, FDD, T and P_r for a given year from the corresponding long term averages over the period of 1994–2010 at a given location, respectively. The deviation in independent variable X is defined using eq (1),

$$\begin{cases} \Delta X = X - \overline{X}_{Lat} \\ \overline{X}_{Lat} = b_0 + b_1 Lat \end{cases} \quad (1)$$

where $\overline{X}_{Lat}$ and ΔX are the long term mean and deviations at corresponding latitude Lat, respectively; b_0 and b_1 are the coefficients.

For SD of ragweed and mugwort, the relationship among pollen season onset, duration, climatic, meteorological and geographical factors can be further formulated using eq (2),

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

$$\begin{cases} Y=\overline{Y}_{Lat}+\Delta Y \\ \overline{Y}_{Lat}=a_0+a_1 Lat \\ \Delta Y=g(\Delta GDD, \Delta Y_p, \Delta FFD, \Delta T, \Delta P_r) \end{cases} \quad (2)$$

where $\overline{Y}_{Lat}$ are the long term mean *SD* and *SL* at latitude *Lat*, ΔY are the deviations in SD and SL from the corresponding long term mean, a_0 and a_1 are the coefficients.

GDD in a fixed period is calculated using eq (3),

$$GDD=\sum_{ID}^{LD}(T_i-T_b), \quad T_i>T_b \quad (3)$$

where *ID* and *LD* are the Initial Date and Last Date to accumulate the temperature difference between daily temperature T_i and base temperature T_b. The parameters in eq (3) are optimized through eq (4) using simulated annealing (Chuine, 2000),

$$(ID, LD, T_b)=\arg \max_{ID<LD,T_b}(|\rho(Y, GDD)|) \quad (4)$$

where ρ is the Spearman correlation coefficient between *Y* and *GDD*. The Spearman correlation coefficient was used because it could handle small nonlinearity, if any, existing between *SD* and *GDD*, likewise for *SL* and *GDD*. *ID* took the value from January 1st, January 15th, February 1st, …, December 1st and December 15th. *LD* took the value from January 15th, January 31st, February. 15th, …, December 15th and December 31st; T_b assumed a value from −2 to 10°C with an interval being 0.25°C.

FFD in a given year is defined as the interval between the last frost day during spring and the first frost day (daily minimum temperature below 0°C) during fall.

2.4. Simplified observation-based model (M2)

A simplified observation-based model can be generally represented as $Y = f(T,P_r,Lat,H,GDD,FFD,\Delta GDD,\Delta FFD,\Delta T,\Delta P_r)$. This simplified model does not incorporate the influence of allergenic pollen season in the previous year, namely Y_p and ΔY_p in section **2.3**. Because of sparse collection of airborne pollen data in some regions for given years, it is usually hard to derive precise Y_p and ΔY_p for these regions. In this situation, the simplified model can provide reasonable approximate estimates of SD and SL.

2.5. GDD model (M3)

The GDD model was adopted to describe the onset (i.e. SD) and end dates of allergenic pollen season. As shown in eq (5),

$$\begin{cases} GDD_{Thr,SD}=\sum_{ID}(T_i-T_b), & T_i>T_b \\ GDD_{Thr,ED}=\sum_{ID}(T_i-T_b), & T_i>T_b \end{cases} \quad (5)$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the start and end dates are defined as the dates, when the accumulated temperature difference between daily temperature T_i and base temperature T_b reached threshold values $GDD_{Thr,SD}$ and $GDD_{Thr,ED}$, respectively. The SL is the interval between start and end dates. The *ID* and T_b are obtained through eq. (4)

The $GDD_{Thr,SD}$ and $GDD_{Thr,ED}$ may change slightly in different years even for same species at same location. In the current study, as shown in eq (6),

$$\begin{cases} GDD_{Thr,SD} \approx \overline{GDD}_{Thr,SD} = c_{SD0} + c_{SD1} Lat \\ GDD_{Thr,ED} \approx \overline{GDD}_{Thr,ED} = c_{ED0} + c_{ED1} Lat \end{cases} \quad (6)$$

$GDD_{Thr,SD}$, and $GDD_{Thr,ED}$, at a given latitude are approximated using their long term averages ($\overline{GDD}_{Thr,SD}$ and $\overline{GDD}_{Thr,ED}$) over the period of 1994–2010 at that latitude. The c_{SD0}, c_{SD1}, c_{ED0} and c_{ED1} are coefficients.

The observation-based model, its simplified version, and GDD model were labeled as M1, M2 and M3 in Table 1, respectively.

2.6. Model parameterization

For observation-based models, first, correlation analyses were conducted between SD and each of the climatic factors on the basis of nationwide observations of airborne pollen data and climate and/or meteorology data in the US, likewise for SL. The factors considered in the current study include annual mean temperature, annual accumulative precipitation, monthly mean temperature and accumulative precipitation from September of previous year to August of current year, GDD, FFD, latitude, elevation, and SD and SL in the previous year. Three or four of the factors with the highest correlation coefficients were prescreened to incorporate the influence of pollen information in previous year, and temperature and precipitation in both current and previous years. Second, the prescreened factors were further selected through stepwise regression, and collinearity analyses based on Variance Inflation Factor (VIF). VIF between climatic factors was required to be less than 5 to insure the independence of the final selected factors. Finally, the selected regression equation was parameterized using nationwide observed SD or SL and climatic factors.

For the GDD model, first, a GDD threshold value at a given station was calculated using observed SD or SL and meteorology data from that station for each available year during 1994–2010. Second, the long term averaged threshold GDD over the period of 1994–2010 (i.e., $\overline{GDD}_{Thr,SD}$ and $\overline{GDD}_{Thr,ED}$) at each station was calculated to parameterize its relationship with latitude (eq (6)).

As shown in Table S1 (Appendix), airborne pollen data are missing for some of the species in some years at some stations. The variable selection and parameterization are only based on the observed airborne pollen data available for each of the species at each of the NAB stations during 1994–2010. The missing data were not accounted towards the variable selection and parameterization of the models. The parameterization processes are depicted using schematic diagrams in Fig. S2 (Appendix).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2.7. Model evaluation

The model performance was evaluated using root mean square errors (RMSE) and deviations between observed and simulated mean SD and SL at each of the NAB stations during 1994–2010. The model performance was also validated using cross-validation by splitting the data into training set and validation set (Brighetti et al., 2014). We used a leave-one-out cross validation procedure, in which one year's data (e.g., 1994) were held out as a validation set, and the data of remaining years (e.g., 1995–2010) were used as a training set. The model was first trained using the training set and then used to predict the SD and SL for the validation year. The process was repeated for each of the years during 1994–2010 to check the validation error.

2.8. Model application

As an example, an observation-based model (M1 in Table 1) was applied to generate the spatially resolved SD and SL of allergenic pollen seasons for 2001 in the US with spatial resolution of 50×50 km using the NARCCAP archived meteorology data. For each of the 50×50 km grid cells, the M1 method for SD and SL was executed using the dependent variables in the corresponding cell. The SD_p and SL_p (i.e. SD and SL in 2000) in each cell required by the observation-based model were approximated using the long term mean SD and SL through equations in Figure 1A and B.

The simulated results were mapped only on cells in which the area coverage of a given allergenic species is greater than zero. The area coverage of birch, oak and grass were sourced from the Biogenic Emissions Land use Database, version 3.1 (BELD3.1) (Kinnee et al., 1997); and the coverage of ragweed and mugwort were generated using methods similar to those in the literature (Skjøth et al., 2010).

3. Results and discussion

3.1. Mean SD, SL and threshold GDD across latitude

Figure 1A shows that allergenic pollen seasons for spring-flowering birch, oak and grass start from the south and then shift gradually to the north in the US, while for summer-flowering ragweed and mugwort the flowering initiates from the north and then proceeds gradually towards the south. The durations of allergenic pollen seasons are longer in the south than those in the north (Figure 1B). Most of the relationships between mean SD or SL and latitude are statistically significant (p value < 0.05). The observed mean SD, SL and the standard deviation at each station are also listed in Table S2 (Appendix).

For the same species, the GDD threshold values for start and end dates of allergenic pollen season generally decrease as latitudes go higher in the US (Figure 1C and D). The exception is the threshold GDD for start date of oak pollen season. It slightly increases at higher latitudes. Oak has more subspecies than other allergenic plants. The observed SD may represent pollen timing from many subspecies. This may influence the general features of threshold GDD values for SD of oak pollen season.

Although ragweed and mugwort start flowering from the cold north towards the warm south (Figure 1A), they flower earlier at a given location if the annual mean temperature is higher

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

at that location (Fig. S3, Appendix). This can be explained by their threshold GDD for SD. At the high latitude (cold north), the threshold GDD for SD ($\overline{GDD}_{Thr,SD}$), required for ragweed and mugwort to flower, is much lower than that at the low latitude (warm south, Figure 1C). At a given location (latitude), the higher the temperature, the earlier the $\overline{GDD}_{Thr,SD}$ can be reached, and therefore the earlier the ragweed or mugwort will bloom.

3.2. Parameterization

Figure 2 presents the correlation coefficients between *SD*, Δ*SD* or *SL* and each individual climatic factor. For *SD* of birch and oak pollen, fixed-period *GDD* has the largest magnitude of correlation coefficient (−0.75 and −0.84 respectively; Figure 2 A1 and B1), and therefore can be used as a factor to account for influence of the current year temperature on *SD*. Monthly mean temperature from November of the previous year to March of the current year (T_{NDJFM}) also have relative larger correlation coefficients, and therefore can be used as a factor to account for influence of the preceding year temperature on SD. Similarly, for SD of grass pollen (Figure 2 C1), the annual mean temperature (T_c) and monthly mean temperature from September to December of the previous year (T_{SOND}) can be used as factors to incorporate influence of temperature on *SD*.

Since ragweed and mugwort flower from the cold north towards the warm south, direct correlation analyses between SD and temperature mask the feature that higher temperature at the same location drives ragweed and mugwort to bloom earlier. Instead, correlation analyses were conducted between the deviations in SD (Δ*SD*) and the deviations in climatic factors. The Δ*SD* of ragweed pollen is closely correlated with deviations in annual mean temperature (ΔT_c) and monthly mean temperature from November of the previous year to March of the current year (ΔT_{NDJFM}, Figure 2 D1). The Δ*SD* of mugwort pollen is closely correlated with deviations in *FFD* (Δ*FFD*) and annual accumulative precipitation (ΔPr_c, Figure 2 E1).

Following similar analyses, relevant climatic factors can be prescreened to account for the influence of temperature, precipitation and other factors on SL (Figure 2 A2–E2). Also shown in Figure 2 is that SD or SL in the previous year exerts important influence on SD or SL in the current year. Thus SD_p or SL_p were also used as prescreened factors for further analyses. Table 1 lists the prescreened factors and optimum parameter values for fixed-period GDD.

The prescreened variables were further selected using stepwise regression and collinearity analyses. The final selected variables and regression equations for the observation-based model and its simplified version are also listed in Table 1. In addition, Table 1 lists the GDD model on the basis of threshold values obtained through Figure 1 C and D.

On the basis of the Root Mean Square Error (RMSE), the observation-based model (M1) performed better than its simplified version (M2) and the GDD model (M3). The simplified observation-based model has the advantage to simulate the SD and SL without being dependent on continuously monitored aerobiological or phenological data. The GDD model performed best at predicting the SL for grass pollen, and SD for mugwort pollen. However,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the GDD threshold values for start and end dates (Figure 1 C and D) may face difficulties for application at latitudes outside the ranges of observations. Particularly, the $\overline{GDD}_{Thr,SD}$ for grass pollen decreases quickly as latitude goes higher; it is equal to the $\overline{GDD}_{Thr,SD}$ at a latitude of around 49.65°N. This makes the GDD model unfeasible for latitudes higher than 49.65°N.

Accumulation of chilling force, masting behavior, photoperiod and soil humidity also affect the SD and SL of allergenic season (Jato et al., 2007; Sofiev and Bergmann, 2013; Zhang et al., 2013a). In the current study, the initial date for accumulating daily temperature for SD is February 1st for birch, ragweed, mugwort and grass, and March 1st for oak. It is reasonable to assume that the chilling requirement was met before these initial dates in the US. Observed data and practical models of masting behavior, photoperiod and soil humidity are rarely available, and extremely limited for large geographic area like the US. These factors should be taken into consideration in future studies.

3.3. Evaluation

On the basis of the minimum RMSE for SD or SL for each species in Table 1, optimum simulation results were used to calculate the deviation in simulated mean SD or SL from the corresponding mean observations during 1994–2010 at each studied station ($SD_{Dev}=\overline{SD}_{Obs}-\overline{SD}_{Sim}$). The results are displayed in Figure 3 according to the locations of the studied NAB monitoring stations. To indicate the pollen abundance of a given allergenic species at a station, the size of each circle in Figure 3 is plotted proportionally to the average seasonal airborne pollen counts during 1994–2010 at that station. The average of seasonal counts and maximum daily counts are also listed in Table S3 (Appendix).

The deviations of simulated mean SDs are within 0–6 days at 63.6%, 72.4%, 68.8%, 94.4% and 56.4% of the studied stations for birch, oak, ragweed, mugwort and grass, respectively. The deviations of simulated mean SLs are within 0–6 days at 60.0%, 69.0%, 66.7%, 25.0% and 74.6% of the studied stations for birch, oak, ragweed, mugwort and grass respectively. In general, the models could simulate the middle range SD and SL reasonably well, but face difficulty in reproducing the extreme large or small observed SD and SL.

The RMSEs derived from cross-validation of the model are comparable with those listed in Table 1, which are based on models trained using the whole dataset from each of the stations during 1994–2010. This indicates that the proposed models are less likely to overfit the data. As an example, Fig. S4 (Appendix) presents the comparison between the observed and predicated SD and SL through the leave-one-out cross validation procedure using the M1 method for all pollen stations during 1994–2010.

3.4. Application

Figure 4 demonstrates the spatially resolved (50×50 km) maps of allergenic pollen season onset and duration for 2001 in US. For birch, oak and grass, the allergenic pollen seasons in 2001 proceeded from the south toward the north in the US. For ragweed and mugwort, the pollen seasons progressed from the north to the south. The durations of allergenic pollen

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

season in 2001 were longer in the south than those in the north in the US. These results are consistent with the long term observations in Figure 1A and B.

The spatially resolved maps of allergenic pollen season onset and duration can be used to drive the operational pollen forecasting model at multiple spatiotemporal scales, and to study the production, release, distribution and health effects of allergenic pollen in the US. Based on the availability of airborne pollen data, the methods presented in the current study can be potentially adapted to other countries, regions and species for studying biogenic aeroallergens in large geographic areas.

Although there are some spatial correlations of SD and SL among different pollen stations, these spatially resolved maps of allergenic pollen season onset and duration are generally hard to derive from geostatistics (e.g. kriging). On one hand, geostatistics generally need sufficient spatial sample points to generate reasonable interpolation values at unknown locations. Specifically, it is hardly convincing to interpolate spatially resolved maps of SD and SL using airborne pollen data from 58 stations across the US. On the other hand, geostatistics generally generate stationary interpolations of a given spatial phenomenon. For the current study, it is challenging to use geostatistics to extrapolate allergenic pollen season onset and duration for the years other than the study period of 1994–2010.

3.5. Uncertainty of the models

Uncertainties generally pervade in the entire modeling process (Beale and Lennon, 2012). They may result from different components of the modeling framework (Regan et al., 2002). The uncertainties in the prediction of allergenic pollen season onset and duration are mainly resulted from the model formulations, parameters and the input data. Table S4 (Appendix) lists the uncertainty sources and the corresponding general treatments regarding prediction of allergenic pollen season onset and duration. Here the discussion is focused on uncertainties introduced by the definition of start and end dates, and the rules used in section **2.2** to exclude potential 'abnormal' pollen observations. Discussion on uncertainties introduced by model formulations and input data are presented in the Appendix.

In the current study, the start and end dates are defined on the basis of percentages (i.e. 5% and 95%) of the annual cumulative airborne pollen count as in the literature (Drzeniecka-Osiadacz et al., 2014; Laaidi et al., 2003). This has been done on the basis of an exploratory graphic method. The accumulated daily percentage of pollen count is approximately a sigmoid curve. The values of 5% and 95% correspond to approximately the two inflection points. While this definition was chosen to reduce the impact of long-range-transport pollen on local pollen season, it could be subjected to uncertainties, because it is difficult to separate the long-range-transport pollen from the local source using airborne pollen count, and the ratio of long-range-transport pollen to local pollen may change station by station. Depending on the aim of studies, geographic areas and pollen species, different definitions of pollen season onset have been proposed on the basis of a threshold of accumulated pollen sums or the constant occurrence of pollen grains in the air (Myszkowska, 2014; Pauling et al., 2014).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

We excluded part of the pollen data for further analyses if the derived pollen season duration is shorter than 7 or longer than 80 days. This choice was based on preliminary data analyses. It shows that pollen season duration shorter than 7 or longer than 80 days generally has either very low daily pollen counts (around 1 to 5 pollen grains per day), or registers zero pollen grains in many consecutive days during a pollen season. This may suggest that there are no allergenic species growing in the surroundings of the corresponding pollen stations. Statistically, extreme values of SL can be approximately treated as “outliers”, which generally cause poor model performance. Around 5% of the observations were excluded for further analyses through application of the exclusion rule. The choice of cutoff pollen season duration (i.e. 7 and 80 days) may introduce uncertainties in model predictions.

In general, the uncertainties introduced by model formulations can be tested and reduced through adoption of new model methods or modification of the existing model formulations. The uncertainties introduced by model parameters can be quantified and potentially reduced on the basis of global uncertainty and sensitivity analyses using algorithms such as Morris’ design (Saltelli et al., 2000). Statistical techniques can be used to recognize and remove the biased data.

4. Conclusion

A modeling framework has been developed and tested based on nationwide observations of airborne pollen counts and climatic factors to predict the onsets and durations of allergenic pollen seasons of representative trees, weeds and grass in the US. The estimated mean SD and SL for birch (*Betula*), oak (*Quercus*), ragweed (*Ambrosia*), mugwort (*Artemisia*) and grass (*Poaceae*) pollen season in 1994–2010 are mostly within 0 to 6 days of the corresponding observations for the majority of the NAB stations across the US.

The simulated spatially resolved maps for onset and duration of allergenic pollen season in the US are consistent with the long term observations. This spatially resolved information of allergenic pollen season onset and duration provides substantial components to advance the operational models studying the emission, transport and health effects of allergenic pollen at multiple spatiotemporal scales.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded in part by USEPA under STAR Grant EPA-RD-83454701-0 to Rutgers University, by the NIEHS sponsored Center for Environmental Exposures and Disease at EOHSI (P30ES005022), and by the Ozone Research Center which is funded by the State of New Jersey Department of Environmental Protection (Contract # AQ05-011). We thank Dr. Ruby Leung (PNNL), Dr Alan Robock (Rutgers), and NARCCAP for help with the meteorology data, NAB for providing airborne pollen data, and Ms. Linda Everett for editorial assistance.

Abbreviations

AAAAI American Academy of Allergy, Asthma and Immunology

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

AAD	allergic airway disease
FFD	frost free day
GDD	growing degree day
NAB	National Allergy Bureau
NARCCAP	North American Regional Climate Change Assessment Program
NOAA	National Oceanic and Atmospheric Administration
SD	start date
SL	season length
WRF	Weather Research & Forecasting model

References

Beale CM, Lennon JJ. Incorporating uncertainty in predictive species distribution modelling. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012; 367:247–258.

Brighetti M, Costa C, Menesatti P, Antonucci F, Tripodi S, Travaglini A. Multivariate statistical forecasting modeling to predict Poaceae pollen critical concentrations by meteoclimatic data. Aerobiologia. 2014; 30:25–33.

Cakmak S, Dales RE, Burnett RT, Judek S, Coates F, Brook JR. Effect of airborne allergens on emergency visits by children for conjunctivitis and rhinitis. The Lancet. 2002; 359:947–948.

Cakmak S, Dales RE, Coates F. Does air pollution increase the effect of aeroallergens on hospitalization for asthma? Journal of Allergy and Clinical Immunology. 2012; 129:228–231. [PubMed: 22035655]

Chuine I. A unified model for budburst of trees. Journal of Theoretical Biology. 2000; 207:337–347. [PubMed: 11082304]

Damialis A, Halley JM, Gioulekas D, Vokou D. Long-term trends in atmospheric pollen levels in the city of Thessaloniki, Greece. Atmospheric Environment. 2007; 41:7011–7021.

Deen W, Swanton CJ, Hunt LA. A mechanistic growth and development model of common ragweed. Weed science. 2001; 49:723–731.

Drzeniecka-Osiadacz A, Krynicka J, Malkiewicz M, Klaczak K, Migała K. Statistical modelling of the main features of the Artemisia pollen season in Wrocław, Poland, during the 2002–2011 time period. Theoretical and Applied Climatology. 2014:1–14.

Emberlin J, Savage M, Woodman R. Annual variations in the concentrations of Betula pollen in the London area, 1961–1990. Grana. 1993; 32:359–363.

García-Mozo H, Galán C, Belmonte J, Bermejo D, Candau P, Díaz de la Guardia C, Elvira B, Gutiérrez M, Jato V, Silva I, Trigo MM, Valencia R, Chuine I. Predicting the start and peak dates of the Poaceae pollen season in Spain using process-based models. Agricultural and Forest Meteorology. 2009; 149:256–262.

Garcia-Mozo H, Galan C, Jato V, Belmonte J, de la Guardia C, Fernandez D, Gutierrez M, Aira M, Roure J, Ruiz L, Trigo M, Dominguez-Vilches E. Quercus pollen season dynamics in the Iberian peninsula: response to meteorological parameters and possible consequences of climate change. Ann Agric Environ Med. 2006; 13:209–224. [PubMed: 17195993]

Jato V, Rodriguez-Rajo FJ, Aira MJ. Use of phenological and pollen-production data for interpreting atmospheric birch pollen curves. Ann Agric Environ Med. 2007; 14:271–280. [PubMed: 18247464]

Kim KH, Jahan SA, Kabir E. A review on human health perspective of air pollution with respect to allergies and asthma. Environment International. 2013; 59:41–52. [PubMed: 23770580]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Kinnee E, Geron C, Pierce T. United states land use inventory for estimating biogenic ozone precursor emissions. Ecological Applications. 1997; 7:46–58.

Laaidi M, Thibaudon M, Besancenot JP. Two statistical approaches to forecasting the start and duration of the pollen season of Ambrosia in the area of Lyon (France). International Journal of Biometeorology. 2003; 48:65–73. [PubMed: 12783292]

Leung LR, Qian Y. Atmospheric rivers induced heavy precipitation and flooding in the western U.S. simulated by the WRF regional climate model. Geophys Res Lett. 2009; 36:L03820.

Li C, Junttila O, Ernstsen A, Heino P, Palva ET. Photoperiodic control of growth, cold acclimation and dormancy development in silver birch (Betula pendula) ecotypes. Physiologia Plantarum. 2003; 117:206–212.

Mearns LO, Gutowski W, Jones R, Leung R, McGinnis S, Nunes A, Qian Y. A Regional Climate Change Assessment Program for North America. Eos Trans AGU. 2009:90.

Myszkowska D. Predicting tree pollen season start dates using thermal conditions. Aerobiologia. 2014:1–15.

Pauling A, Gehrig R, Clot B. Toward optimized temperature sum parameterizations for forecasting the start of the pollen season. Aerobiologia. 2014; 30:45–57.

Rasmussen A. The effects of climate change on the birch pollen season in Denmark. Aerobiologia. 2002; 18:253–265.

Regan HM, Colyvan M, Burgman MA. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecological Applications. 2002; 12:618–628.

Saltelli, A.; Chan, K.; Scott, EM. Sensitivity analysis. Wiley; New York: 2000.

Schaber J, Badeck FW. Physiology-based phenology models for forest tree species in Germany. International Journal of Biometeorology. 2003; 47:193–201. [PubMed: 12698325]

Siljamo P, Sofiev M, Filatova E, Grewling L, Jager S, Khoreva E, Linkosalo T, Jimenez SO, Ranta H, Rantio-Lehtimaki A, Svetlov A, Veriankaite L, Yakovleva E, Kukkonen J. A numerical model of birch pollen emission and dispersion in the atmosphere. Model evaluation and sensitivity analysis. International Journal of Biometeorology. 2013; 57:125–136. [PubMed: 22434484]

Skjøth CA, Smith M, Šikoparija B, Stach A, Myszkowska D, Kasprzyk I, Radišić P, Stjepanović B, Hrga I, Apatini D, Magyar D, Páldy A, Ianovici N. A method for producing airborne pollen source inventories: An example of Ambrosia (ragweed) on the Pannonian Plain. Agricultural and Forest Meteorology. 2010; 150:1203–1210.

Smith M, Skjøth CA, Myszkowska D, Uruska A, Puc M, Stach A, Balwierz Z, Chlopek K, Piotrowska K, Kasprzyk I, Brandt J. Long-range transport of Ambrosia pollen to Poland. Agricultural and Forest Meteorology. 2008; 148:1402–1411.

Sofiev, M.; Bergmann, K-C. Allergenic Pollen: A Review of the Production, Release, Distribution and Health Impacts. Springer; 2013.

Sofiev M, Siljamo P, Ranta H, Linkosalo T, Jaeger S, Rasmussen A, Rantio-Lehtimaki A, Severova E, Kukkonen J. A numerical model of birch pollen emission and dispersion in the atmosphere. Description of the emission module. International Journal of Biometeorology. 2013; 57:45–58. [PubMed: 22410824]

Voukantsis D, Niska H, Karatzas K, Riga M, Damialis A, Vokou D. Forecasting daily pollen concentrations using data-driven modeling methods in Thessaloniki, Greece. Atmospheric Environment. 2010; 44:5101–5111.

Zhang R, Duhl T, Salam M, House J, Flagan R, Avol E, Gilliland F, Guenther A, Chung S, Lamb B. Development of a regional-scale pollen emission and transport modeling framework for investigating the impact of climate change on allergic airway disease. Biogeosciences. 2014a; 11:1461–1478.

Zhang Y, Bielory L, Georgopoulos P. Climate change effect on Betula (birch) and Quercus (oak) pollen seasons in the United States. International Journal of Biometeorology. 2013a:1–11.

Zhang Y, Bielory L, Mi Z, Cai T, Robock A, Georgopoulos P. Allergenic pollen season variations in the past two decades under changing climate in the United States. Global Change Biology. 2014b10.1111/gcb.12755

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Zhang Y, Isukapalli SS, Bielory L, Georgopoulos PG. Bayesian analysis of climate change effects on observed and projected airborne levels of birch pollen. Atmospheric Environment. 2013b; 68:64–73. [PubMed: 23526049]

Zink K, Vogel H, Vogel B, Magyar D, Kottmeier C. Modeling the dispersion of Ambrosia artemisiifolia L. pollen with the model system COSMO-ART. International Journal of Biometeorology. 2012; 56:669–680. [PubMed: 21744099]

Ziska L, Knowlton K, Rogers C, Dalan D, Tierney N, Elder MA, Filley W, Shropshire J, Ford LB, Hedberg C, Fleetwood P, Hovanky KT, Kavanaugh T, Fulford G, Vrtis RF, Patz JA, Portnoy J, Coates F, Bielory L, Frenz D. Recent warming by latitude associated with increased length of ragweed pollen season in central North America. Proceedings of the National Academy of Sciences. 2011; 108:4248–4251.

Appendix: Supplementary Data

Supplementary data to this article can be found online at http://_____.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

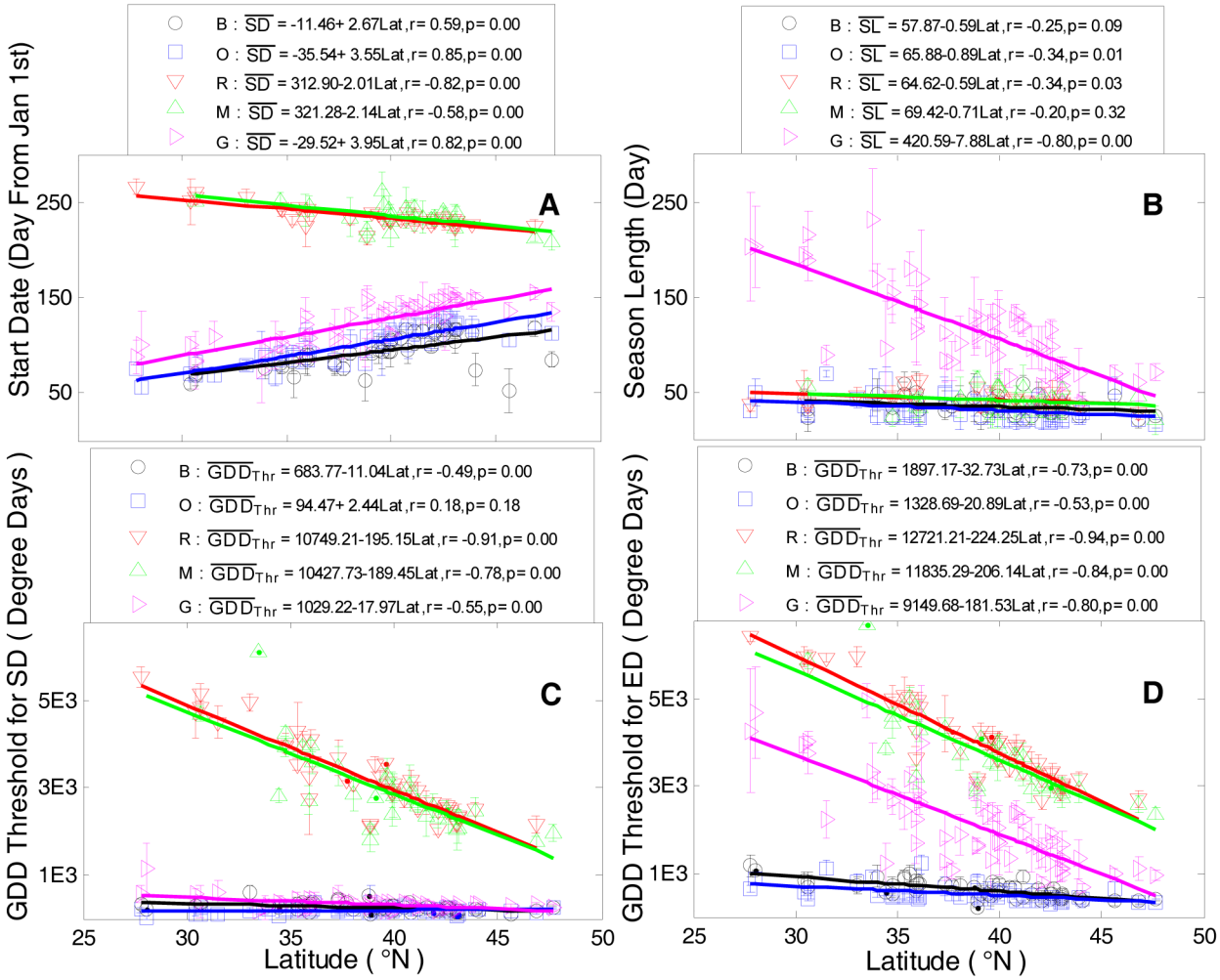

Figure 1.

Mean Start Date (SD), Season Length (SL), threshold GDDs and their corresponding standard deviations for birch (B), oak (O), ragweed(R), mugwort (M) and grass (G) across the latitudes. The mean is defined as the average over the observation period of 1994–2010 at a station. Regression equations for the trends are presented in the legends. (A) Start Date, (B) Season Length, (C) GDD threshold for SD, and (D) GDD threshold for ED.

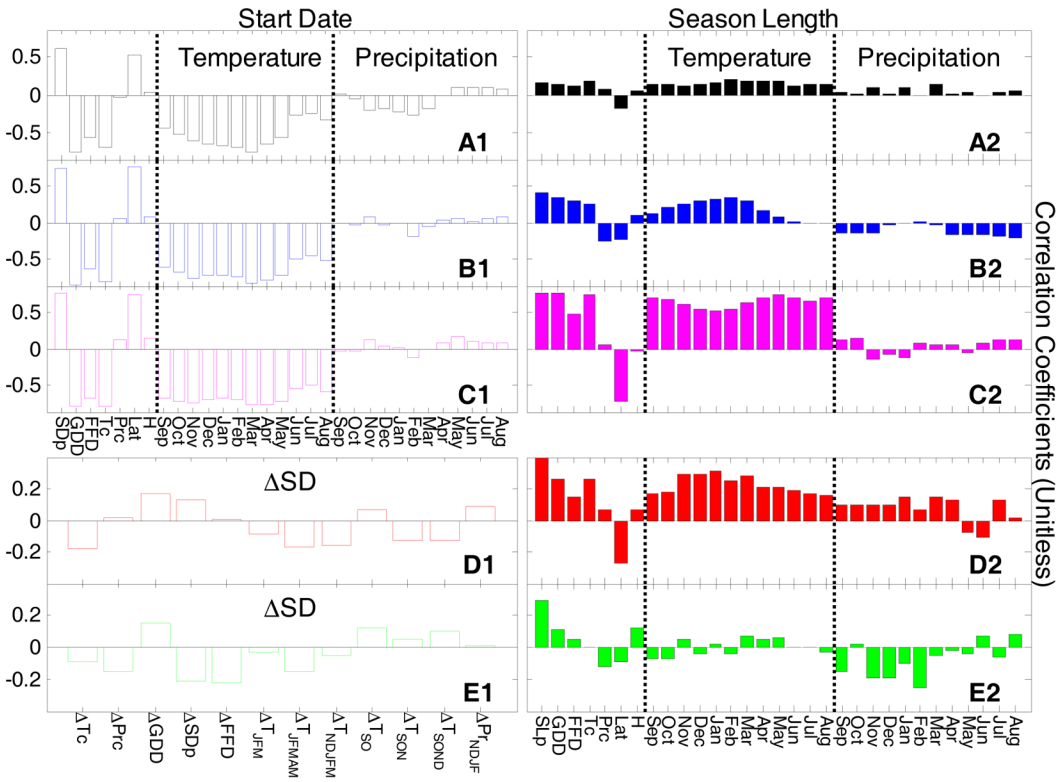

Figure 2.

Correlation coefficients between Start Date (SD), Deviation in SD (ΔSD), or Season Length (SL) and an individual climatic factor. These factors include SD and SL in the previous year (SD_P, SL_P), Growing Degree Days (GDD), Frost Free Days (FFD), annual accumulative precipitation (Pr_C) and mean temperature (T_C) in the current year, latitude (Lat), height (H), and monthly accumulative precipitation and mean temperature from September in the preceding year to August in the current year. Subscripts in the aggregated climatic factors indicate the consecutive months, in which the mean temperatures or total precipitation are calculated. (A1–E1) correlation coefficients of SD or Δ*SD* for birch, oak, grass, ragweed and mugwort, respectively. (A2–E2) correlation coefficients of SL for birch, oak, grass, ragweed and mugwort, respectively.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

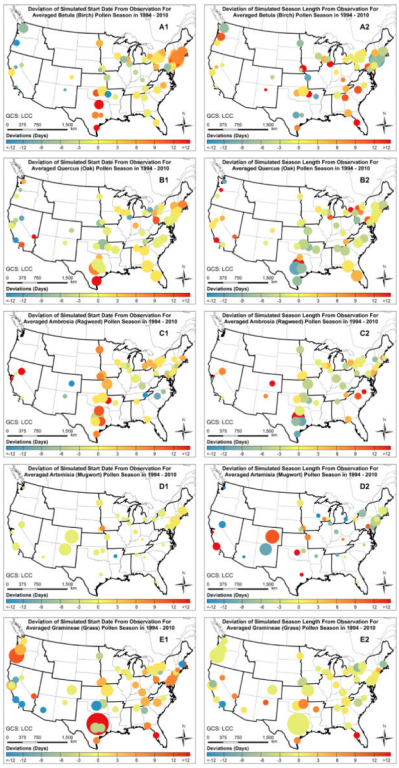

Figure 3.

Deviation of simulated mean Start Date (SD) and Season Length (SL) from the mean observations in 1994–2010 at each NAB station, i.e. $SD_{Dev} = \overline{SD}_{Obs} - \overline{SD}_{Sim}$. The size of the circle indicates the average pollen abundance for a given species during the same period. (A1–E1) deviations in SD for birch, oak, ragweed, mugwort and grass pollen, respectively. (A2–E2) deviations in SL for birch, oak, ragweed, mugwort and grass pollen, respectively.

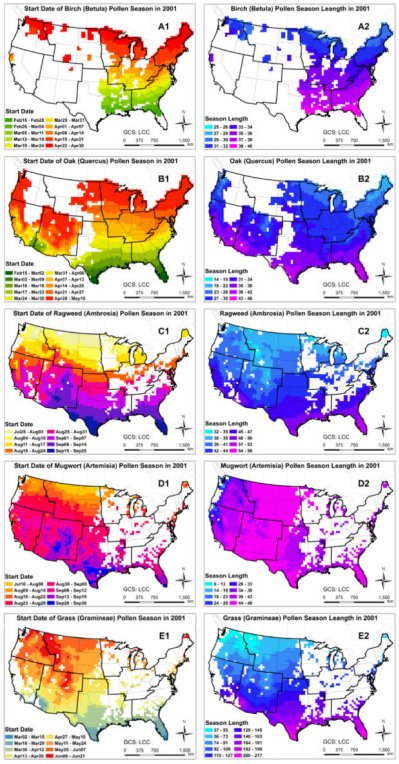

Figure 4.

Simulated Start Date (SD) and Season Length (SL) of allergenic pollen season using method M1 (Table 1) based on NARCCAP archived meteorology simulation data in 2001. Data were mapped only on cells in which the area coverage of a given allergenic species is greater than zero. (A–E1) SD for birch, oak, ragweed, mugwort and grass, respectively. (A2–E2) SL for birch, oak, ragweed, mugwort and grass, respectively.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Summary of three models simulating the Start Date (SD) and Season Length (SL) of pollen season. Listed in the table are the prescreened variables, optimum parameters for Growing Degree Day (GDD), selected regression equation, the corresponding *P* value, adjusted coefficient of determination (R^2), and Root Mean Square Error (RMSE). Subscripts in the aggregated climatic factors indicate the consecutive months in which the mean temperatures or total precipitations are calculated.

			Birch	Oak	Ragweed	Mugwort	Grass
Start Date	Variables		GDD, SD_p, T_{NDJFM}	GDD, SD_p, T_{NDJFM}	ΔT_c, ΔSD_p, ΔT_{NDJFM}	ΔFFD, ΔSD_p, ΔPr_c	T_c, SD_p, T_{SOND}
	ID, LD, T_b (°C)		Feb. 1st, Apr. 15th, 3	Mar. 1st, Apr. 15th, 5	Feb. 1st, Feb. 28th, 0.5	Feb. 1st, Feb. 28th, 1.8	Feb. 1st, May 15th, 7.3
	M1	Equation	SD=95.24-0.05GDD+0.20SD_p, P = 0.00, R^2 = 0.59	SD=106.35-0.09GDD +0.18SD_p, P = 0.00, R^2 = 0.73	ΔSD=7.9×10^{-3}-1.61ΔTc + 0.11ΔSD_p, P = 0.00, R^2 = 0.03	ΔSD= −0.25-0.19ΔFFD-0.17ΔSD_p, P = 0.00, R^2 = 0.08	SD=134.54-3.69T_c+0.31SD_p, P = 0.00, R^2 = 0.64
		RMSE (days)	**12.9**	**10.4**	**10.0**	15.1	**15.1**
	M2	Equation	SD=118.66-0.06GDD, P = 0.00, R^2 = 0.58	SD=128.42-0.10GDD, P = 0.00, R^2 = 0.72	ΔSD=2.9×10^{-3}-1.53ΔTc, P = 0.01, R^2 = 0.02	ΔSD=−0.20-0.20ΔFFD, P = 0.01, R^2 = 0.05	SD=193.33-5.23T_c, P = 0.00, R^2 = 0.60
		RMSE (days)	13.3	10.7	10.0	15.2	16.0
	M3	Equation	Eqs. (5) and (6)	Eqs. (5) and (6)	Eqs. (5) and (6)	Eqs. (5) and (6)	Eqs. (5) and (6)
		RMSE (days)	26.1	29.4	12.3	**5.3**	33.8
Season Length	Variables		T_{JFMAM}, SL_p, T_{NDJFM}	T_{JFM}, SL_p, T_{NDJFM}, Pr_c	T_{JFM}, SL_p, T_{NDJFM}	GDD, SL_p, Pr_{NDJF}	GDD, SL_p, T_{SO}
	ID, LD, T_b (°C)		Jan. 1st, Apr. 30th, 9.5	Feb. 1st, Mar. 31st, 6.3	Mar. 1st, Mar. 15th, 1.3	Feb. 1st, Apr. 15th, 9.5	Feb. 15th, Jun 30th, 10
	M1	Equation	SL=24.13+0.47T_{JFMAM}+0.16SL_p, P = 0.00, R^2 = 0.05	SL=23.76+0.60T_{JFM} -0.01Pr_c +0.29SL_p, P = 0.00, R^2 = 0.23	SL=28.14+0.58T_{NDJFM}+0.27SL_p, P = 0.00, R^2 = 0.18	SL=38.42-0.02Pr_{NDJF}+0.21SL_p, P = 0.00, R^2 = 0.10	SL=19.64+0.05GDD +0.46SL_p, P = 0.00, R^2 = 0.64
		RMSE (days)	**14.2**	**11.9**	**9.9**	**16.9**	29.8
	M2	Equation	SL=28.55+0.60T_{JFMAM}, P = 0.00, R^2 = 0.03	SL=33.93+0.76T_{JFM} -0.01Pr_c, P = 0.00, R^2 = 0.15	SL=38.92+0.72T_{NDJFM}, P = 0.00, R^2 = 0.11	SL=48.48-0.03Pr_{NDJF}, P = 0.00, R^2 = 0.06	SL=37.46+0.10GDD, P = 0.00, R^2 = 0.54
		RMSE (days)	14.4	12.5	10.4	17.3	33.8
	M3	Equation	Eqs. (5) and (6)	Eqs. (5) and (6)	Eqs. (5) and (6)	Eqs. (5) and (6)	Eqs. (5) and (6)
		RMSE (days)	15.5	15.0	29.9	33.4	**5.7**