Geoscientific
Model Development

A prognostic pollen emissions model for climate models (PECM1.0)

Matthew C. Wozniak and Allison L. Steiner

Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109, USA

Correspondence to: Matthew C. Wozniak (mcwoz@umich.edu)

Abstract. We develop a prognostic model called Pollen Emissions for Climate Models (PECM) for use within regional and global climate models to simulate pollen counts over the seasonal cycle based on geography, vegetation type, and meteorological parameters. Using modern surface pollen count data, empirical relationships between prior-year annual average temperature and pollen season start dates and end dates are developed for deciduous broadleaf trees (*Acer, Alnus, Betula, Fraxinus, Morus, Platanus, Populus, Quercus, Ulmus*), evergreen needleleaf trees (Cupressaceae, Pinaceae), grasses (Poaceae; C_3, C_4), and ragweed (*Ambrosia*). This regression model explains as much as 57 % of the variance in pollen phenological dates, and it is used to create a "climate-flexible" phenology that can be used to study the response of wind-driven pollen emissions to climate change. The emissions model is evaluated in the Regional Climate Model version 4 (RegCM4) over the continental United States by prescribing an emission potential from PECM and transporting pollen as aerosol tracers. We evaluate two different pollen emissions scenarios in the model using (1) a taxa-specific land cover database, phenology, and emission potential, and (2) a plant functional type (PFT) land cover, phenology, and emission potential. The simulated surface pollen concentrations for both simulations are evaluated against observed surface pollen counts in five climatic subregions. Given prescribed pollen emissions, the RegCM4 simulates observed concentrations within an order of magnitude, although the performance of the simulations in any subregion is strongly related to the land cover representation and the number of observation sites used to create the empirical phenological relationship. The taxa-based model provides a better representation of the phenology of tree-based pollen counts than the PFT-based model; however, we note that the PFT-based version provides a useful and "climate-flexible" emissions model for the general representation of the pollen phenology over the United States.

1 Introduction

Pollen grains are released from plants to transmit the male genetic material for reproduction. When lofted into the atmosphere, they represent a natural source of coarse atmospheric aerosols, ranging typically from 15 to 60 µm in diameter, while sometimes exceeding 100 µm (Cecchi, 2014; Sofiev et al., 2014). In the midlatitudes, much of the vegetation relies dominantly on anemophilous, or wind-driven, pollination (Lewis et al., 1983), representing a closely coupled relationship of pollen emissions to weather and climate. Anemophilous pollinators include woody plants such as trees and shrubs, as well as other non-woody vascular plants such as grasses and herbs. Pollen emissions are directly affected by meteorological (e.g., temperature, wind, relative humidity) and climatological (e.g., temperature, soil moisture) factors (Weber, 2003). Aerobiology studies indicate that after release, pollen can be transported on the order of 10–1000 km (Sofiev et al., 2006; Schueler and Schlünzen, 2006; Kuparinen et al., 2007) but there are still large uncertainties regarding emissions and transport of pollen.

Prognostic pollen emissions are useful for the scientific community and public, specifically for forecasting allergenic conditions or predicting the flow of genetic material. The interest and growing wealth of knowledge of allergenic pollen have been recently reviewed by Beggs et al. (2017). To date, most pollen emissions models focus on relatively short, seasonal timescales and smaller locales for a limited selection of taxa (Sofiev et al., 2013; Liu et al., 2016; R. Zhang et al., 2014). Climatic changes in large-scale pollen distributions are mostly absent from scientific literature, though mul-

tiple studies on phenological changes in the pollen season have been published (Ziska, 2016; Yue et al., 2015; Y. Zhang et al., 2015a). Only recently have regional-scale modeling studies of pollen dispersion been conducted for Europe, and they have been used to assess the impacts of climate change on airborne pollen distributions (Sofiev and Prank, 2016; Lake et al., 2017). In contrast to most meteorological pollen models, climate models require long-term (e.g., decadal- to century-scale) emissions at a range of resolutions covering continental regions up to the global scale. This distinction in both time and space requires a flexible model that can account for emissions without taxon-specific emission data (i.e., differentiation between genera or species) and can be used within aggregated vegetation descriptions, such as plant functional types (PFTs). Given recent interest in airborne biological particles and their role in climate (Després et al., 2012; Myriokefalitakis et al., 2017), an emissions model that captures longer temporal scales and broader spatial scales is key to developing global inventories and understanding pollen's role in the climate system. Here, we develop a model for use in the climate modeling community that can be used specifically to simulate pollen emissions on the decadal or centurial timescale for large regions using conventional climate or Earth system models.

Existing pollen forecasting models are often classified as either process-based phenological models or observation-based models (Scheifinger et al., 2013). Process-based phenological models employ a parameterization of plant physiology and climatic conditions (e.g., relating to the timing of flowering to a chilling period, photoperiod, or water availability). Pollen season phenology in an anemophilous species is inherently connected to its environment via relationships in the growing season dynamics (e.g., bud burst and temperature; Fu et al., 2012), and many models apply the same techniques to flowering as for bud burst (Chuine et al., 1999). This approach to phenology could be suited to climate models, given its flexibility for adaptive phenological events and regional-scale studies. Typically, these types of phenological models are taxa specific as well as regionally dependent, e.g., *Betula* in Europe or ragweed in California (Sofiev et al., 2013; Siljamo et al., 2013; R. Zhang et al., 2014). These models are usually calibrated to local data only even though distinct geographic differences exist for pollen phenology. Thus, such models may not perform equally well in other locations. Though process-based models draw a connection between an atmospheric state variable, i.e., temperature, and pollen emissions, at least three parameters are required for optimization and they are susceptible to overfitting (Linkosalo et al., 2008). While some process-based models may be scaled up to larger regions while maintaining appreciable accuracy (García-Mozo et al., 2009), such models are generally not practical for implementation in larger-scale climate modeling with regional climate models (RCMs) and global climate models (GCMs) because sufficient land cover data are not available at the appropriate taxonomic level.

In contrast to process-based models, observation-based methods determine the phenology of vegetation with statistical–empirical approaches (e.g., relating the start of the pollen season with mean temperatures preceding the pollen event) and often rely on regression models or time series modeling (Scheifinger et al., 2013). Time series modeling utilizes observations to define the deterministic and stochastic variability of pollen count observations and is frequently used in aerobiological studies (Moseholm et al., 1987; Box et al., 1994). Regression models, either using a single or multiple explanatory variable(s), exploit past relationships to define the magnitude of emissions as well as timing variables such as the start date and duration of the pollen season (Emberlin et al., 1999; Smith and Emberlin, 2005; Galán et al., 2008). Using local pollen count data, Y. Zhang et al. (2015b) completed a regional phenological analysis using multiple linear regressions for pollen in southern California for six taxa. Olsson and Jönsson (2014) show that empirical models based solely on spring temperature perform just as well as process-based models using the temperature forcing concept, and better than those including a chilling or dormancy-breaking requirement.

Observation-based methods assume stationarity, or the likelihood that the statistics of pollen counts or climate variables are not changing over time. For these models to apply outside of calibration period, they require that the driving pattern or relationship is maintained in the future (or past). For example, as the Earth's climate changes, these models do not represent the complex connections between pollen emissions and a warming world aside from the relationships determined empirically. However, these models provide clear and often simple formulations that have predictable behaviors and forgo the nuance of fitting ambiguous and uncertain parameters. We therefore choose to employ elements of the observational methods for this pollen emissions model formulation, as described in Sect. 4.

In addition to understanding the release of pollen grains, a second consideration is the large-scale transport of pollen. Once emitted to the atmosphere, pollen is mixed within the atmospheric boundary layer by turbulence and, depending on large-scale conditions, can be transported far from the emission source. Prior studies have used both Lagrangian (Hunt et al., 2002; Hidalgo et al., 2002) and Eulerian techniques to simulate the transport of pollen, with the former typically used for studies of crop germination and the latter primarily for allergen forecasting. For example, Helbig et al. (2004) used the meteorological model KAMM (Karlsruher Meteorologisches Modell) with the DRAIS (Dreidimensionales Ausbreitungs- und Immissions-Simulationsmodell) turbulence component to simulate daily pollen counts for regions over Europe. Schueler and Schlünzen (2006) use a mesoscale atmospheric model (METRAS) to quantify the release, transport, and deposition of oak pollen for a 2-day period over Europe. Sofiev et al. (2013) includes the long-range transport of birch pollen over western Europe by developing
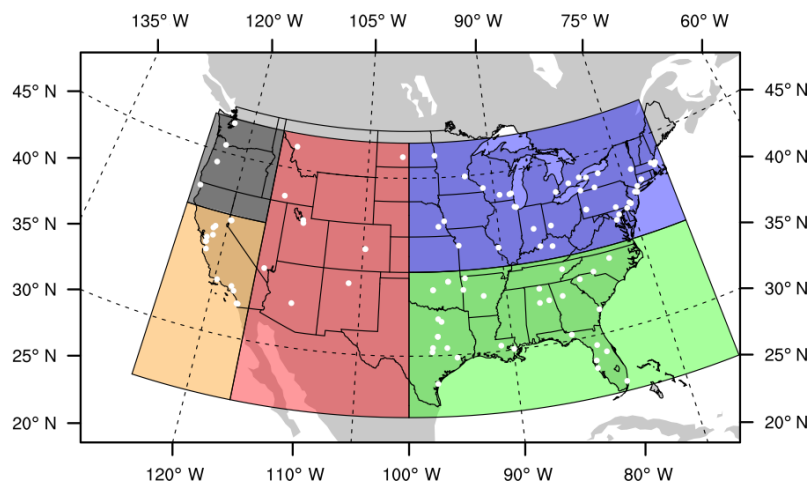
Figure 1. Locations of American Academy of Allergy, Asthma and Immunology (AAAAI) station and geographic regions used in this study: Northeast (NE; 38–48° N, 70–100° W) in blue, Southeast (SE; 25–38° N, 70–100° W) in green, Mountain (MT; 25–48° N, 100–116° W) in red, California (CA; 25–40° N, 116–125° W) in orange, and Pacific Northwest (PNW; 40–48° N, 116–125° W) in dark grey.

a birch pollen map and a flowering model to trigger release in the Finnish System for Integrated modeling of Atmospheric coMposition (SILAM). Efstathiou et al. (2011) developed a pollen emissions model for use within the regional air quality model (the Community Multi-scale Air Quality model; CMAQ) and tested their model with birch and ragweed taxa. R. Zhang et al. (2014) implemented a similar pollen emissions scheme with a regional numerical weather prediction model (the Weather Research and Forecasting (WRF) modeling system). Zink et al. (2013) developed a generic pollen modeling parameterization for use with a numerical weather prediction model (COSMO-ART) that is flexible to include differing pollen taxa. Collectively, these relatively new developments suggest a growing interest in the prognostic estimation of pollen in the short term for seasonal allergen forecasting on the weather (e.g., 1- to 2-week) timescale.

In this paper, we build on these coupled emissions–transport models and develop a comprehensive emissions model (Pollen Emissions for Climate Models; PECM) for use at climate model timescales that covers the majority of pollen sources in subtropical to temperate climes, including woody plants, grasses, and ragweed. First, we summarize the spatial distribution and seasonality of pollen counts for various taxa in the United States based on current observations (Sect. 2). Then, we develop new pollen emissions parameterization for climate studies (Sect. 4), transport these emissions over the continental United States (CONUS) using the Regional Climate Model version 4 (RegCM4) (Giorgi et al., 2012), and evaluate the results using 8 years of observed pollen count data (Sect. 5). We implement two different land cover classification schemes to illustrate the uncertainties associated with vegetation representation for trees including (1) detailed family- or genus-level tree distributions over CONUS, and (2) the use of PFT level distribu-

tions, which group vegetation types by physiological characteristics (Sect. 3). As the latter provides a greater opportunity for expansion into regional- and global-scale climate models over multiple domains, we discuss the effects that the PFT-based categorization has on the total estimated source strength of pollen. Finally, the limitations of this emissions framework and suggestions for future developments are included (Sect. 6).

2 Observed pollen phenology

2.1 Data description

The National Allergy Bureau (NAB) of the American Academy of Allergy, Asthma and Immunology (AAAAI) conducts daily pollen counts at 96 sites in cities across the United States (US), its territories, and several locations in southern Canada. All NAB sites implement a volumetric air sampler and certified pollen count experts to conduct daily pollen counts (grains m^{-3}) for up to 42 plant taxa at either the family level (e.g., Cupressaceae, Poaceae), genus level (e.g., *Acer*, *Quercus*), or for four generic categories termed "other grass pollen", "other tree pollen", "other weed pollen", or "unidentified". We use NAB pollen count data ranging from 2003 to 2010 at all stations in the continental United States (Fig. 1) for selected taxa to develop and evaluate PECM, and to determine the phenology of wind-driven pollen. Individual station locations and descriptions are included in Table S1 in the Supplement.

We evaluate the observed pollen counts to determine the vegetation types that emit the largest magnitude of pollen over the continental United States. Since many of the taxa reported at the 96 NAB sites frequently have very low pollen counts (e.g., less than 10 grains m^{-3}), a threshold for the
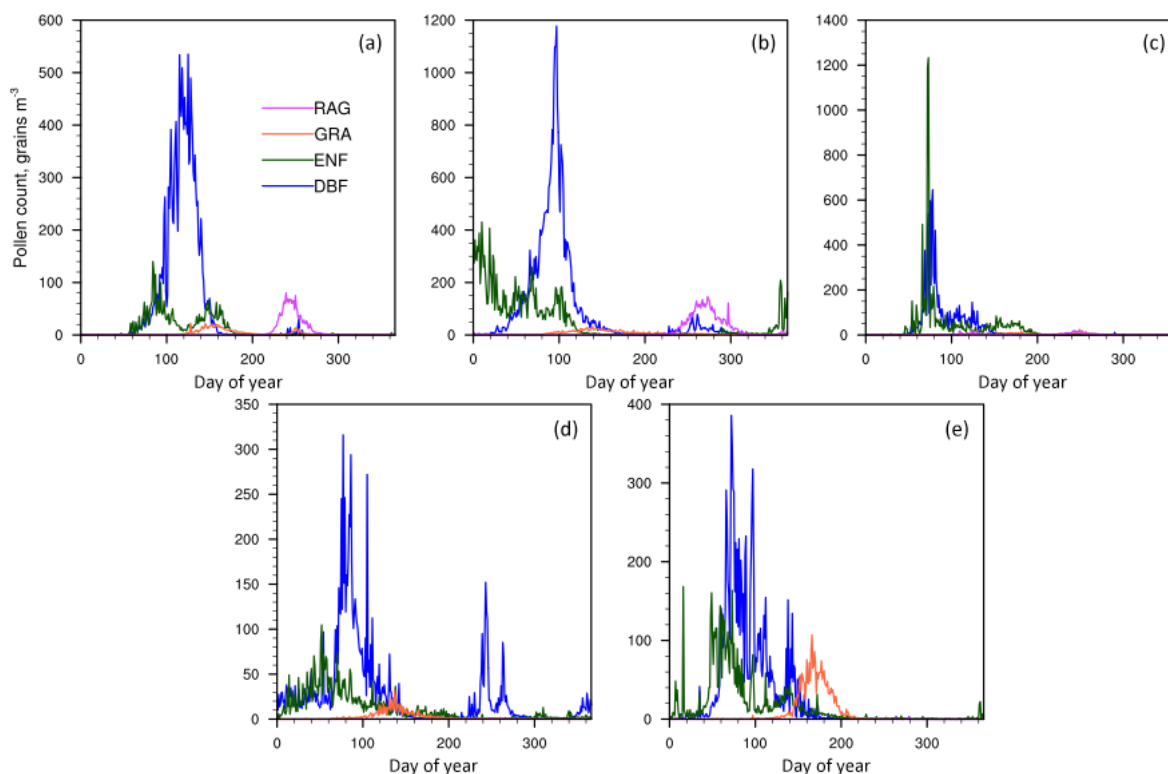
Figure 2. Daily observed average time series of daily pollen count data (2003–2010) for the four representative plant functional types (deciduous broadleaf forest, DBF; evergreen needleleaf forest, ENF; grasses, GRA; and ragweed, RAG) averaged over the five regions in Fig. 1: **(a)** Northeast, **(b)** Southeast, **(c)** Mountain, **(d)** California, and **(e)** Pacific Northwest.

grain count is set to select the taxa with the highest pollen counts. We calculate the average of the annual maximum pollen count across all years (2003–2010), P_{avgmax}, at each site for each counted taxon. We then select taxa to include in PECM using two criteria: (1) the maximum of P_{avgmax} among all stations exceeds 100 grains m^{-3}, and (2) the average P_{avgmax} among all stations exceeds 70 grains m^{-3} (Table S2). Using these two criteria, 13 taxa are selected for inclusion in the model, including *Acer, Alnus, Ambrosia, Betula*, Cupressaceae, *Fraxinus*, Poaceae, *Morus*, Pinaceae, *Platanus, Populus, Quercus*, and *Ulmus*. These 13 taxa account for about 77 % of the total pollen counted across the United States during 2003–2010.

The 13 dominant pollen types are grouped into four main categories by plant functional type: deciduous broadleaf forest (DBF), evergreen needleleaf forest (ENF), grasses (GRA), and ragweed (RAG). Plant functional type is a land cover classification commonly used in the land surface component of climate models, and this categorization will allow flexibility to apply the emissions model to other climate models. The DBF category includes nine genus-level taxa (*Acer, Alnus, Betula, Fraxinus, Morus, Platanus, Populus, Quercus*, and *Ulmus*) and the ENF category includes two family-level taxa (Cupressaceae and Pinaceae). The grass PFT utilizes pollen count data from the Poaceae family, although we note

that the grass PFT classification may include herbs and other non-woody species that may emit pollen as well. *Ambrosia* (ragweed) is segregated as its own category (RAG), due to its high pollen counts in the early autumn and unique land cover features. Daily pollen counts were summed for each PFT prior to calculating an 8-year average pollen time series.

2.2 Observed seasonality of pollen emissions

Pollen counts are analyzed over five subregions based on their climatic differences (Fig. 1; Table S1) to identify emissions patterns over the continental United States. These five subregions are the "Northeast" (temperate; 38–48° N and 70–100° W; 34 stations), "Southeast" (temperate, subtropical; 25–38° N and 70–100° W; 29 stations), "Mountain" (varied climate; 25–48° N and 100–116° W; 9 stations), "California" (Mediterranean, varied climate; 25–40° N and 116–125° W; 13 stations), and "Pacific Northwest" (temperate rainforest; west of 116° W and north of 40° N; 4 stations). Figure 2 shows the observed average PFT daily pollen counts averaged over all stations within the defined subregions.

For deciduous broadleaf forest (DBF) taxa, the Southeast has the highest average pollen maximum, reaching up to about 700–1200 grains m^{-3} around day 100. In the Northeast, DBF is the dominant PFT, reaching up to an average

of $400\,\mathrm{grains\,m^{-3}}$ and peaking slightly later (around day 120) than in the Southeast. California sites show an average peak around $150\,\mathrm{grains\,m^{-3}}$ occurring slightly earlier around day 80. A sharp maximum of $775\,\mathrm{grains\,m^{-3}}$ appears in the Mountain subregion at about day 80, with a secondary emission reaching around $150\,\mathrm{grains\,m^{-3}}$ on day 125. In the Northwest, DBF pollen has the earliest maximum (day 70) at about the same magnitude as California ($\sim 200\,\mathrm{grains\,m^{-3}}$). In some locations, there is a secondary DBF peak in the late summer and early fall due to the late flowering of *Ulmus crassifolia* and *Ulmus parvifolia*, located predominantly in the Southeast and California (Lewis, et al.1983). In the Southeast, this occurs between days 225 and 300, while in California this occurs twice around day 245 and day 265.

The two ENF families exhibit pollen release at two distinct but overlapping times, with Cupressaceae peaking before Pinaceae. Cupressaceae in the Southeast emits pollen earlier than in other subregions, with a maxima at just over $400\,\mathrm{grains\,m^{-3}}$ around day 10 and counts above $200\,\mathrm{grains\,m^{-3}}$ in December of the prior year. Cupressaceae dominates the total emissions for the Southeast, with a smaller maximum from Pinaceae of about $180\,\mathrm{grains\,m^{-3}}$ near day 110. In the Northeast, the bimodality of ENF is evident with the Cupressaceae family reaching a maximum of $100\,\mathrm{grains\,m^{-3}}$ near day 85 with a secondary Pinaceae maximum approximately 65 days later at about half the magnitude ($\sim 50\,\mathrm{grains\,m^{-3}}$). In the Mountain and Pacific Northwest subregions, the maximum occurs around days 50–80 and can reach up to $350\,\mathrm{grains\,m^{-3}}$ in the Mountain subregion, but in both subregions it is generally much lower than in the eastern United States (approximately $50\,\mathrm{grains\,m^{-3}}$). In the California subregion, ENF emissions are comparatively low ($< 50\,\mathrm{grains\,m^{-3}}$) which is likely due to the bias in sampling locations.

The grasses (Poaceae) have comparatively low average pollen counts ($< 25\,\mathrm{grains\,m^{-3}}$) throughout the season in all subregions except the Northwest, where the maximum reaches $75\,\mathrm{grains\,m^{-3}}$. However, the average maximum Poaceae pollen count at individual stations is close to $100\,\mathrm{grains\,m^{-3}}$, with the individual annual maxima reaching several hundreds of pollen grains $\mathrm{m^{-3}}$. In the AAAAI data, there are two distinct maxima in the Northeast Poaceae count, and we attribute the first seasonal maximum to C_3 grasses (peak around day 155) and the second grass maximum mainly to C_4 grasses (peak around day 250). Observations by Craine et al. (2012) of Poaceae in an American prairie have indicated that C_3 and C_4 grass flowering occurs at distinctly different times, with C_3 in the late spring and C_4 in mid-to-late summer. Similarly, Medek et al. (2016) observed two grass pollen peaks in Australia, with a stronger, late-summer peak at lower southern latitudes where there is higher incidence of C_4 grass. However, the authors note that sometimes this may be due to a second flowering of some C_3 grass species. Although the C_3–C_4 separation cannot be confirmed in the AAAAI pollen count data because they are not distinguished during pollen identification, this distinction is included in the model as discussed in Sects. 3.1 and 4.2 below. In the Southeast, this separation of the Poaceae pollen counts is less apparent because both of the emission maxima are broader and intersect one another. In the Southeast, the first observed pollen maximum (assessed as C_3 grass pollen) peaks earlier around day 140, while the second maximum (assessed as C_4 grasses) has a similar, yet smaller, value around day 250. In the Mountain subregion, the first grass maximum occurs later in the year (day 175), and the second grass maximum occurs around day 250 in the late summer. Pollen counts in California are only substantial during the earlier flowering time (C_3 grasses) and have a similar duration to those in the Northeast, peaking at around day 135. For the Pacific Northwest, there is one strong early peak of grass pollen in the middle of the summer (day 170) and a secondary maximum is negligible, although counts below $10\,\mathrm{grains\,m^{-3}}$ register around days 250–270.

Ragweed (*Ambrosia*) pollen is segregated from other grasses and herbs because of the strong allergic response in humans to this specific species and the unique timing of emissions. Because it is a short-day plant (i.e., its phenology driven by a shortening photoperiod and cold temperatures; Deen et al., 1998), ragweed pollen seasons are generally constrained to the late summer with the exception of the Mountain region where some counts occur in the spring. Emissions in the Northeast reach a maximum around day 240 at $60\,\mathrm{grains\,m^{-3}}$, while they occur slightly later in the Southeast, peaking around day 270 with twice the magnitude ($120\,\mathrm{grains\,m^{-3}}$). In the Mountain subregion, ragweed pollen has an expected peak at around day 245, but also an earlier peak at around day 130 with no confirmed cause. *Ambrosia* is not detected in the station averages for California and the Pacific Northwest, although some individual sites in these regions record relatively low counts on the order of $10\,\mathrm{grains\,m^{-3}}$.

3 Model input data

3.1 Land cover data

With a goal of developing regional to global pollen emissions, one of the greatest limitations is the description of vegetation at the appropriate taxonomic level and spatial resolution. While land cover databases specific to species level are available for some regions, they are not available globally. Alternatively, vegetation land cover in regional to global models can be represented by classifications based on biophysical characteristics. For climate models, a common approach to represent land cover is with PFTs, and global PFT data are readily available and used by many regional and global climate models to describe a variety of terrestrial emissions (Guenther et al., 2006) and biophysical processes in land–atmosphere exchange models. The creation
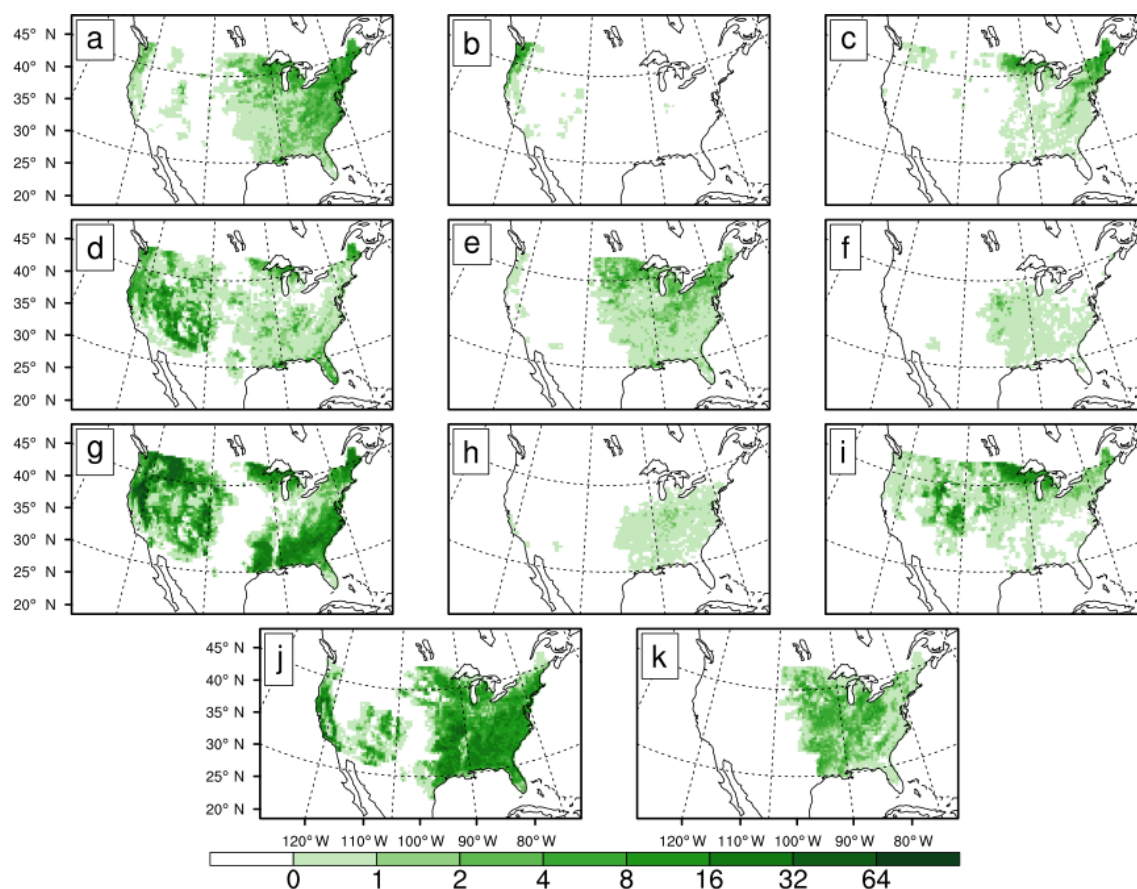
Figure 3. Land cover fraction (% coverage) for 11 tree taxa from the Biogenic Emissions Landuse Database version 3 (BELD) regridded to a 25 km resolution grid, including **(a)** *Acer* (maple), **(b)** *Alnus* (alder), **(c)** *Betula* (birch), **(d)** Cupressaceae (cedar/juniper), **(e)** *Fraxinus* (ash), **(f)** *Morus* (mulberry), **(g)** Pinaceae (pine), **(h)** *Platanus* (sycamore), **(i)** *Populus* (poplar/aspen), **(j)** *Quercus* (oak), and **(k)** *Ulmus* (elm).

of a pollen emissions model with PFT categorization would be of use at a broad range of spatial scales and domains while integrating more readily with climate models. In the pollen emissions model development and evaluation (Sects. 4 and 5), we compare two different vegetation descriptions of broadleaf deciduous and evergreen needleleaf trees including (1) family- or genus-specific land cover and (2) land cover categorized by PFT.

The Biogenic Emissions Landuse Database version 3 (BELD) provides vegetation species distributions at 1 km resolution over the continental United States based on satellite imagery, aerial photography, and ground surveys, as well as other land cover classification data such as geographical boundaries (Kinnee et al., 1997; https://www.epa.gov/air-emissions-modeling/biogenic-emissions-landuse-database-version-3-beld3). The BELD database includes 230 different tree, shrub, and crop taxa across the United States as a fraction of the grid cell area at either the genus or species level. For family- and genus-level pollen emissions, the BELD land cover fraction for the 11 dominant pollen-emitting tree taxa identified in

Sect. 2.1 is utilized (Table 1; Fig. 3). For species-level land cover data, land cover fraction is calculated as the aggregate of all species within a family or genus.

For the PFT land cover, we use the Community Land Model 4 (CLM4) (Oleson et al., 2010) surface dataset that employs a 0.05° resolution satellite-derived land cover fraction from the International Geosphere-Biosphere Programme (IGBP) classification (Lawrence and Chase 2007). We sum all three biome PFT categories (temperate, tropical, and boreal) for deciduous broadleaf forests (DBF) and two biome PFT categories (boreal and temperate) for evergreen needleleaf forests (ENF) to produce the model PFT land cover.

Figure 4a–d compares the BELD land cover (summed by PFT) and CLM4 land cover for the two tree PFTs. Region-by-region comparison of land cover for all BELD taxa and each tree PFT (from both BELD and CLM4) is given in Table 2. An important distinction is that CLM4 land cover extends beyond US borders because it is derived from a global dataset, whereas BELD is constrained to the continental United States. BELD and CLM4 land cover show general agreement on the regional distribution of both tree PFTs.

Table 1. Production factors (P) and phenological regression coefficients for the start day of year (sDOY) and end day of year (eDOY) as a function of temperature for the 13 individual pollen-producing taxa. Individual taxa and families are organized into the four PFTs, with the two aggregated tree PFTs denoted as DBF and ENF. Regression slope (days $°C^{-1}$) and coefficient of determination are provided for both sDOY and eDOY (slope/R^2).

Taxon or PFT	p_{annual} 10^7 grains m^{-2} $year^{-1}$	Reference for p_{annual}	sDOY (slope/R^2) days $°C^{-1}$	eDOY (slope/R^2) days $°C^{-1}$
		Deciduous broadleaf forest (DBF)		
Acer	89.1	Tormo Molina et al. (1996)	−1.78/0.15	−1.56/0.06
Alnus	210	Helbig et al. (2004)	−8.82/0.46	−4.88/0.26
Betula	140	Jato et al. (2007)	−3.46/0.54	−3.45/0.35
Fraxinus	45.1	Tormo Molina et al. (1996)	−4.69/0.50	−2.92/0.32
Morus	10	N/A	−4.00/0.53	−2.97/0.29
Platanus	121	Tormo Molina et al. (1996)	−4.47/0.40	−2.65/0.20
Populus	24.2	Tormo Molina et al. (1996)	−2.23/0.24	−0.31/< 0.01
Quercus	78	Tormo Molina et al. (1996)	−4.09/0.53	−2.03/0.19
Ulmus (early, late)	3.55	Tormo Molina et al. (1996)	−4.61/0.59, 3.06/0.12	−2.37/0.16, 5.12/0.29
DBF	80.1	This paper	−4.55/0.46	−1.94/0.13
		Evergreen needleleaf forest (ENF)		
Cupressaceae	363	Hidalgo et al. (1999)	−5.67/0.48	−2.67/0.17
Pinaceae	22.2	Tormo Molina et al. (1996)	−5.72/0.45	−5.03/0.41
ENF	193	This paper	−5.95/0.40	−4.96/0.33
		Grasses (GRA)		
Poaceae (C_3, C_4)	8.5, 0.85	Prieto-Baena et al. (2003)	−4.76/0.48, 0.05/< 0.01	−1.08/0.04, 2.96/0.32
		Ragweed (RAG)		
Ambrosia	119*	Fumanal et al. (2007)	1.08/0.08	3.42/0.37

* *Ambrosia* production factor in 10^7 grains plant^{-1}.

Table 2. Total spatial coverage (km^2) of tree taxa and PFTs from BELD and CLM4 land cover datasets in the five US subregions (Northeast, NE; Southeast, SE; Mountain, MT; California, CA; Pacific Northwest, PNW). All individual tree taxa are from the BELD database. BELD DBF and ENF land cover are the sums of the land cover of the taxa belonging to each PFT.

Land cover class	NE	SE	MT	CA	PNW
Acer	6.79E+04	2.88E+04	1.89E+03	1.97E+02	3.09E+03
Alnus	3.37E+00	1.23E-01	6.49E+01	1.71E+02	9.56E+03
Betula	2.99E+04	2.68E+03	2.78E+02	2.64E+00	4.82E+02
Fraxinus	3.96E+04	1.10E+04	3.14E+03	3.94E+01	2.76E+02
Morus	3.99E+03	2.25E+03	3.89E+01	0.00E+00	0.00E+00
Platanus	3.18E+03	3.38E+03	1.33E+01	1.44E+02	0.00E+00
Populus	5.48E+04	1.23E+03	4.37E+04	1.96E+02	1.55E+03
Quercus	1.30E+05	2.25E+05	2.51E+04	2.82E+04	1.40E+04
Ulmus	4.96E+04	2.81E+04	1.37E+03	0.00E+00	0.00E+00
BELD DBF	3.79E+05	3.03E+05	7.56E+04	2.90E+04	2.90E+04
CLM4 DBF	6.67E+05	4.03E+05	1.72E+05	7.93E+03	4.18E+04
Cupressaceae	1.85E+04	2.11E+04	7.84E+04	9.64E+03	2.35E+04
Pinaceae	8.34E+04	1.58E+05	1.79E+05	2.95E+04	1.10E+05
BELD ENF	1.02E+05	1.79E+05	2.58E+05	3.91E+04	1.34E+05
CLM4 ENF	1.44E+06	4.26E+05	4.66E+05	4.57E+04	5.34E+05
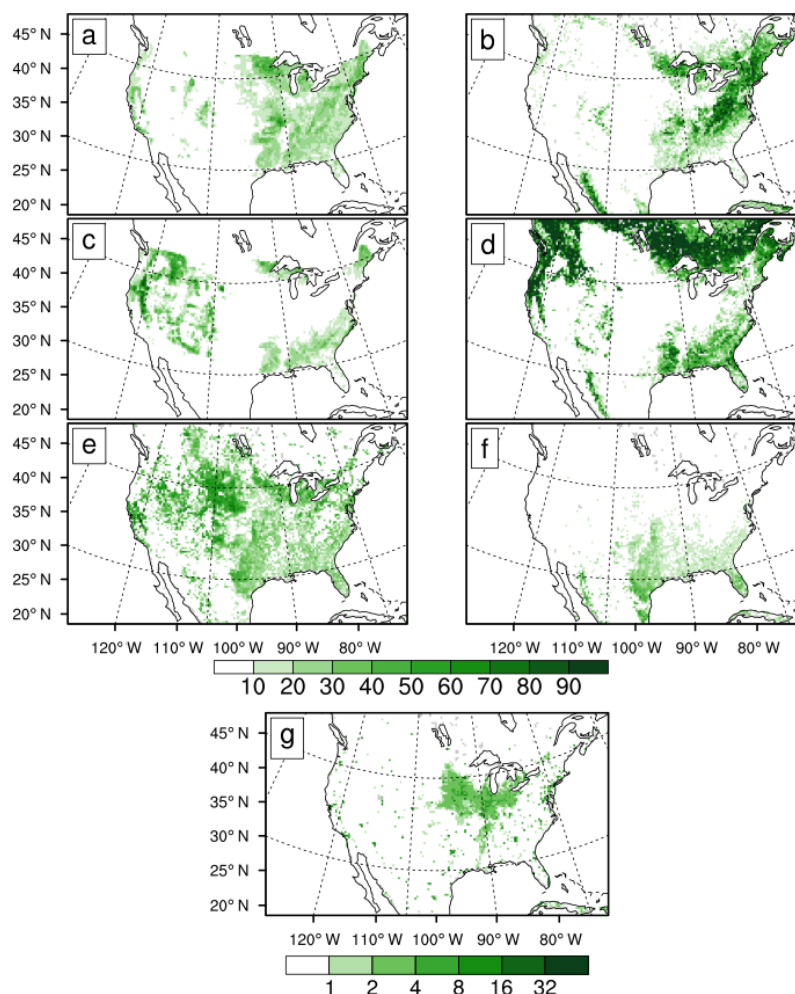
Figure 4. BELD (**a, c**) and CLM4 (**b, d, e, f**) land cover for the four PFT categories that produce pollen emissions, including (1) deciduous broadleaf forest for (**a**) BELD and (**b**) CLM4, (2) evergreen needleleaf forest for (**c**) BELD and (**d**) CLM4, (3) grasses, including (**e**) C_3 grasses and (**f**) C_4 grasses, and (**g**) ragweed, represented by crop and urban CLM4 categories.

DBF is predominantly in the eastern portion of the United States with a gap in the midwestern corn belt. ENF is present in the Southeast, the Northeast along the US–Canadian border, along the Cascade and Coastal mountain ranges, and throughout the northern Rockies. A notable difference is the CLM4 representation of ENF, which shows a strong, dense band extending from the Sierra Nevada through the Canadian Rockies. The BELD ENF broadly covers the Rocky Mountain range, yet more diffusely (land cover percentage up to 76 %), whereas the CLM4 dataset shows sparser and dense ENF land cover (e.g., up to 100 %) in the same range. For the DBF category, another notable difference is that the strong band of oaks around the Central Valley of California, which is evident in BELD but missing from the CLM4 dataset. Additionally, the CLM4 has far greater densities of DBF along the Appalachian range than BELD. Overall, the CLM4 land cover fractions for forest PFTs are higher on average than the summed BELD taxa, about 2 to 10 times as much in each re-

gion, with the exception of California subregion DBF where CLM4 land cover is about half of that in the BELD dataset (Table 2).

Grass spatial distributions are given by C_3 (non-arctic) and C_4 grass PFT land cover classes from CLM4 (Fig. 4e, f), which correspond to the observed family-level Poaceae pollen subdivided into C_3 and C_4 categories (described in Sect. 2.2). C_3 coverage is evident across the United States, with broad coverage throughout the Southeast, Midwest, and Northern Great Plains (Fig. 4e). C_4 coverage is concentrated in the Southeast and Southern Great Plains at lower densities (Fig. 4f).

Ragweed requires a different land cover treatment, as land cover distributions are not available for ragweed across the entire continental United States. Ragweed is known to arise in areas of human disturbances (Forman and Alexander 1998; Larson 2003) and is found mainly in disturbed or developed areas such as cities and farms (Katz et al., 2014; Clay

et al., 2006). *Ambrosia* land cover (Fig. 4g) is derived from the urban and crop categories of the CLM4 land cover, which are sourced from LandScan 2004 (Jackson et al., 2010) and the CLM4 datasets, respectively. The urban data are subdivided by urban intensity, which is determined by population density. We assume that ragweed is unlikely to grow in the densest of urban areas (such as city centers) and utilize the lowest urban density category that is also the most widespread. Ragweed land cover (plants m^{-2}) in urban areas is determined by multiplying the average urban ragweed stem density given by Katz et al. (2014) by the urban land cover fraction. For crops, the CLM4 subdivides land cover fraction into categories including corn and soybean crops, and Clay et al. (2006) provide ragweed stem densities in soybean and corn cropland. Thus, we calculate the ragweed land cover in stems m^{-2} (f_{rag}):

$$f_{\text{rag}} = \alpha \left(d_{\text{soy}} f_{\text{soy}} + d_{\text{corn}} f_{\text{corn}} \right) + \beta \left(d_{\text{urb}} f_{\text{urb}} \right), \qquad (1)$$

where d_{soy}, d_{corn}, and d_{urb} represent the stem density (stems m^{-2}) of ragweed in soybean, corn, and urban areas, respectively, and the f$_{\text{soy}}$, f$_{\text{corn}}$, and f$_{\text{urb}}$ represent the fractional land cover for soybean, corn, and urban, respectively. α and β are tuning parameters to that are determined by a preliminary evaluation between modeled and observed ragweed pollen counts, where $\alpha = 0.01$ for crop and $\beta = 0.1$. Zink et al. (2017) show that a ragweed land cover representation developed by combining land use and local pollen count information evaluates better against observed pollen counts than even ragweed ecological models, giving confidence to this choice of land cover representation.

All land cover data are regridded to a 25 km resolution across the United States to provide emissions at the same spatial resolution as the regional climate model (see Sect. 5).

3.2 Meteorological data for phenology

To develop the emissions model, we use two sources of meteorological data. The first is a high-resolution meteorological dataset to develop the phenological relationships for the timing of pollen release. Because reliable measurements are not available at all pollen count stations and there is uncertainty in the siting of these stations (e.g., they may be in urban areas with highly heterogeneous temperature), we use a gridded observational meteorological product for consistency across all sites (Maurer et al., 2002). The gridded Maurer dataset interpolates station data to a 1/8° grid across the continental United States on a daily basis, representing a high spatial resolution gridded data product where data from each meteorological station have been subject to consistent quality control. Higher-resolution DayMet temperatures (daily 1 km) (Thornton et al., 2014) were used in lieu of Maurer data at NAB sites where the Maurer dataset did not provide information at the collocated grid cell (Table S1). For offline emission calculations input into the regional climate model, we use annual-average temperatures computed from monthly

Climate Research Unit (CRU) temperature data (Harris et al., 2014). These data were interpolated from a 0.5° × 0.5° grid to the 25 km regional climate model grid used for pollen transport.

4 PECM model description

4.1 Emission potential

The pollen emissions model is a prognostic description of the potential emissions flux of pollen (E_{pot}; grains m^{-2} d^{-1}) for an individual taxon i:

$$E_{\text{pot},i}(x, y, t) = f_i(x, y) \frac{p_{\text{annual},i}}{\int_0^{365} \gamma_{\text{phen},i}(x, y, t)\, \text{d}t} \gamma_{\text{phen},i}(x, y, t), \quad (2)$$

for a model grid cell of location x and y at time t. In this expression, $f(x, y)$ is the vegetation land cover fraction (Sect. 2.1; m^2 vegetated m^{-2} total area), p_{annual} is the daily production factor (grains m^{-2} year^{-1}), and γ_{phen} is the phenological evolution of pollen emissions that controls the release of pollen (description below). Equation (2) can apply to either a single taxa or PFT, depending on the prescription of land cover through $f(x, y)$. In the simulations described here, emissions are calculated offline based on this equation and provided as input to a RCM. This emission potential is later adjusted based on meteorological factors in the RCM where the pollen grains are transported as aerosol tracers (Sect. 5.1.1). In the future, Eq. (2) can be coupled directly within the climate model for online calculation of emissions. The phenological and production factors are described in greater detail below.

4.2 Phenological factor (γ_{phen})

Based on the observed pollen counts, a Gaussian distribution is used to model the phenological timing of pollen release (γ_{phen}):

$$\gamma_{\text{phen},i}(x, y, t) = e^{-\frac{(t - \mu(x,y))^2}{2\sigma(x,y)^2}}, \qquad (3)$$

where $\mu(x, y)$ and $\sigma(x, y)$ are the mean and half width of the Gaussian, respectively, and can be determined based on the start day of year (sDOY) and end day of year (eDOY) calculated by an empirical phenological model:

$$\mu(x, y) = \frac{\text{sDOY}(x, y) + \text{eDOY}(x, y)}{2} \qquad (4)$$

$$\sigma(x, y) = \frac{\text{eDOY}(x, y) - \text{sDOY}(x, y)}{a}. \qquad (5)$$

The fit parameter, a, accounts for the conversion between the empirical phenological dates based on a pollen count threshold and the equivalent width of the emissions curve. Based on evaluation versus observations, $a = 3$ was selected for initial offline simulations.

Linear regressions of observed sDOY and eDOY from individual pollen count stations versus temperature are used to empirically determine sDOY and eDOY that drive γ_{phen}. An important criterion is the grain count used to determine the sDOY and eDOY, and we utilize a count threshold adaptable to bimodal emission patterns such as those noted for *Ulmus* and Poaceae. Sofiev et al. (2013) selected dates on which the 5th and 95th percentiles of the annual index (annual sum of pollen counts) were reached, while Liu et al. (2016) combined a 5 grains m^{-3} threshold with the additional condition that 2.5 % (97.5 %) of the annual sum of pollen was reached before the start (end) date. Here, we implement a pollen count threshold of 5 grains m^{-3} and found this was sufficient to reproduce the observed seasonal cycle. To account for smaller signals that may be due to count errors (e.g., an exceedance of the 5 grains m^{-3} threshold but not followed by an increase in emissions), we used a moving window with a threshold of 25 grains m^{-3} for the sum of pollen counts in the nearest 10 neighboring days; when the sum of the neighbors failed to meet this threshold, the data point was omitted. In this manner, we calculated the sDOY and eDOY for the full 8-year time series for each taxon at each station. If more than one start or end date was found in a single year at a single station for a taxon that was not clearly bimodal, only the first set of dates was retained for the linear regression. For taxa with an observed bimodal peak, the second peak was treated as a separate taxon (e.g., early and late *Ulmus*, C$_3$ and C$_4$ Poaceae) with a separate phenology. Once the sDOY and eDOY were determined, outliers in these dates were determined by bounding the data for each taxon at 4 times the mean absolute deviation of sDOY and eDOY.

Near-surface atmospheric temperature (e.g., 2 m height) is an important factor of vegetation phenology. In the interest of having a regional model of emissions that prognostically calculates the start dates, the previous-year annual average temperature (PYAAT) based on near-surface atmospheric temperature from Maurer et al. (2002) and Thornton et al. (2014) (Sect. 2.2) is the explanatory variable in the linear regressions. For example, for a start date of 2 February 2007, the PYAAT would be the mean temperature for the year 2006. For *Pinus* and Cupressaceae, PYAAT is calculated differently from 1 July 2005 to 30 June 2006 because emissions of these families begin in the early winter (December). Prior studies have shown that the meteorology of the year previous to the pollen season influences pollen production, especially temperature, suggesting that PYAAT may be a good predictor variable (Menzel and Jochner, 2016). While emissions in this study are calculated using offline meteorological data, this also could be coupled to a dynamic land surface model to predict reasonably accurate pollen phenological dates.

To exemplify this method, Fig. 5 shows the phenological dates and regression lines for the *Betula* (birch) genus, with all 13 modeled taxon shown in Figs. S1 and S2 in the Supplement. The sDOY and eDOY of the pollen season show a moderate and considerable trend with temperature for most
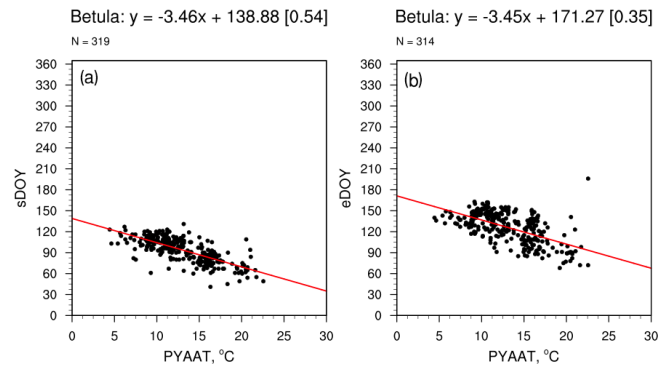

Figure 5. Phenological regressions for *Betula* (birch) pollen for **(a)** sDOY and **(b)** eDOY versus previous-year annual average temperature (PYAAT; °C). Each point signifies one station per year for pollen count data from 2003 to 2010 (total denoted as N).

taxa and PFTs (Table 1; Figs. S1, S2). The linear regression models for sDOY explain 41 % of the variance on average for DBF taxa, 47 % on average for ENF taxa, 48 % for C$_3$ Poaceae, and 8 % for *Ambrosia*, while having a negligible R^2 for C$_4$ Poaceae. For eDOY, the linear regression models explain 21 % of the variance on average for DBF taxa, 29 % for ENF, 4 % for C$_3$ Poaceae, 32 % for C$_4$ Poaceae, and 37 % for *Ambrosia*. All trends except C$_4$ Poaceae, late elm, and *Ambrosia* are negative, indicating that warmer previous-year temperatures result in earlier start and end dates. For most tree taxa, the trend of both sDOY and eDOY are negatively correlated with PYAAT, with a steeper negative slope for sDOY. The correlation for the duration of the pollen season (eDOY–sDOY) is then positive for all taxa except Cupressaceae. This suggests that warmer climates have earlier pollen season start and end dates but longer season lengths.

Trends for grass in Australasia show that the correlation of the end date of the pollen season with average spring temperature is positive, while the same relationship for the start date is negative, suggesting also that season start dates are earlier and season duration increases with warmer climates (Medek et al., 2016). The apparent trend in the season end date for *Ambrosia* with PYAAT could be due to the increased number of frost-free days, consistent with global warming, and a strong relationship between frost-free days and changes of ragweed season length (Easterling 2002; Ziska et al., 2011).

This agrees with earlier findings that suggest the pollen season will, on average, start earlier with a warmer global climate and have a longer duration (Confalonieri et al., 2007). The spatiotemporal heterogeneity of climate change may affect which regions and seasons will be most influenced by climate change (Ziska 2016). In fact, there is imperfect agreement that earlier start dates and longer seasons will occur unanimously throughout the United States region, at least for trees (Yue et al., 2015). It is understood that photoperiod and the dormancy-breaking process controlled by chilling temperatures play a significant role in the phenol-
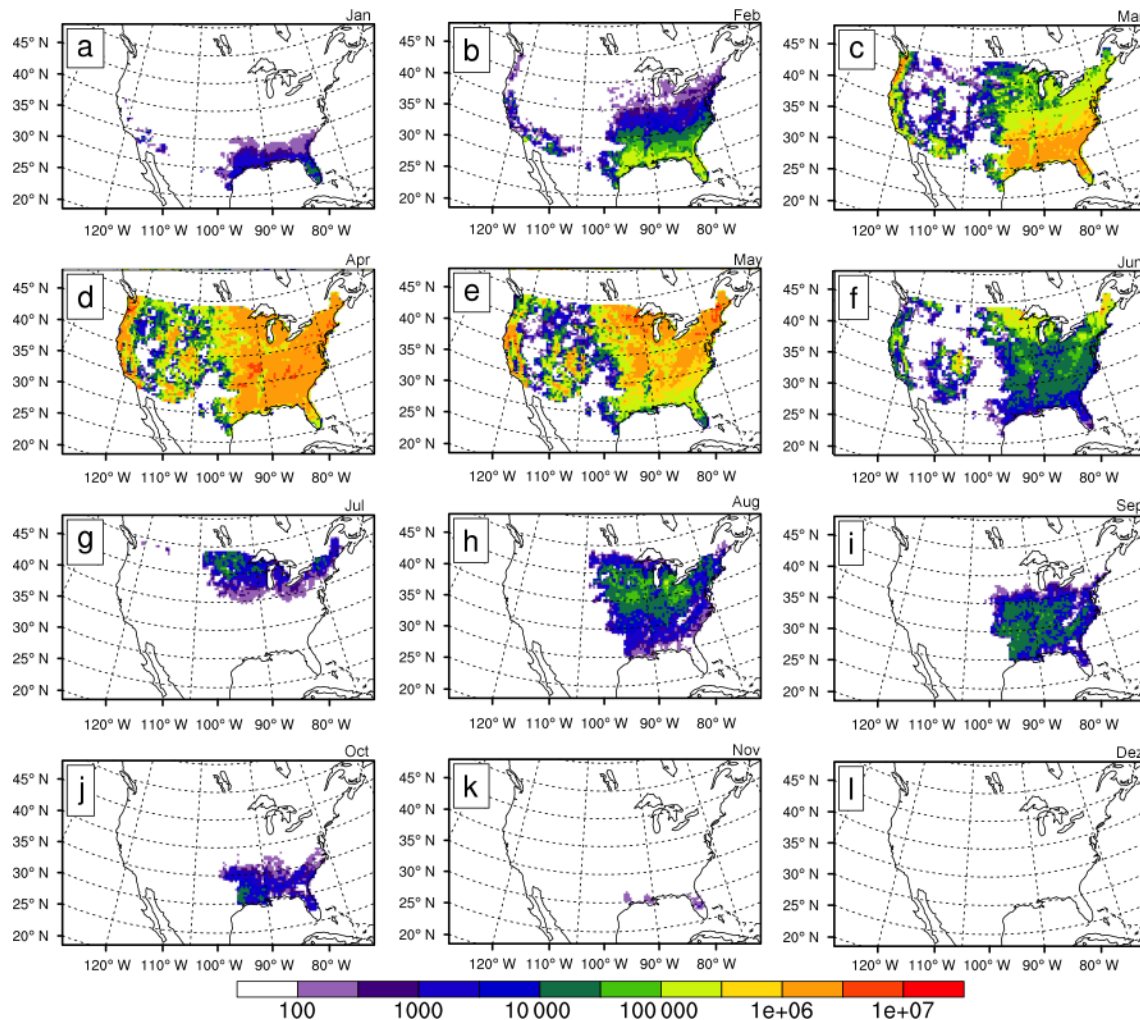
Figure 6. Monthly average emissions potential (E; Eq. 1) for BELD model DBF (2003–2010) in grains m^{-2} day^{-1}. **(a)** January, **(b)** February, **(c)** March, **(d)** April, **(e)** May, **(f)** June, **(g)** July, **(h)** August, **(i)** September, **(j)** October, **(k)** November, and **(l)** December.

ogy of trees (Myking and Heide 1995; Ziska 2016), and it is generally accepted that a plethora of other factors, such as plant age, mortality, and nutrient availability, also affect observed phenological dates (Jochner et al., 2013). However, even without these factors, the current phenological model is applicable to large regions and provides a clear response of plants to interannual climate variability as well as long-term climate changes. For this first assessment of PECM, we assume that the pollen production factor (p_{annual}) does not change with time and that the phenological model described above captures the main features of pollen emissions.

4.3 Annual pollen production (p_{annual})

Annual production factors (grains m^{-2} $year^{-1}$, where m^{-2} refers to vegetated area, or grains $stem^{-1}$ $year^{-1}$ for ragweed) for each modeled taxon are provided in Table 1. The annual pollen production factor (p_{annual}) defines the amount of pollen produced per vegetation biomass per year based

on literature values. Tormo Molina et al. (1996) report the annual pollen productivity in grains $tree^{-1}$ $year^{-1}$ measured from three representative trees from several taxa. *Morus* has no known reference for production factor and was assumed to be 10×10^7 grains m^{-2} $year^{-1}$, conservatively at the low end of the range for other deciduous broadleaf taxa. Other tree taxa and grasses are reported in grains m^{-2} $year^{-1}$, while ragweed is reported in grains $stem^{-1}$ $year^{-1}$ (Helbig et al., 2004; Jato et al., 2007; Hidalgo et al., 1999; Prieto-Baena et al., 2003; Fumanal et al., 2007). To convert the production factors from Tormo Molina et al. (1996) (grains $tree^{-1}$ $year^{-1}$), the production factors for each representative tree are multiplied by the tree crown area, calculated as the circular area of the tree crown diameter given in Table II of Tormo Molina et al. (1996). The resulting individual production factors (grains m^{-2} $year^{-1}$) are then averaged for each taxa.
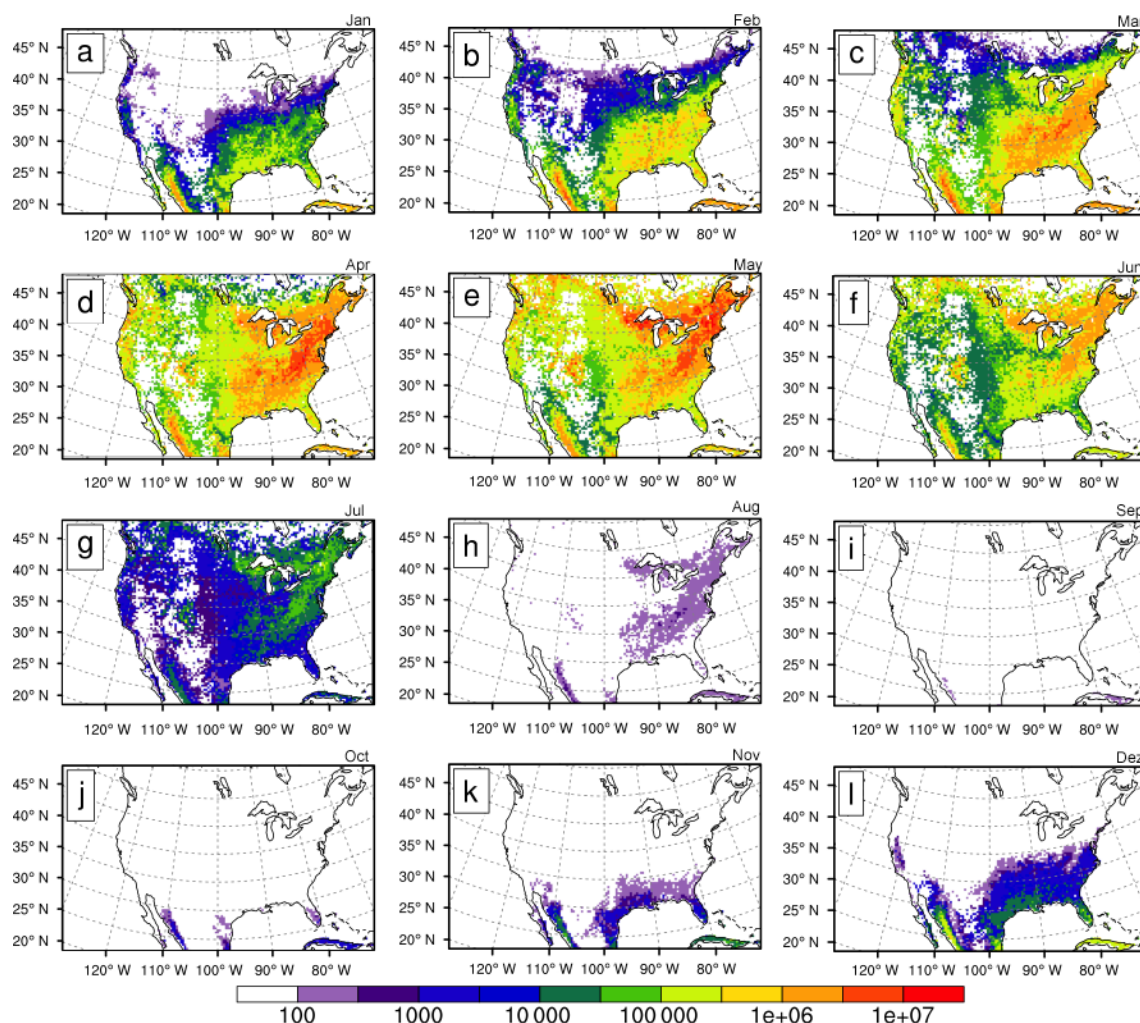
Figure 7. Same as Fig. 6 but for PFT model DBF.

After sensitivity experiments of running pollen emissions in RegCM4, we find that the literature value of p_{annual} for Poaceae provides better agreement with observations for C_4 grass when reduced by a factor of 10; thus, we use this value. To obtain the coefficient of daily pollen production over the duration of the phenological curve, γ_{phen}, the integral of the daily pollen production is normalized to p_{annual} as demonstrated by Eq. (2).

4.4 Offline emission simulations

We calculate emissions offline for two versions of PECM that differ in the land cover input data for woody plants. The first uses the detailed BELD tree database (Fig. 3) for tree pollen emissions (hereinafter, the "BELD" simulation), and the second uses globally based PFT data for tree pollen emissions (Fig. 4b, d) (hereinafter, the "PFT" simulation). For the grass and ragweed taxa, the emissions calculations are identical between the two simulations as the input land

cover is the same for these two categories. While the family and genus levels are useful for the allergen community, the respective taxon land cover databases needed to develop a global, adaptable model are not always available. While many plant traits are found to vary quite strongly within individual PFTs (Reichstein et al., 2014), the PFT convention is accepted and remains in use in climate models, particularly because of the lack of species-level land cover data at large scales. For the PFT version, pollen counts from individual taxa were summed within each PFT prior to calculation of the phenological regression (Table 1). We exclude the bimodality in *Ulmus* for the PFT version because it is the only tree taxon that exhibits this behavior, and late *Ulmus* pollen emissions are relatively small compared to the major DBF season. The production factors for each PFT are calculated as the unweighted average of the production factors for all the taxa within the PFT (Table 1).

Figures 6–11 show the monthly averages of the 2003–2010 emissions potential calculated by the offline models
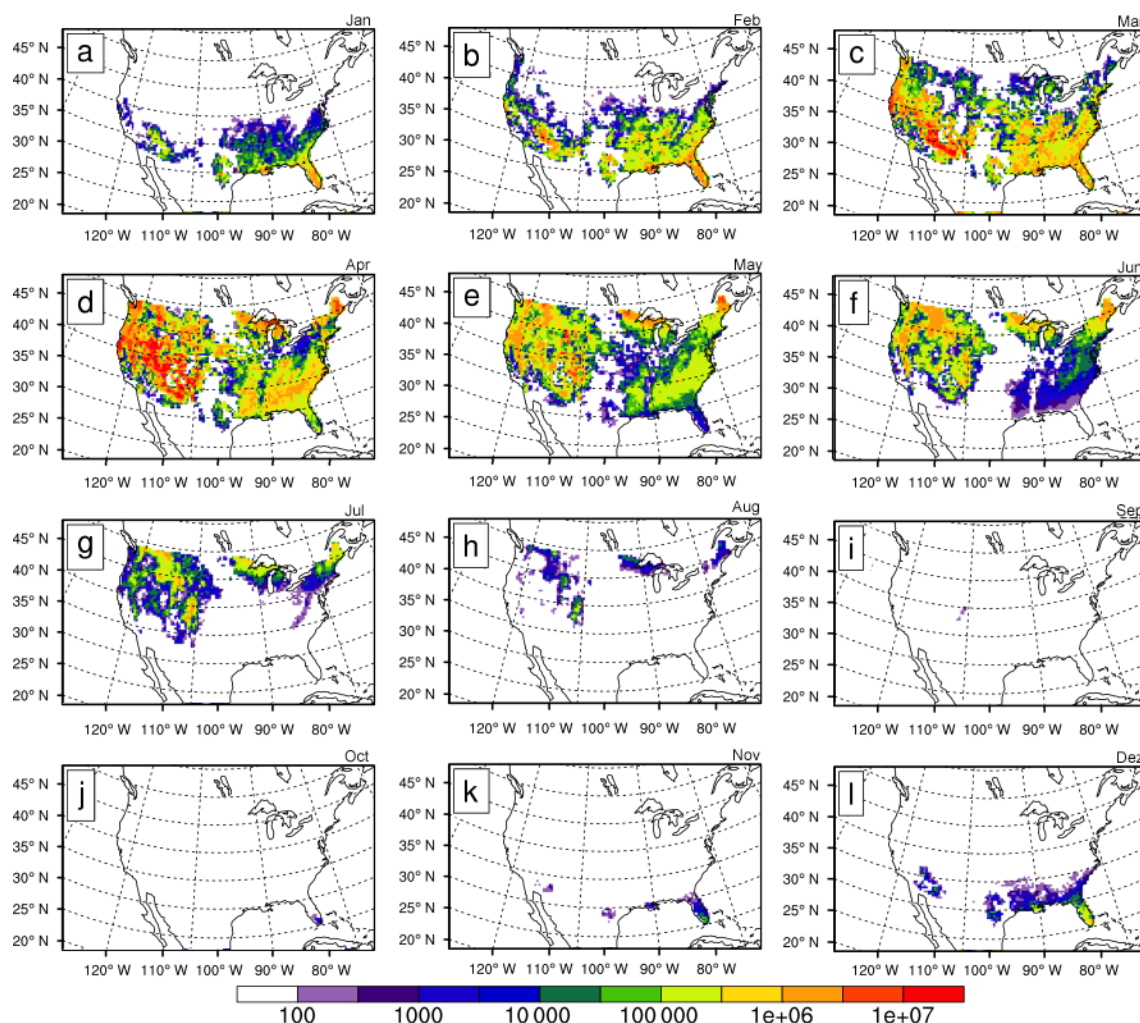
Figure 8. Same as Fig. 6 but for BELD model ENF.

described in Sect. 4.1 (E_{pot}; Eq. 2). The seasonal cycle can be clearly identified in the emissions' potential, with the onset of pollen emissions beginning in the warmer south and moving northward along the gradient of annual average temperature. Colder locales such as those at high elevations can interrupt this general trend. Though pollen seasons generally end later in the colder parts of the domain just as they start later, modeled pollen emission seasons tend to be shorter at colder locations for most taxa (about 1 day per 1 °C, on average). The highest maximum emissions for DBF occur over the Appalachian range between April and May for both the BELD and PFT versions (Figs. 6, 7). For ENF, the maximum occurs in April in the American west for the BELD version where Cupressaceae land cover is dominant, while it is consistent in magnitude between the Southeast and west coast for the PFT-based version (Figs. 8, 9). The grass PFT maximum emissions occur in June in the northern Rockies for C_3 and in September in the south-central Great Plains for C_4 (Fig. 10). Ragweed pollen emissions reach their maximum

during September throughout the Corn Belt where soybean and corn crops dominate the land surface, with local maxima apparent in urban centers (Fig. 11).

5 Emissions' implementation and evaluation

5.1 Emissions' implementation in a regional climate model

To evaluate PECM, emissions calculated offline are included within a regional climate model to compare simulated atmospheric pollen concentrations with ground-based observations from the NAB pollen network. The two phenological pollen emissions estimates (BELD and PFT) described above are prescribed as daily emissions, after which they are scaled by meteorological factors and undergo atmospheric transport. We use the Regional Climate Model version 4 (Giorgi et al., 2012), which is a limited-area climate model that includes a coupled aerosol tracer module (Solmon et al., 2006)
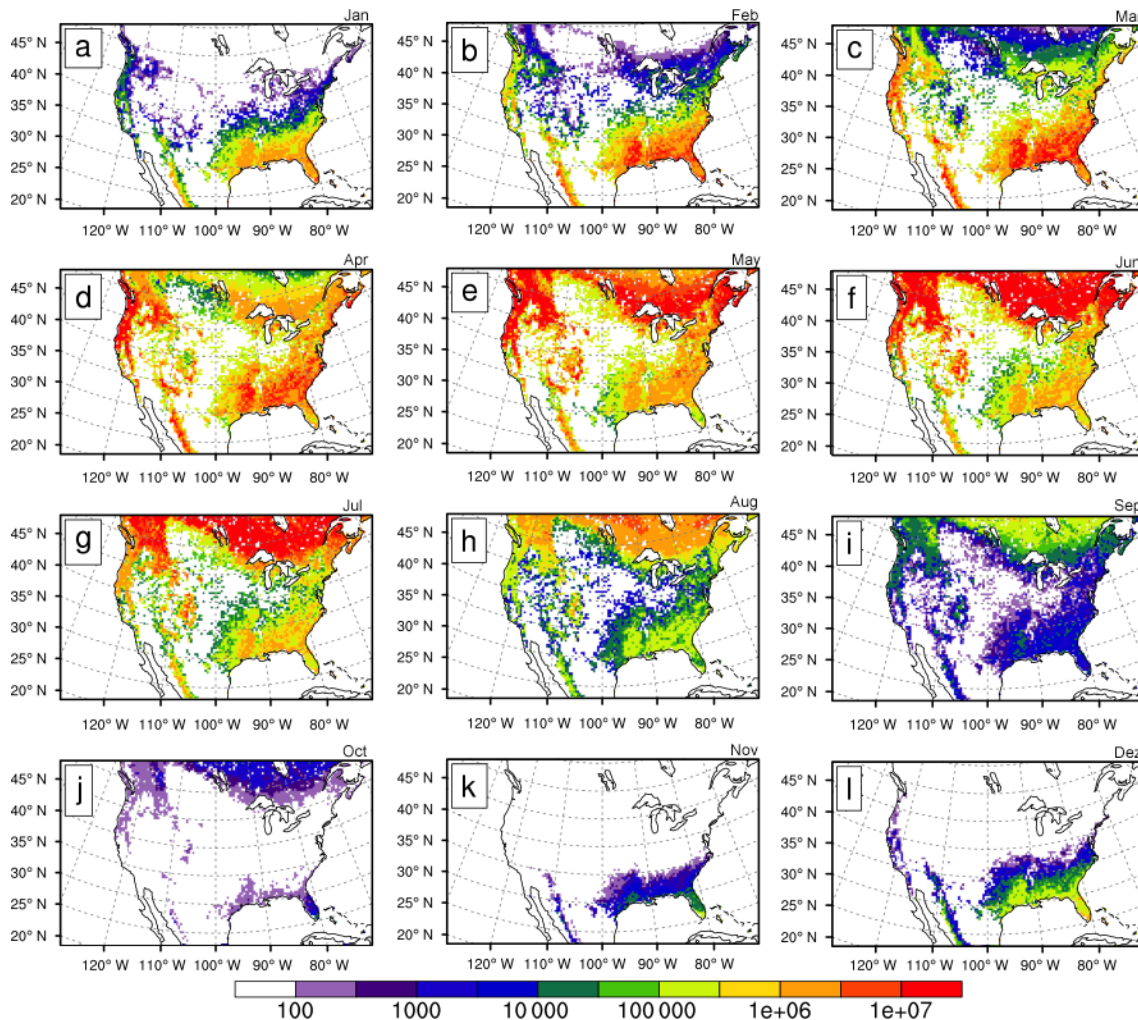
Figure 9. Same as Fig. 6 but for PFT model ENF.

that readily accommodates pollen tracers (Liu et al., 2016). The pollen tracer transport scheme is extended from one to four bins in this study to simulate the four PFTs (DBF, ENF, GRA, and RAG), with tracer bin particle effective diameters of 28, 40, 35 and 20 µm, respectively. Additionally, the temporal emissions input is updated to accommodate daily pollen emissions (grains m^{-2} day^{-1}).

RegCM4 is based on the hydrostatic version of the Penn State/NCAR mesoscale model MM5 (Grell et al., 1994) and configured for long-term climate simulations. In our RegCM4 configuration, we use the Community Land Model version 4.5 (CLM4.5; Oleson et al., 2010), the Emanuel cumulus precipitation scheme over land and ocean (Emanuel, 1991), and the SUBEX resolvable scale precipitation (Pal et al., 2000). The horizontal resolution is 25 km with 144×243 grid cells on a Lambert conformal projection, centered on 39° N, 100° W, with parallels at 30 and 60° N (Fig. 1). The vertical resolution includes 18 vertical sigma levels. Boundary conditions are driven by ERA-Interim reanalysis while

sea surface temperatures are prescribed from NOAA optimum interpolation SSTs (Dee et al., 2011; Smith et al., 2008). Two 8-year simulations of pollen emissions and transport in RegCM4 were conducted from 2003 to 2010 with the BELD and PFT versions of the offline emissions model. A total of 6 months of spin-up (July–December 2002) are run for both simulations that we exclude from the following analysis.

In the model, we calculate the fate of four pollen tracers corresponding to the four PFTs (DBF, ENF, GRA, and RAG) from the PECM offline emissions. Because individual tracers add to the computational cost of the simulations, BELD-based tree emissions are summed into DBF and ENF PFTs before they are emitted into the model atmosphere. To calculate the emissions, the emission potential calculated offline for each PFT (E_{pot}) is scaled according to surface meteorology following the methods of Sofiev et al. (2013):

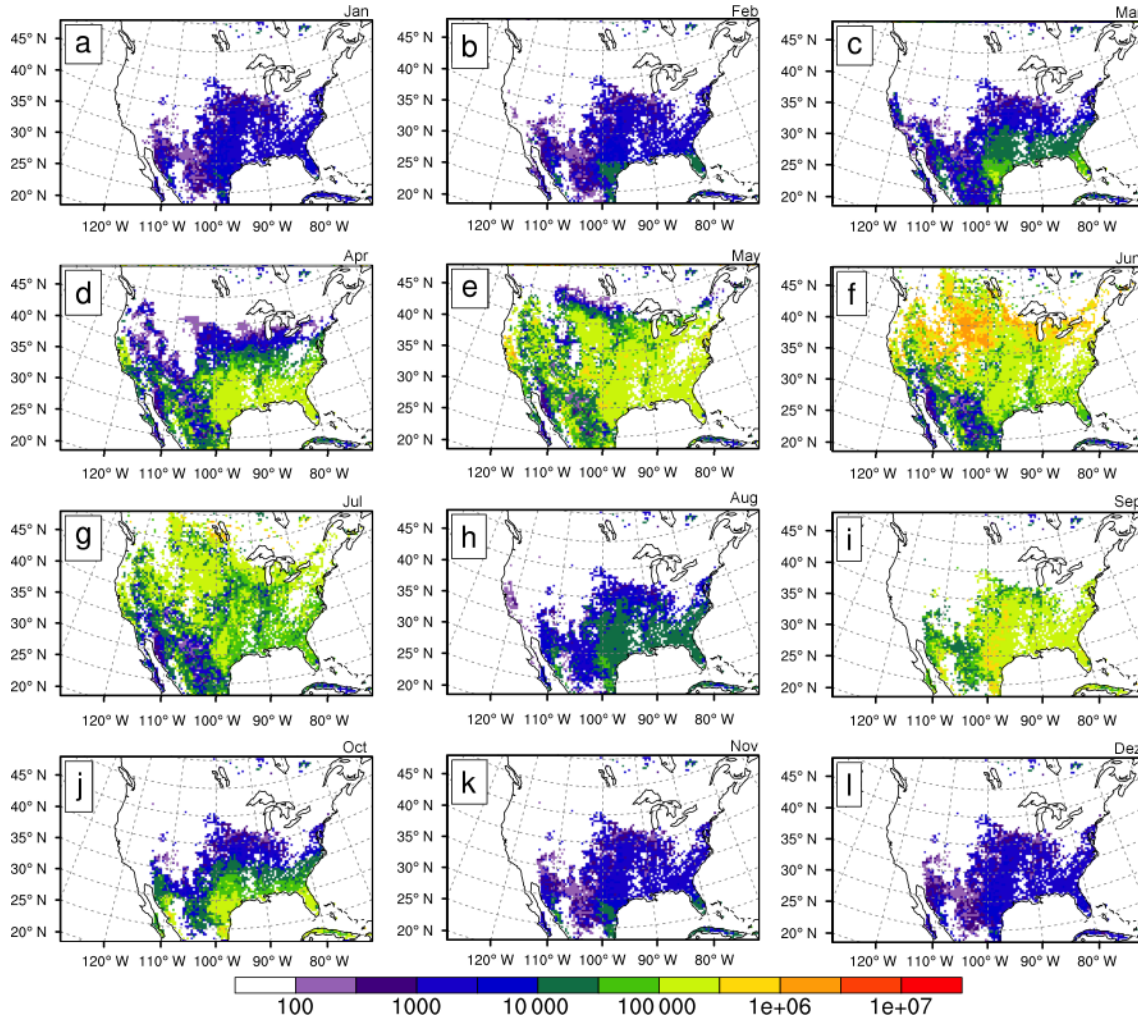$$E_{\text{pollen},i}(x, y, t) = E_{\text{pot},i}(x, y, t)\, f_w f_r f_h \qquad (6)$$

Figure 10. Same as Fig. 6 but for C_3 and C_4 grass.

$$f_w = 1.5 - e^{-\frac{u_{10}+u_{conv}}{5}} \qquad (7)$$

$$f_r = \begin{cases} 1, \, pr < pr_{low} \\ \dfrac{pr_{high} - pr}{pr_{high} - pr_{low}}, \, pr_{low} < pr < pr_{high} \\ 0, \, pr > pr_{high} \end{cases} \qquad (8)$$

$$f_h = \begin{cases} 1, \, rh < rh_{low} \\ \dfrac{rh_{high} - rh}{rh_{high} - rh_{low}}, \, rh_{low} < rh < rh_{high} \, , \\ 0, \, rh > rh_{high} \end{cases} \qquad (9)$$

where f_w, f_r, and f_h are the wind, precipitation, and humidity factors, respectively. The meteorological parameters in these equations are from online RegCM4 variables, including u_{10} and u_{conv} as the 10 m horizontal wind speed and vertical wind speed, and pr and rh are precipitation and relative humidity with low and high thresholds. These scaling factors account for the effects of wind, precipitation, and humidity on the emission of pollen from flowers and cones. The humidity and precipitation factors are piecewise linear func-

tions of the near-surface (10 m) RH and total precipitation and range from 0 (high precipitation or humidity) to 1 (no precipitation or low humidity). The wind factor ranges from 0.5 to 1.5, as even in calm conditions turbulent motions can trigger pollen release with high winds releasing more pollen. These scaled emissions are then transported according to the tracer transport equation (Eq. 10) of Solmon et al. (2006) that includes advection, horizontal and vertical diffusion (F_H and F_V), and convective transport (T_c), as well as wet (R_{Wls} and R_{Wc}, representing large-scale and convective precipitation removal) and dry deposition (D_d) of an individual tracer (χ), represented by $i = 1$ to 4 for each PFT pollen emission:

$$\frac{\partial \chi^i}{\partial t} = \bar{V} \cdot \nabla \chi^i + F_H^i + F_V^i + T_C^i + S^i - R_{Wls}^i - R_{Wc}^i - D_d^i. \qquad (10)$$
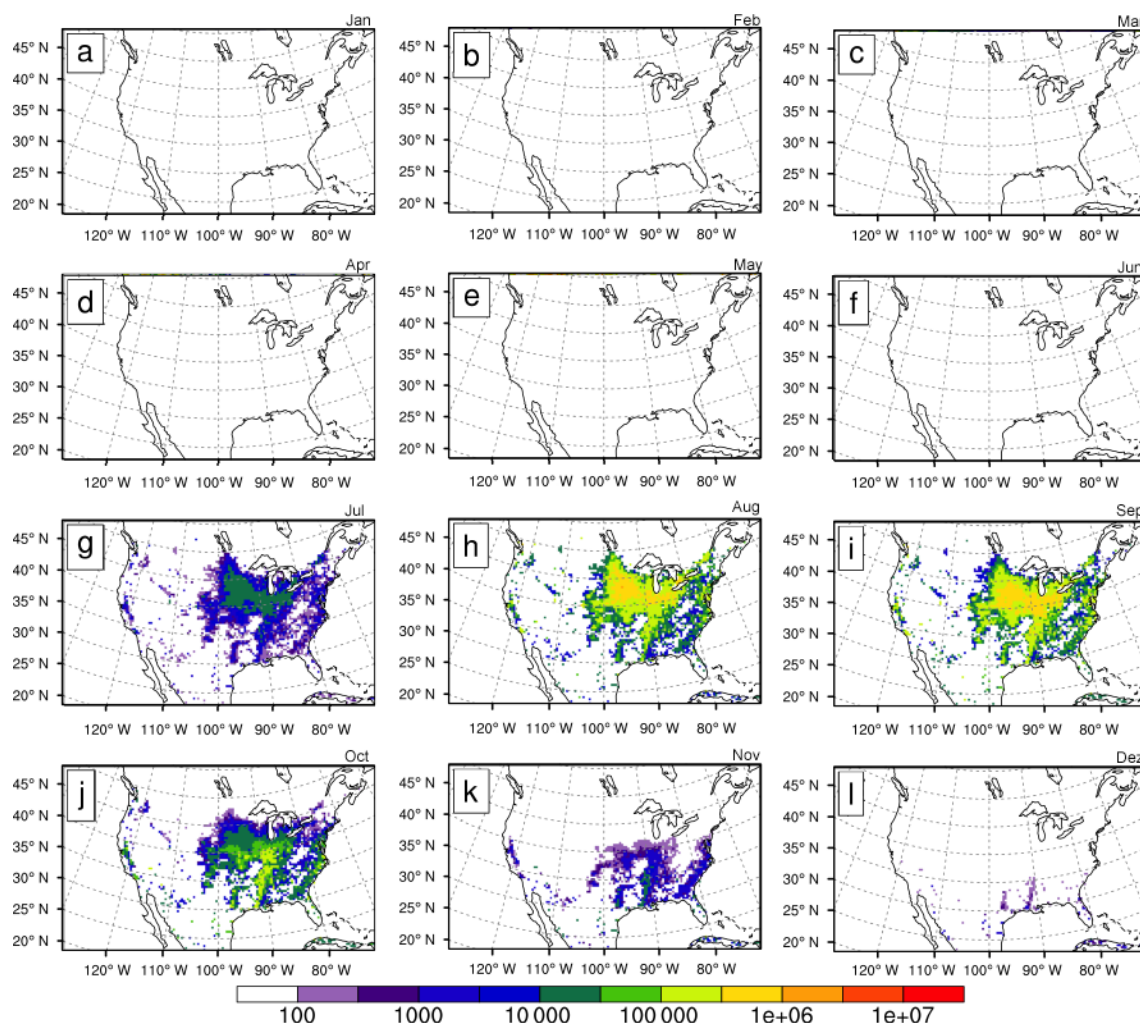
Figure 11. Same as Fig. 6 but for ragweed.

5.2 Model evaluation against observations

We evaluate the efficacy of PECM in simulating the tim-
ing and magnitude of pollen emissions across the continental
United States by evaluating RegCM4 tracer concentrations
versus observations. We compare the average daily simu-
lated near-surface pollen counts and observed, ground-based
pollen counts for each of the four modeled PFTs (Fig. 12).
The observed pollen time series in Fig. 12 are the spatial av-
erages of the average daily pollen counts at all pollen count-
ing stations comprising each of the five major US subregions
(Sect. 2.2) and are compared with the modeled average daily
pollen counts, which averages the individual grid cells that
contain the pollen counting stations. Interannual variability is
assessed using the relative mean absolute deviation for each
day of the average time series. The interannual variability in
observed daily pollen counts throughout the year is, on av-
erage, 81, 78, 78, and 77 % of the mean (DBF, ENF, GRA,
and RAG, respectively), while this variability from the sim-
ulations is 53 % for the BELD version of the DBF model
and 61 % for the PFT version, 55 and 92 % for the BELD
and PFT versions of the ENF model, 43 % for grasses, and
49 % for ragweed (Fig. 12). This indicates that the model
is capturing the relative interannual variability of the pollen
counts between PFTs but not all of the variability in pollen
counts from season to season. The unexplained variability
in pollen concentrations could be due to the lack of sensi-
tivity of annual pollen production factor to the environment,
as this may be closely tied with precipitation (Duhl et al.,
2013) or temperature (Jochner et al., 2013). Additionally, the
average observed and simulated pollen counts are analyzed
using box-and-whisker plots to assess the models' represen-
tativeness of pollen count magnitude in spite of phenology
(Fig. 13). These metrics are discussed in detail by PFT and
US subregion below.
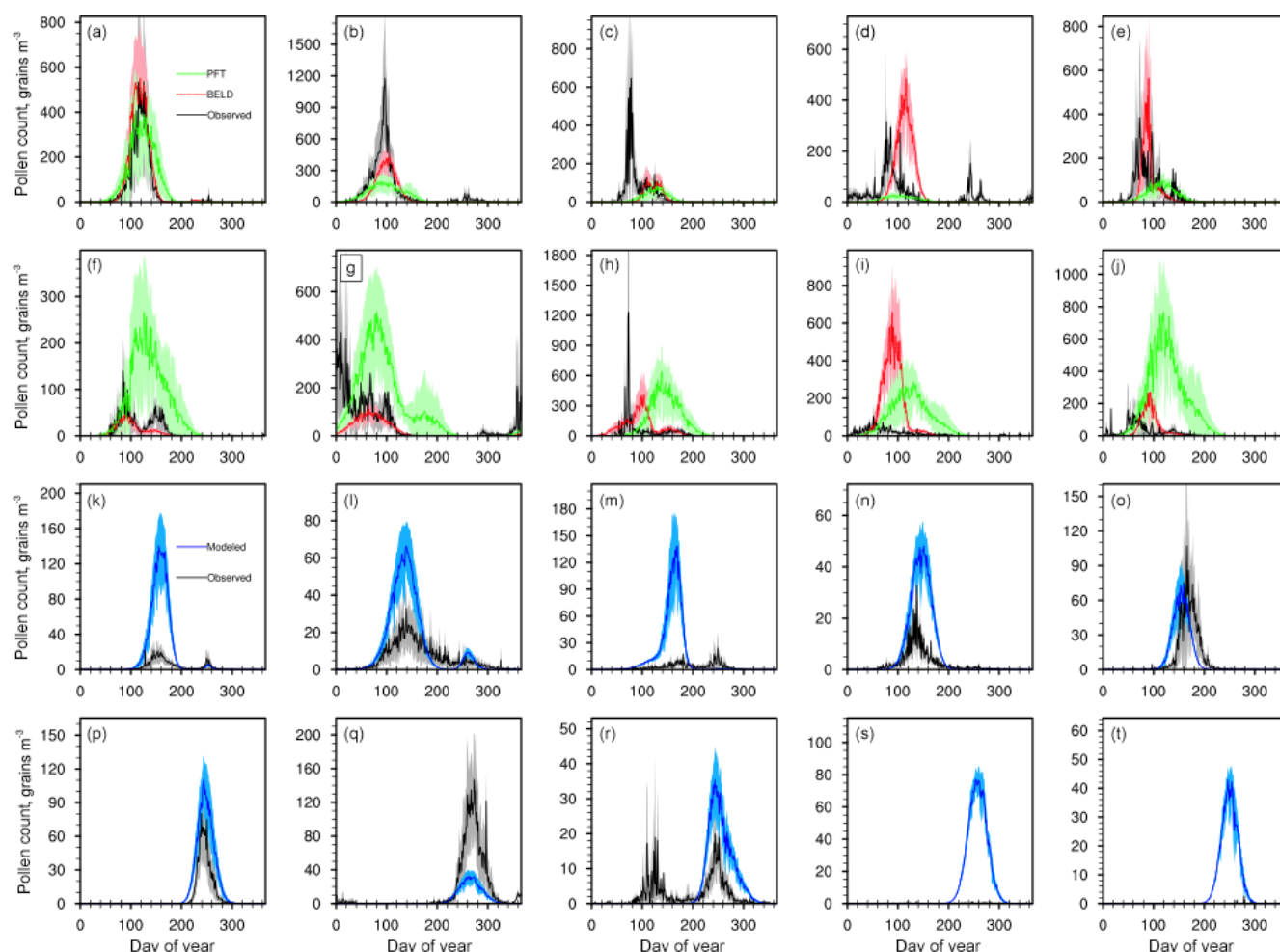
Figure 12. Average (2003–2010) time series of daily pollen counts comparing model and observations for four PFTs (**a–e**, deciduous broadleaf, DBF; **f–j**, evergreen needleleaf, ENF; **k–o**, grasses, GRA; **p–t**, ragweed, RAG) across five US subregions (columns from left to right: Northeast, NE; Southeast, SE; Mountain, MT; California, CA; Pacific Northwest, PNW). Shading for the observations and model represents the mean absolute deviation from the average for each day of the time series. Note that the scale of the *y* axes varies by region and PFT.

5.2.1   DBF

In the Northeast, the BELD model captures both the observed seasonal timing and the magnitude of DBF pollen counts (Fig. 12a). Observed DBF phenology is also simulated by the PFT-based emissions with even greater statistical accuracy in reproducing the observed pollen counts, though the BELD model more accurately reproduces the annual maximum (Fig. 13a). The accuracy in this subregion is not surprising, as northeastern pollen counting stations contributed the greatest number of data points to the phenological regression analyses. Observed DBF pollen counts in the Southeast have a large maximum that is greater than the average seasonal maximum of all four other subregions and all three other PFTs (Fig. 12b), which is predominantly from *Quercus*. Neither the BELD nor PFT version of the simulation recreates this sharp peak, but they do simulate a large majority of

the pollen count distribution (Fig. 13b), especially the PFT-based model for which the lower 75 % of simulated average pollen counts agrees well with the lower 75 % of observed average pollen counts. The PFT model does not specifically resolve *Quercus*, and while the BELD model does resolve *Quercus*, it fails to model this maximum. This may be because the linear regression producing the phenological dates is an average, where a longer season may result from earlier start dates and/or later end dates that will reduce the maximum of the Gaussian distribution of pollen counts in the time series. In the Mountain region, there is an observed maximum early in the spring that is not simulated by either model because the DBF phenology at several cold Mountain sites is exceptionally early and falls well below the regression lines (Figs. S1, S2). However, both the BELD and PFT model simulate the second Mountain subregion peak with the correct magnitude. The BELD-simulated maximum
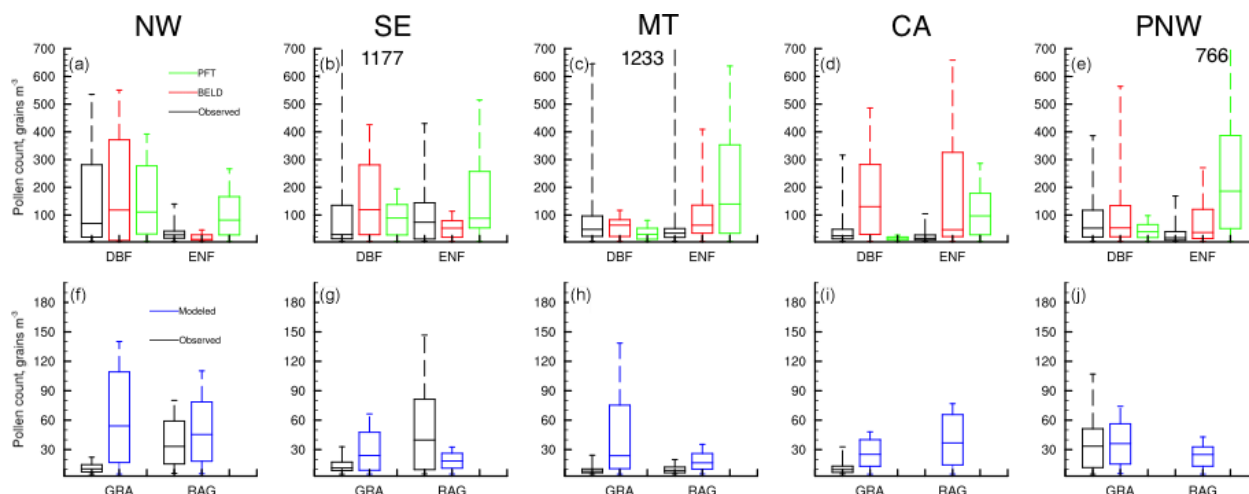
Figure 13. Box-and-whisker plots showing the statistical spread of the pollen count magnitudes from the regional averages presented in Fig. 12. Columns from left to right: Northeast, NE (**a, f**); Southeast, SE (**b, g**); Mountain, MT (**c, h**); California, CA (**d, i**); Pacific Northwest, PNW (**e, j**). DBF and ENF PFTs are shown in the top row (**a–e**) and grass and ragweed PFTs are shown in the bottom row (**f–j**). Boxes and whiskers from bottom to top represent the minimum, lower quartile, median, upper quartile, and maximum. Maxima that are not visible in panels (**b**, **c**, and **e**) are 1177, 1233, and 766 gains m^{-3}, respectively. All y axes are the same scale for each row.

DBF in California is about 40 days later than the observed peak, also due to the regionally anomalous phenology in California as compared with the rest of the US, and though the PFT model peaks much closer to the observations, it underestimates DBF pollen counts. In the Pacific Northwest, the observed pattern is quite similar to the DBF pollen phenology in the Mountain subregion, with only a slightly weaker early spring peak due to low-elevation pollen. The observed phenological pattern (Fig. 12e) and pollen count magnitudes (Fig. 12e) are both more accurately simulated by the BELD model, likely due to the earlier spring maximum that does not appear in the PFT simulation.

5.2.2   ENF

Like DBF, the BELD ENF in the Northeast is well represented by simulating two distinct Cupressaceae and Pinaceae maxima, although the model slightly underestimates observed Pinaceae pollen counts (Fig. 12f). The PFT model ENF phenology emits from the start of the earlier Cupressaceae season to the end of the later Pinaceae season, while overestimating the maximum pollen count by about a factor of 2. In the Southeast, the winter peak is not captured by the model phenology (Fig. 12g). However, the spring Pinaceae maximum is accurately captured by the BELD simulation. The PFT model follows the observed Pinaceae phenology more closely, though overestimating pollen counts by a factor of 2 to 3 and estimating a later ending date by about 40 days. In the Mountain subregion, ENF start and end dates are simulated by the BELD model with improved accuracy than the DBF phenology in this subregion, though the predicted spring maximum is later than observed (Fig. 12h). As with

DBF, there is good agreement between the BELD model with the later part of the season in this subregion. The PFT model, again, simulates the peak ENF emissions in the later part of the season and overpredicts the pollen counts by a factor of 2 to 3. In the California subregion, the tails of the pollen distributions by both models closely resemble the pollen count magnitudes, yet the majority of these pollen counts (the top 75 %, Fig. 13i) lie above the observed maximum (Fig. 12i). Finally, in the Pacific Northwest, the BELD model phenology shows some agreement with the model mean (Fig. 13j), with the simulated pollen count showing a stronger Gaussian distribution than observed (Fig. 12j). In contrast, the PFT model grossly overpredicts the observed pollen counts by up to a factor of 10 at its maximum, likely due to the greater representation of the ENF PFT than the BELD model in this region. The simulated average start date of the PFT model is within a few days of the observed average start date, while the end date is about 20 days later than observed.

5.2.3   Grasses

Grass phenology across all subregions for both C$_3$ and C$_4$ types is captured by the emissions estimates (Fig. 12k–o). However, the pollen count magnitude in northeastern C$_3$ grass peak is overestimated by about a factor of 7, even when using the minimum value of the annual production factor in the range estimated by Prieto-Baena (2003) (Fig. 12k). The secondary peak, which we attribute to C$_4$ grasses and is only about half as large, is well represented. In the Southeast, the simulated pollen count magnitudes are much closer to observations, while the C$_3$ peak is overestimated here by only a factor of 2 and the C$_4$ peak is within 5 grains m^{-3} (Fig. 12l).

In this region, the observed duration of the pollen emissions is not fully captured by the simulated grass phenology in the Southeast, and this is probably due to the non-Gaussian shape of the observed time series. In the Mountain subregion, the C_3 pollen count is overestimated by the model, but the phenology is represented by a gradual rise in low emissions beginning in March to match the maximum burst of emissions in June (Fig. 12m). C_4 grass pollen counts are not simulated in the Mountain region due to the relatively low C_4 land cover in the CLM4 dataset (Fig. 4f). In California, there is a single observed grass peak, which the model attributes to C_3 pollen, and the peak count in the simulation is about 5 days late and about 2 to 10 times too large (Fig. 12n). In the Pacific Northwest, the average C_3 season is accurately simulated with the exception that the phenology is shifted 20 days earlier than observed (Fig. 12o). A small C_4 peak in the observations at around day 260 is not simulated in this region due to negligible land cover for C_4 grasses in the CLM4 land cover data (Fig. 4f).

5.2.4 Ragweed

Simulated ragweed phenology in the Northeast, Southeast, and Mountain subregions follows the observed phenology of late-summer ragweed very closely, where the peaks of both the simulated and the observed time series averages occur within a day of each other (Fig. 12p–t). Close evaluation of each regional phenological time series reveals that many of the observed features, like those determined by the rate of increase or decrease of the pollen count, are reproduced by the model. The magnitude of the modeled ragweed maxima in the Northeast and Mountain subregions is slightly greater than observed (Fig. 12p and r), while there is a clear underestimation by a factor of 4 or 5 in the Southeast (Fig. 12q). There is a yet unidentified observed spring peak of ragweed pollen at about day 125 in the Mountain subregion, possibly due to an identification error. The observed average ragweed pollen counts in California and the Pacific Northwest are negligible, though the simulation predicts them to be similar in magnitude and timing to the other three subregions (Fig. 12s and t). These discrepancies may be due to the land use description developed for ragweed (Sect. 3.1), which may overestimate the ragweed potential in the western United States, or potentially the relatively spare observational stations in these regions may be poorly placed relative to emissions sources.

6 Conclusions

We have developed a climate-flexible pollen emissions model (PECM) for the 13 most prevalent wind-pollinating taxa in the United States based on observed pollen counts. PECM was adapted to the PFT categorization common to climate and Earth system models with four major temperate-

zone PFTs (DBF, ENF, grasses and ragweed); thus, it is possible to apply this model to larger geographic regions where specific taxon-level data are unavailable. We evaluated PECM using a regional climate model (RegCM4) to transport emissions and evaluated resulting pollen counts versus observations. PECM generally captures the observed phenology, and observed surface pollen concentrations can be simulated within an order of magnitude. While many emissions models to date have focused on smaller geographical regions with more detailed land cover information and pollen information, this model represents the first of its kind to simulate multiple taxa over broad spatial areas. This transition to a larger scale does have its disadvantages, and we define several major sources of uncertainty to consider when scaling up pollen emissions to the regional or global scale: (1) pollen production factors, (2) climatic sensitivities in phenological timing, (3) land cover data, and (4) taxa specificity. We discuss each of these uncertainties in greater detail.

A large source of uncertainty is the use of a constant annual production factor for pollen (Sect. 4.3). It has been reported that wind-driven pollen production has increased historically and is expected, potentially, to increase in the near future (R. Zhang et al., 2014; Lake et al., 2017; Confalonieri et al., 2007; Ziello et al., 2012). Some of more effective improvements to the emissions model would be to create a pollen production model that is sensitive to multiple environmental factors such as soil moisture, temperature, and nutrient status (Jochner et al., 2013). The interannual variability in observed daily pollen counts is, on average, substantially greater than that of the modeled pollen counts, which is likely due to this lack of production sensitivity. The current production factors for woody plants could be enhanced by studies that extend the number of representative units (i.e., individual trees) of vegetation used to determine the average pollen production. In a PFT representation, there is an inevitable limitation to the accuracy of any single PFT's ability to account for taxa differences within the PFT. Furthermore, the current model also assumes that there are no interspecies differences that affect the performance of the BELD model as well as the PFT model, whereas in reality it may vary by an order of magnitude within a genus (Duhl et al., 2013). However, despite the assumption of a constant production factor, observed surface pollen counts for all PFTs are typically reproduced within a single order of magnitude, as apparent in emissions model evaluation.

Second, the use of observed relationships between pollen count and temperature to determine the phenological pollen start and end dates also adds uncertainty to our modeling framework. Firstly, we assume stationarity in the phenological relationships, and this assumption may be violated. Secondly, based on the subregions defined for the analysis, there appears to be a bias in the linear regressions toward subregions with more available pollen counting stations, therefore affecting performance differences in these regions. Lastly, even though generally the Gaussian time series model of the

pollen phenology performs well in our analysis, in the PFT representation, the Gaussian absorbs or misses some of the phenological details in the observed pollen seasonality, and in some cases taxa (e.g., grasses in the Southeast subregion) may not be captured by the existing phenology.

Third, the specificity of land cover data provides an important constraint in the overall simulation of emissions. The representation of land cover is a key factor to accurately capturing regional features, especially in areas with a high degree of topographical variation and therefore greater variance in the land cover. For example, we notice large differences in the two model simulations when considering tree-specific taxa, such as in the western United States for ENF (Sect. 5.2.2). Also, our definition of the land cover available for ragweed used assumptions based on crop cover and urban area, which overestimated emissions in the western United States (Sect. 5.2.4). Interestingly, even though ragweed lacks an exact spatial distribution, distinct observed features of the ragweed phenology in three of the five subregions emerged using the current ragweed land cover parameterization.

Fourth, the aggregation of emissions to the PFT level affects the representativeness of the production factors, phenology, and land cover. When comparing the two models of the tree pollen (BELD versus PFT), the individual phenology of each of the 11 tree taxa is resolved by the BELD simulation, whereas they are either folded into or excluded from the single phenology modeled by the PFT simulation. This results from either treating the taxa in the phenological regressions individually, as in the BELD model, or as a sum, as in the PFT model. With a few exceptions (e.g., the ENF family distinctions), the PFT model does generally reproduce the regional phenology throughout the United States domain, which is a priority of this study.

Despite these limitations, the empirical formulation presented here is the first of its kind to predict a broad range of different pollen emissions across a large geographic region. Even with univariate phenology and invariable pollen production factors, the model includes seasonal dynamics sensitive to climate change consistent with observations and is also able to simulate observed pollen magnitudes. As a result, the model can be useful for estimation of how allergenic risk or plant reproductive potential will be redistributed by climate change, as well as studying pollen as an aerosol in the climate system. While the empirical phenological models can be reproduced for any set of regional pollen counting stations, PECM as a whole can be easily adapted to various community climate and Earth system models, global and regional, to extend research on the relationships and interactions between pollen and climate.

References

Beggs, P. J., Šikoparija, B., and Smith, M.: Aerobiology in the International Journal of Biometeorology, 1957–2017, 61, 51–58, https://doi.org/10.1007/s00484-017-1374-5, 2017.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M.: Time Series Analysis: Forecasting and Control, Upper Saddle River, Prentice Hall, NJ, 1994.

Cecchi, L.: Introduction, in: Allergenic Pollen: A Review of the Production, Release, Distribution and Health Impacts, edited by: Sofiev, M. and Bergmann, K.-C., Springer Science+Business Media, Dordrecht, New York, London, 1–7, https://doi.org/10.1007/978-94-007-4881-1, 2014.

Chuine, I., Cour, P., and Rousseau, D. D.: Selecting Models to Predict the Timing of Flowering of Temperate Trees: Implications for Tree Phenology Modelling. Plant Cell Environ. 22, 1–13, https://doi.org/10.1046/j.1365-3040.1999.00395.x, 1999.

Clay, S. A., Kreutner, B., Clay, D. E., Reese, C., Kleinjan, J., and Forcella, F.: Spatial distribution, temporal stability, and yield loss estimates for annual grasses and common ragweed (*Ambrosia artimisiifolia*) in a corn/soybean production field over nine years, Weed Sci. 54, 380–390, https://doi.org/10.1614/WS-05-090R1.1, 2006.

Confalonieri, U., Menne, B., Akhtar, R., Ebi, K. L., Hauengue, M., Kovats, R. S., Revich, B., and Woodward, A.: Human Health, in: Climate Change 2007: Impacts, Adaptation and Vulnerability, Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J., and Hanson, C. E., Cambrige University Press, Cambridge, UK, chap. 8, 391–431, 2007.

Craine, J. M., Wolkovich, E. M., Gene Towne, E., and Kembel, S. W.: Flowering phenology as a functional trait in a tallgrass prairie, New Phytol., 193, 673–682, https://doi.org/10.1111/j.1469-8137.2011.03953.x, 2012.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. Roy. Meteor. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Deen, W., Hunt, T., and Swanton, C. J.: Influence of Temperature, Photoperiod, and Irradiance on the Phenological Development of Common Ragweed (*Ambrosia artemisiifolia*), Weed Science, 46, 555–560, 1998.

Després, V. R., Huffman, J. A., Burrows, S. M., Hoose, C., Safatov, A. S., Buryak, G., Fröhlich-Nowoisky, J., Elbert, W., Andreae, M. O., Pöschl, U., and Jaenicke, R.: Primary biological aerosol particles in the atmosphere: a review, Tellus B, https://doi.org/10.3402/tellusb.v64i0.15598, 2012.

Duhl, T. R., Zhang, R., Guenther, A., Chung, S. H., Salam, M. T., House, J. M., Flagan, R. C., Avol, E. L., Gilliland, F. D., Lamb, B. K., VanReken, T. M., Zhang, Y., and Salathé, E.: The Simulator of the Timing and Magnitude of Pollen Season (STaMPS) model: a pollen production model for regional emission and transport modeling, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmdd-6-2325-2013, 2013.

Easterling, D. R.: Recent Changes in Frost Days and the Frost-Free Season in the United States, B. Am. Meteorol. Soc., 83, 1327–1332, https://doi.org/10.1175/1520-0477(2002)083<1327:RCIFDA>2.3.CO;2, 2002.

Efstathiou, C., Isukapalli, S., and Georgopoulos, P.: A Mechanistic Modeling System for Estimating Large-Scale Emissions and Transport of Pollen and Co-Allergens, Atmos. Environ., 45, 2260–2276, https://doi.org/10.1016/j.atmosenv.2010.12.008, 2011.

Emanuel, K. A.: A Scheme for Representing Cumulus Convention in Large-Scale Models, J. Atmos. Sci., 48, 2313–2335, 1991.

Emberlin, J., Mullins, J., Corden, J., Jones, S., Millington, W., Brooke, M., and Savage, M.: Regional Variations in Grass Pollen Seasons in the UK, Long-Term Trends and Forecast Models, Clinical and Experimental Allergy, 29, 347–356, https://doi.org/10.1046/j.1365-2222.1999.00369.x, 1999.

Forman, R. T. T. and Alexander, L. E.: Roads and their major ecological effects. Annu. Rev. Ecol. Syst., 29, 207–231, https://doi.org/10.1146/annurev.ecolsys.29.1.207, 1998.

Fu, Y. H., Campioli, M., Deckmyn, G., and Janssens, I. A.: The Impact of Winter and Spring Temperatures on Temperate Tree Budburst Dates: Results from an Experimental Climate Manipulation, PLoS ONE, 7, e47324, https://doi.org/10.1371/journal.pone.0047324, 2012.

Fumanal, B., Chauvel, B., and Bretagnolle, F.: Estimation of pollen and seed production of common ragweed in france. Ann. Agric. Environ. Med., 14, 233–236, https://doi.org/10.1093/annonc/mdw163, 2007.

Galán, C., García-Mozo, H., Vázquez, L., Ruiz, L., Díaz de la Guardia, C., and Domínguez-Vilches, E.: Modeling Olive Crop Yield in Andalusia, Spain, Agron. J., 100, 98–104, https://doi.org/10.2134/agronj2006.0345, 2008.

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, M. M., Valencia, R., and Chuine, I.: Predicting the start and peak dates of the Poaceae pollen season in Spain using process-based models, Agr. Forest. Meteorol., 149, 256–262, https://doi.org/10.1016/j.agrformet.2008.08.013, 2009.

Giorgi, F., Coppola, E., Solmon, F., Mariotti, L., Sylla, M. B., Bi, X., Elguindi, N., Diro, G. T., Nair, V., Giuliani, G., Turuncoglu, U. U., Cozzini, S., Güttler, I., O'Brien, T. A., Tawfik, A. B., Shalaby, A., Zakey, A. S., Steiner, A. L., Stordal, F., Sloan, L. C., and Brankovic, C.: RegCM4: Model Description and Preliminary Tests over Multiple CORDEX Domains, Climate Research, 52, 7–29, https://doi.org/10.3354/cr01018, 2012.

Grell, G. A., Dudhia, J., and Stauffer, D. R.: A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5), NCAR Technical Note NCAR/TN-398+STR, NCAR, Boulder, USA, https://doi.org/10.5065/D60Z716B, 1994.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181–3210, https://doi.org/10.5194/acp-6-3181-2006, 2006.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34, 623–642, https://doi.org/10.1002/joc.3711, 2014.

Helbig, N., Vogel, B., Vogel, H., and Fiedler, F.: Numerical Modelling of Pollen Dispersion on the Regional Scale, Aerobiologia, 3, 3–19, 2004.

Hidalgo, P. J., Galán, C., and Domínguez, E.: Pollen Production of the Genus Cupressus, Grana, 38, 296–300, https://doi.org/10.1080/001731300750044519, 1999.

Hidalgo, P. J., Mangin, A., Galán, C., Hembise, O., Vázquez, L. M., and Sanchez, O.: An Automated System for Surveying and Forecasting Olea Pollen Dispersion, Aerobiologia, 18, 23–31, 2002.

Hunt, J. C. R., Higson, H. L., Walklate, P. J., and Sweet, J. B.: Modelling the Dispersion and Cross-Fertilisation of Pollen from GM Crops, 2002.

Jackson, T. L., Feddema, J. J., Oleson, K. W., Bonan, G. B., and Bauer, J. T.: Parameterization of Urban Characteristics for Global Climate Modeling, https://doi.org/10.1080/00045608.2010.497328, 2010.

Jato, V., Rodríguez-Rajo, F. J., and Aira, M. J.: Use of Phenological and Pollen-Production Data for Interpreting Atmospheric Birch Pollen Curves, Ann. Agr. Env. Med., 14, 271–280, 2007.

Jochner, S., Höfler, J., Beck, I., Göttlein, A., Ankerst, D. P., Traidl-Hoffmann, C., and Menzel, A.: Nutrient status: a missing factor in phenological and pollen research?, J. Exp. Bot., 64, 2081–2092, https://doi.org/10.1093/jxb/ert061, 2013.

Katz, D. S. W., Barrie, B. T. C., and Carey, T. S.: Urban ragweed populations in vacant lots: An ecological perspective on management, Urban For. Urban Gree., 13, 756–760, https://doi.org/10.1016/j.ufug.2014.06.001, 2014.

Kinnee, E., Geron, C., and Pierce, T.: United States Land Use Inventory For Estimating Biogenic Ozone Precursor Emissions, Ecol. Appl., 7, 46–58, 1997.

Kuparinen, A., Markkanen, T., Riikonen, H., and Vesala, T.: Modeling Air-Mediated Dispersal of Spores, Pollen and

Seeds in Forested Areas, Ecol. Model., 208, 177–188, https://doi.org/10.1016/j.ecolmodel.2007.05.023, 2007.

Lake, I. R., Jones, N. R., Agnew, M., Goodess, C. M., Giorgi, F., Hamaoui-Laguel, L., Semenov, M. A., Solomon, F., Storkey, J., Vautard, R., and Epstein, M. M.: Climate Change and Future Pollen Allergy in Europe, Environ. Health Persp., 125, 385–391, https://doi.org/10.1289/EHP173, 2017.

Larson, D. L.: Native weeds and exotic plants: relationships to disturbance in mixed-grass prairie, Plant Ecol., 169, 317–333, https://doi.org/10.1023/A:1026046810307, 2003.

Lawrence, P. J. and Chase, T. N.: Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0), J. Geophys. Res.-Biogeo., 112, G01023, https://doi.org/10.1029/2006JG000168, 2007.

Lewis, W. H., Vinay, P., and Zenger, V. E.: Airborne and allergenic pollen of North America, Johns Hopkins University Press, Baltimore, 1983.

Linkosalo, T., Lappalainen, H. K., and Hari, P.: A comparison of phenological models of leaf bud burst and flowering of boreal trees using independent observations, Tree Physiol., 28, 1873–1882, https://doi.org/10.1093/treephys/28.12.1873, 2008.

Liu, L., Solmon, F., Vautard, R., Hamaoui-Laguel, L., Torma, C. Z., and Giorgi, F.: Ragweed pollen production and dispersion modelling within a regional climate system, calibration and application over Europe, Biogeosciences, 13, 2769–2786, https://doi.org/10.5194/bg-13-2769-2016, 2016.

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., and Nijssen, B.: A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions, J. Climate, 15, 3237–3251, https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2, 2002.

Medek, D. E., Beggs, P. J., Erbas, B., Jaggard, A. K., Campbell, B. C., Vicendese, D., Johnston, F. H., Godwin, I., Huete, A. R., Green, B. J., Burton, P. K., Bowman, D. M. J. S., Newnham, R. M., Katelaris, C. H., Haberle, S. G., Newbigin, E., and Davies, J. M.: Regional and Seasonal Variation in Airborne Grass Pollen Levels between Cities of Australia and New Zealand, Aerobiologia, 32, 289–302, https://doi.org/10.1007/s10453-015-9399-x, 2016.

Menzel, A. and Jochner, S.: Impacts of Climate Change on Aeroallergen Production and Atmospheric Concentration, in: Impacts of Climate Change on Allergens and Allergic Diseases, edited by: Beggs, P. J., Cambridge University Press, Cambridge, UK, 10–28, 2016.

Moseholm, L., Weeke, E. R., and Petersen, B. N.: Forecast of Pollen Concentrations of Poaceae (Grasses) in the Air by Time Series Analysis, Pollen et Spores, 29, 305–321, 1987.

Myking, T. and Heide O. M.: Dormancy Release and Chilling Requirement of Buds of Latitudinal Ecotypes of *Betula pendula* and *B. ubescens*, Tree Physiol., 15, 697–704, https://doi.org/10.1093/treephys/15.11.697, 1995.

Myriokefalitakis, S., Fanourgakis, G., and Kanakidou, M.: The Contribution of Bioaerosols to the Organic Carbon Budget of the Atmosphere, in: Perspectives on Atmospheric Sciences, edited by: Karacostas, T., Bais, A., and Nastos, P. T., Springer International Publishing, 845–851, https://doi.org/10.1007/978-3-319-35095-0_121, 2017.

Oleson, K., Lawrence, D., Bonan, G., Flanner, M., and Kluzek, E.: Technical Description of Version 4.0 of the Community Land Model (CLM), NCAR Technical Note NCAR/TN-478+STR, NCAR, Boulder, USA, 257 pp., https://doi.org/10.5065/D6FB50WZ, 2010.

Olsson, C. and Jönsson, A. M.: Process-Based Models Not Always Better than Empirical Models for Simulating Budburst of Norway Spruce and Birch in Europe, Global Change Biol., 20, 3492–3507, https://doi.org/10.1111/gcb.12593, 2014.

Pal, J. S., Small, E. E., and Eltahir, E. A. B.: Simulation of regional-scale water and energy budgets: Representation of subgrid cloud and precipitation processes within RegCM, J. Geophys. Res.-Atmos., 105, 29579–29594, https://doi.org/10.1029/2000JD900415, 2000.

Prieto-Baena, J. C., Hidalgo, P. J., Galán, C., and Domínguez, E.: Pollen Production in the Poaceae Family, Grana, 42, 153–159, 2003.

Reichstein, M., Bahn, M., Mahecha, M. D., Kattge, J., and Baldocchi, D. D.: Linking Plant and Ecosystem Functional Biogeography, P. Natl. Acad. Sci. USA, 111, 13697–13702, https://doi.org/10.1073/pnas.1216065111, 2014.

Scheifinger, H., Belmonte, J., Buters, J., Celenk, S., Damialis, A., Dechamp, C., García-Mozo, H., Gehrig, R., Grewling, L., Halley, J. M., Hogda, K.-A., Jäger, S., Karatzas, K., Karlsen, S.-R., Koch, E., Pauling, A., Peel, R., Sikoparija, B., Smith, M., Galán-Soldevilla, C., Thibaudon, M., Vokou, D., and de Weger, L. A.: Monitoring, Modelling and Forecasting of the Pollen Season, in: Allergenic Pollen: A Review of the Production, Release, Distribution and Health Impacts, edited by: Sofiev, M. and Bergmann, K.-C., Springer Science+Business Media, Dordrecht, New York, London, 71–126, https://doi.org/10.1007/978-94-007-4881-1_4, 2013.

Schueler, S. and Heinke Schlünzen, K.: Modeling of Oak Pollen Dispersal on the Landscape Level with a Mesoscale Atmospheric Model, Environ. Model. Assess., 11, 179–194, https://doi.org/10.1007/s10666-006-9044-8, 2006.

Siljamo, P., Sofiev, M., Filatova, E., Grewling, Ł., Jäger, S., Khoreva, E., Linkosalo, T., Ortega Jimenez, S., Ranta, H., Rantio-Lehtimäki, A., Svetlov, A., Veriankaite, L., Yakovleva, E., and Kukkonen, J.: A numerical model of birch pollen emission and dispersion in the atmosphere. Model evaluation and sensitivity analysis, Int. J. Biometeorol., 57, 125–136, https://doi.org/10.1007/s00484-012-0539-5, 2013.

Smith, M. and Emberlin J.: Constructing a 7-Day Ahead Forecast Model for Grass Pollen at North London, United Kingdom, Clin. Exp. Allergy, 35, 1400–1406, https://doi.org/10.1111/j.1365-2222.2005.02349.x, 2005.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis (1880–2006), J. Climate, 21, 2283–2296, https://doi.org/10.1175/2007JCLI2100.1, 2008.

Sofiev, M. and Prank, M.: Impacts of Climate Change on Aeroallergen Dispersion, Transport, and Deposition, in: Impacts of Climate Change on Allergens and Allergic Diseases, edited by: Beggs, P. J., Cambrige University Press, Cambridge, UK, 50–73, 2016.

Sofiev, M., Siljamo, P., Ranta, H., and Rantio-Lehtimäki, A.: Towards Numerical Forecasting of Long-Range Air Transport of Birch Pollen: Theoretical Considerations and

a Feasibility Study, Int. J. Biometeorol., 50, 392–402, https://doi.org/10.1007/s00484-006-0027-x, 2006.

Sofiev, M., Siljamo, P., Ranta, H., Linkosalo, T., Jaeger, S., Rasmussen, A., Rantio-Lehtimaki, A., Severova, E., and Kukkonen, J.: A Numerical Model of Birch Pollen Emission and Dispersion in the Atmosphere. Description of the Emission Module, Int. J. Biometeorol., 57, 45–58, https://doi.org/10.1007/s00484-012-0532-z, 2013.

Sofiev, M., Belmonte, J., Gehrig, R., Izquierdo, R., Smith, M., Dahl, A., and Siljamo, P.: Airborne Pollen Transport, in: Allergenic Pollen: A Review of the Production, Release, Distribution and Health Impacts, edited by: Sofiev, M. and Bergman, K.-C., Springer Science+Business Media Dordrecht, New York, London, 127–159, https://doi.org/10.1007/978-94-007-4881-1, 2014.

Solmon, F., Giorgi, F., and Liousse, C.: Aerosol modelling for regional climate studies: application to anthropogenic particles and evaluation over a European/African domain, Tellus B, 58, 51–72, https://doi.org/10.1111/j.1600-0889.2005.00155.x, 2006.

Thornton, P. E., Thornton, M. M., Mayer, B. W., Wilhelmi, N., Wei, Y., Devarakonda, R., and Cook, R. B.: Daymet: Daily Surface Weather Data on a 1-Km Grid for North America, Version 2. Data Set, Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, https://doi.org/10.3334/ORNLDAAC/1219, 2014.

Tormo Molina, R., Muñoz Rodríguez, A., Palacios, S., and Gallardo Lopes, F.: Pollen Production in Anemophilous Trees, Grana, 35, 38–46, https://doi.org/10.1080/00173139609430499, 1996.

Weber, R. W.: Meteorologic Variables in Aerobiology, Immunol. Allergy Clin., 23, 411–422, https://doi.org/10.1016/S0889-8561(03)00062-6, 2003.

Yue, X., Unger, N., Keenan, T. F., Zhang, X., and Vogel, C. S.: Probing the past 30-year phenology trend of US deciduous forests, Biogeosciences, 12, 4693–4709, https://doi.org/10.5194/bg-12-4693-2015, 2015.

Zhang, R., Duhl, T., Salam, M. T., House, J. M., Flagan, R. C., Avol, E. L., Gilliland, F. D., Guenther, A., Chung, S. H., Lamb, B. K., and VanReken, T. M.: Development of a regional-scale pollen emission and transport modeling framework for investigating the impact of climate change on allergic airway disease, Biogeosciences, 11, 1461–1478, https://doi.org/10.5194/bg-11-1461-2014, 2014.

Zhang, Y., Bielory, L., Mi, Z., Cai, T., Robock, A., and Georgopoulos, P.: Allergenic Pollen Season Variations in the Past Two Decades under Changing Climate in the United States, Global Change Biol., 21, 1581–1589, https://doi.org/10.1111/gcb.12755, 2015a.

Zhang, Y., Bielory, L., Cai, T., Mi, Z., and Georgopoulos, P.: Predicting Onset and Duration of Airborne Allergenic Pollen Season in the United States, Atmos. Environ., 103, 297–306, https://doi.org/10.1016/j.atmosenv.2014.12.019, 2015b.

Ziello, C., Sparks, T., Estrella, N., Belmonte, J., Bergmann, K. C., Bucher, E., Brighetti, M. A., Damialis, A., Detandt, M., Galán, C., Gehrig, R., Grewling, L., Gutiérrez Bustillo, A., Hallsdóttir, M., Kockhans-Bieda, M.-C., De Linares, C., Myszkowska, D., Pàldy, A., Sánchez, A., Smith, M,. Thibaudon, M,. Travaglini, A., Uruska, A., Valencia-Barrera, R. M., Vokou, D., Wachter, R., de Weger, L. A., and Menzel, A.: Changes to Airborne Pollen Counts across Europe, PLoS ONE, 7, 1–8, https://doi.org/10.1371/journal.pone.0034076, 2012.

Zink, K., Pauling, A., Rotach, M. W., Vogel, H., Kaufmann, P., and Clot, B.: EMPOL 1.0: a new parameterization of pollen emission in numerical weather prediction models, Geosci. Model Dev., 6, 1961–1975, https://doi.org/10.5194/gmd-6-1961-2013, 2013.

Zink, K., Kaufmann, P., Petitpierre, B., Broennimann, O., Guisan, A., Gentilini, E., and Rotach, M. W.: Numerical ragweed pollen forecasts using different source maps: a comparison for France, Int. J. Biometeorol., 61, 23–33, https://doi.org/10.1007/s00484-016-1188-x, 2017.

Ziska, L. H: Impacts of Climate Change on Allergen Seasonality, in: Impacts of Climate Change on Allergens and Allergic Diseases, edited by: Beggs, P. J., Cambrige University Press, Cambridge, UK, 93–112, 2016.

Ziska, L., Knowlton, K., Rogers, C., Dalan, D., Tierney, N., Elder, M. A., Filley, W., Shropshire, J., Ford, L. B., Hedberg, C., Fleetwood, P., Hovanky, K. T., Kavanaugh, T., Fulford, G., Vrtis, R. F., Patz, J. A., Portnoy, J., Coates, F., Bielory, L., and Frenz, D.: Recent warming by latitude associated with increased length of ragweed pollen season in central North America, P. Natl. Acad. Sci. USA, 108, 4248–4251, https://doi.org/10.1073/pnas.1014107108, 2011.