

# 1 **A prognostic pollen emissions model for climate models** 2 **(PECM1.0)**

3 Matthew C. Wozniak<sup>1</sup>, Allison L. Steiner<sup>1</sup>

4 <sup>1</sup>Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109, USA

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

6 **Abstract.** We develop a prognostic model of Pollen Emissions for Climate Models (PECM) for use within regional  
7 and global climate models to simulate pollen counts over the seasonal cycle based on geography, vegetation type  
8 and meteorological parameters. Using modern surface pollen count data, empirical relationships between prior-year  
9 annual average temperature and pollen season start dates and end dates are developed for deciduous broadleaf trees  
10 (*Acer*, *Alnus*, *Betula*, *Fraxinus*, *Morus*, *Platanus*, *Populus*, *Quercus*, *Ulmus*), evergreen needleleaf trees  
11 (Cupressaceae, Pinaceae), grasses (Poaceae; C<sub>3</sub>, C<sub>4</sub>), and ragweed (*Ambrosia*). This regression model explains as  
12 much as 57% of the variance in pollen phenological dates, and it is used to create a “climate-flexible” phenology  
13 that can be used to study the response of wind-driven pollen emissions to climate change. The emissions model is  
14 evaluated in a regional climate model (RegCM4) over the continental United States by prescribing an emission  
15 potential from PECM and transporting pollen as aerosol tracers. We evaluate two different pollen emissions  
16 scenarios in the model, using: (1) a taxa-specific land cover database, phenology and emission potential, and (2) a  
17 plant functional type (PFT) land cover, phenology and emission potential. The simulated surface pollen  
18 concentrations for both simulations are evaluated against observed surface pollen counts in five climatic subregions.  
19 Given prescribed pollen emissions, the RegCM4 simulates observed concentrations within an order of magnitude,  
20 although the performance of the simulations in any subregion is strongly related to the land cover representation and  
21 the number of observation sites used to create the empirical phenological relationship. The taxa-based model  
22 provides a better representation of the phenology of tree-based pollen counts than the PFT-based model, however we  
23 note that the PFT-based version provides a useful and “climate-flexible” emissions model for the general  
24 representation of the pollen phenology over the United States.

25

## 26 **1 Introduction**

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

37 Prognostic pollen emissions are useful for the scientific community and public, specifically for forecasting  
38 allergenic conditions or predicting the flow of genetic material. The interest and growing wealth of knowledge of  
39 allergenic pollen has been recently reviewed by *Beggs et al.* (2017). To date, most pollen emissions models focus on  
40 relatively short, seasonal time scales and smaller locales for a limited selection of taxa (Sofiev et al. 2013; Liu et al.  
41 2016; R. Zhang et al. 2014). Climatic changes in large-scale pollen distributions are mostly absent from scientific  
42 literature, though multiple studies on phenological changes in the pollen season have been published (Ziska 2016;  
43 Yue et al. 2015; Y. Zhang et al. 2015a). Only recently have regional-scale modeling studies of pollen dispersion  
44 been conducted for Europe, and they have been used to assess the impacts of climate change on airborne pollen  
45 distributions (Sofiev and Prank 2016; Lake et al. 2017). In contrast to most meteorological pollen models, climate  
46 models require long-term (e.g., decadal to century scale) emissions at a range of resolutions covering continental  
47 regions up to the global scale. This distinction in both time and space requires a flexible model that can account for  
48 emissions without taxon-specific emission data (i.e. differentiation between genera or species) and can be used  
49 within aggregated vegetation descriptions, such as plant functional types (PFTs). Given recent interest in airborne  
50 biological particles and their role in climate (Despres et al. 2012; Myriokefalitakis et al. 2017), an emissions model  
51 that captures longer temporal scales and broader spatial scales is key to developing global inventories and  
52 understanding pollen's role in the climate system. Here we develop a model for use in the climate modeling  
53 community that can be used specifically to simulate pollen emissions on the decadal or centurial time scale for large  
54 regions using conventional climate or Earth system models.

55 Existing pollen forecasting models are often classified as either process-based phenological models or observation-  
56 based models (Scheifinger et al. 2013). Process-based phenological models employ a parameterization of plant  
57 physiology and climatic conditions (e.g., relating the timing of flowering to a chilling period, photoperiod, or water  
58 availability). Pollen season phenology in an anemophilous species is inherently connected to its environment via  
59 relationships in the growing season dynamics (e.g. bud burst and temperature, (Fu et al. 2012)), and many models  
60 apply the same techniques to flowering as for bud burst (Chuine et al. 1999). This approach to phenology could be  
61 suited to climate models, given its flexibility for adaptive phenological events and regional-scale studies. Typically,  
62 these types of phenological models are taxa specific as well as regionally dependent, e.g., *Betula* in Europe or

63 ragweed in California (Sofiev et al. 2013; Siljamo et al. 2013; R. Zhang et al. 2014). These models are usually  
64 calibrated to local data only even though distinct geographic differences exist for pollen phenology. Thus, such  
65 models may not perform equally well in other locations. Though process-based models draw a connection between  
66 an atmospheric state variable, i.e. temperature, and pollen emissions, at least three parameters are required for  
67 optimization and they are susceptible to overfitting (Linkosalo et al. 2008). While some process-based models may  
68 be scaled up to larger regions while maintaining appreciable accuracy (García-Mozo et al. 2009), such models are  
69 generally not practical for implementation in larger-scale climate modeling with regional climate models (RCMs)  
70 and global climate models (GCMs) because sufficient land cover data is not available at the appropriate taxonomic  
71 level.

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

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

92 In addition to understanding the release of pollen grains, a second consideration is the large-scale transport of pollen.  
93 Once emitted to the atmosphere, pollen is mixed within the atmospheric boundary layer by turbulence, and  
94 depending on large-scale conditions, can be transported far from the emission source. Prior studies have used both  
95 Lagrangian (Hunt et al. 2002; Hidalgo et al. 2002) and Eulerian techniques to simulate the transport of pollen, with  
96 the former typically used for studies of crop germination and the latter primarily for allergen forecasting. For  
97 example, *Helbig et al.* (2004) used the meteorological model KAMM (Karlsruher Meteorologisches Modell) with  
98 the DRAIS (Dreidimensionales Ausbreitungs- und Immissions-Simulationsmodell) turbulence component to  
99 simulate daily pollen counts for region over Europe. *Schueler and Schlünzen* (2006) use a mesoscale atmospheric

100 model (METRAS) to quantify the release, transport and deposition of oak pollen for a two-day period over Europe.  
101 *Sofiev et al.* (2013) includes the long-range transport of birch pollen over Western Europe by developing a birch  
102 pollen map and a flowering model to trigger release in the Finnish System for Integrated modeling of Atmospheric  
103 coMposition (SILAM). *Efstathiou et al.* (2011) developed a pollen emissions model for use within the regional air  
104 quality model (the Community Multi-scale Air Quality model (CMAQ)), and tested their model with birch and  
105 ragweed taxa. *Zhang et al.* (2014) implements a similar pollen emissions scheme with a regional numerical weather  
106 prediction model (the Weather Research and Forecasting (WRF) modeling system). *Zink et al.* (2013) developed a  
107 generic pollen modeling parameterization for use with a numerical weather prediction model (COSMO-ART) that is  
108 flexible to include differing pollen taxa. Collectively, these relatively new developments suggest a growing interest  
109 in the prognostic estimation of pollen on the short-term for seasonal allergen forecasting on the weather (e.g., one to  
110 two weeks) time scale.

111 In this manuscript, we build on these coupled emissions-transport models and develop a comprehensive emissions  
112 model (Pollen Emissions for Climate Models; PECM) for use at climate model time scales that covers the majority  
113 of pollen sources in sub-tropical to temperate climates, including woody plants, grasses and ragweed. First, we  
114 summarize the spatial distribution and seasonality of pollen counts for various taxa in the United States based on  
115 current observations (Section 2). Then we develop new pollen emissions parameterization for climate studies  
116 (Section 4), transport these emissions over the continental United States (CONUS) using the Regional Climate  
117 Model version 4 (RegCM4) (Giorgi et al. 2012), and evaluate the results using eight years of observed pollen count  
118 data (Section 5). We implement two different land cover classification schemes to illustrate the uncertainties  
119 associated with vegetation representation for trees including: (1) detailed family- or genus- level tree distributions  
120 over CONUS, and (2) the use of plant functional type (PFT) level distributions, which groups vegetation types by  
121 physiological characteristics (Section 3). As the latter provides a greater opportunity for expansion into regional and  
122 global scale climate models over multiple domains, we discuss the effects that the PFT-based categorization has on  
123 the total estimated source strength of pollen. Finally, the limitations of this emissions framework and suggestions for  
124 future developments are included (Section 6).

## 125 **2 Observed pollen Phenology**

### 126 **2.1 Data description**

127 The National Allergy Bureau (NAB) of the American Academy of Allergy, Asthma and Immunology (AAAAI)  
128 conducts daily pollen counts at 96 sites in cities across the United States (US), its territories and several locations in  
129 southern Canada. All NAB sites implement a volumetric air sampler and certified pollen count experts to conduct  
130 daily pollen counts (grains m<sup>-3</sup>) for up to 42 plant taxa at either the family level (e.g., Cupressaceae, Poaceae), genus  
131 level (e.g., *Acer*, *Quercus*), or for four generic categories termed “Other Grass Pollen,” “Other Tree Pollen,” “Other  
132 Weed Pollen” or “Unidentified.” We use NAB pollen count data ranging from 2003-2010 at all stations in the  
133 continental United States (Figure 1) for selected taxa to develop and evaluate PECM, and to determine the  
134 phenology of wind-driven pollen. Individual station locations and descriptions are included in Table S1.

135 We evaluate the observed pollen counts to determine the vegetation types that emit the largest magnitude of pollen  
136 over the continental United States. Since many of the taxa reported at the 96 NAB sites frequently have very low  
137 pollen counts (e.g., less than 10 grains m<sup>-3</sup>), a threshold for the grain count is set to select the taxa with the highest  
138 pollen counts. We calculate the average of the annual maximum pollen count across all years (2003-2010),  $P_{avgmax}$ ,  
139 at each site for each counted taxon. We then select taxa to include in PECM using two criteria: (1) the maximum of  
140  $P_{avgmax}$  among all stations exceeds 100 grains m<sup>-3</sup>, and (2) the average  $P_{avgmax}$  among all stations exceeds 70 grains m<sup>-3</sup>  
141 (Table S2). Using these two criteria, 13 taxa are selected for inclusion in the model, including *Acer*, *Alnus*,  
142 *Ambrosia*, *Betula*, Cupressaceae, *Fraxinus*, Poaceae, *Morus*, Pinaceae, *Platanus*, *Populus*, *Quercus* and *Ulmus*.  
143 These thirteen taxa account for about 77% of the total pollen counted across the United States during 2003-2010.  
144 The 13 dominant pollen types are grouped into four main categories by plant functional type: deciduous broadleaf  
145 forest (DBF), evergreen needle-leaf forest (ENF), grasses (GRA) and ragweed (RAG). Plant functional type is a  
146 land cover classification commonly used in the land surface component of climate models, and this categorization  
147 will allow flexibility to apply the emissions model to other climate models. The DBF category includes 9 genus-  
148 level taxa (*Acer*, *Alnus*, *Betula*, *Fraxinus*, *Morus*, *Platanus*, *Populus*, *Quercus*, and *Ulmus*) and the ENF category  
149 includes two family-level taxa (Cupressaceae and Pinaceae). The grass PFT utilizes pollen count data from the  
150 Poaceae family, although we note that the grass PFT classification may include herbs and other non-woody species  
151 that may emit pollen as well. *Ambrosia* (ragweed) is segregated as its own category (RAG), due to its high pollen  
152 counts in the early autumn and unique land cover features. Daily pollen counts were summed for each PFT prior to  
153 calculating an 8-year average pollen time series.

154

## 155 **2.2 Observed seasonality of pollen emissions**

156 Pollen counts are analyzed over five subregions based on their climatic differences (Figure 1; Table S1) to identify  
157 emissions patterns over the continental United States. These five subregions are the Northeast (temperate; 38°-48°N  
158 and 70°-100°W; 34 stations), the Southeast (temperate, subtropical; 25-38°N and 70°-100°W; 29 stations), Mountain  
159 (varied climate; 25°-48°N and 100°-116°W; 9 stations), California (Mediterranean, varied climate; 25°-40°N and  
160 116°-125°W; 13 stations) and the Pacific Northwest (temperate rainforest; west of 116°W and north of 40°N; 4  
161 stations). Figure 2 shows the observed average PFT daily pollen counts averaged over all stations within the defined  
162 subregions.

163 For deciduous broadleaf forest (DBF) taxa, the Southeast has the highest average pollen maximum reaching up to  
164 about 700-1200 grains m<sup>-3</sup> around day 100. In the Northeast, DBF is the dominant PFT, reaching up to an average  
165 of 400 grains m<sup>-3</sup> and peaking slightly later (around day 120) than the Southeast. California sites show an average  
166 peak around 150 grains m<sup>-3</sup> occurring slightly earlier around day 80. A sharp maximum of 775 grains m<sup>-3</sup> appears in  
167 the Mountain subregion at about day 80, with a secondary emission reaching around 150 grains m<sup>-3</sup> on day 125. In  
168 the Northwest, DBF pollen has the earliest maximum (day 70) at about the same magnitude as California (~200  
169 grains m<sup>-3</sup>). In some locations, there is a secondary DBF peak in the late summer and early fall due to the late  
170 flowering of *Ulmus crassifolia* and *Ulmus parvifolia*, located predominantly in the Southeast and California (Lewis,

171 et al.1983). In the Southeast this occurs between days 225 and 300, while in California this occurs twice around day  
172 245 and day 265.

173 The two ENF families exhibit pollen release at two distinct but overlapping times, with Cupressaceae peaking before  
174 Pinaceae. Cupressaceae in the Southeast emits pollen earlier than in other subregions, with a maxima at just over  
175 400 grains  $m^{-3}$  around day 10 and counts above 200 grains  $m^{-3}$  in December of the prior year. Cupressaceae  
176 dominates the total emissions for the Southeast, with a smaller maximum from Pinaceae of about 180 grains  $m^{-3}$   
177 near day 110. In the Northeast, the bimodality of ENF is evident with the Cupressaceae family reaching a  
178 maximum of 100 grains  $m^{-3}$  near day 85 with a secondary Pinaceae maximum approximately 65 days later at about  
179 half the magnitude ( $\sim 50$  grains  $m^{-3}$ ). In the Mountain and Pacific Northwest subregions, the maximum occurs around  
180 day 50-80 and can reach up to 350 grains  $m^{-3}$  in the Mountain subregion, but in both subregions is generally much  
181 lower than the eastern United States (approximately 50 grains  $m^{-3}$ ). In the California subregion, ENF emissions are  
182 comparatively low ( $< 50$  grains  $m^{-3}$ ) which is likely due to the bias in sampling locations.

183 The grasses (Poaceae) have comparatively low average pollen counts ( $< 25$  grains  $m^{-3}$ ) throughout the season in all  
184 subregions except the Northwest, where the maximum reaches 75 grains  $m^{-3}$ . However, the average maximum  
185 Poaceae pollen count at individual stations is close to 100 grains  $m^{-3}$ , with the individual annual maxima reaching  
186 several hundreds of pollen grains  $m^{-3}$ . In the AAAAI data, there are two distinct maxima in the Northeast Poaceae  
187 count, and we attribute the first seasonal maximum to  $C_3$  grasses (peak around day 155) and the second grass  
188 maximum mainly to  $C_4$  grasses (peak around day 250). Observations by *Craine et al.* (2011) of Poaceae in an  
189 American prairie have indicated that  $C_3$  and  $C_4$  grass flowering occurs at distinctly different times, with  $C_3$  in the  
190 late spring and  $C_4$  in mid- to late summer. Similarly, *Medek et al.* (2016) observed two grass pollen peaks in  
191 Australia, with a stronger, late-summer peak at lower Southern latitudes where there is higher incidence of  $C_4$  grass.  
192 However, the authors note that sometimes this may be due to a second flowering of some  $C_3$  grass species. Although  
193 the  $C_3$ - $C_4$  separation cannot be confirmed in the AAAAI pollen count data because they are not distinguished during  
194 pollen identification, this distinction is included in the model as discussed in Sections 3.1 and 4.2 below. In the  
195 Southeast, this separation of the Poaceae pollen counts is less apparent because both of the emission maxima are  
196 broader and intersect one another. In the Southeast, the first observed pollen maximum (assessed as  $C_3$  grass pollen)  
197 peaks earlier around day 140, while the second maximum (assessed as  $C_4$  grasses) have a similar, yet smaller value  
198 around day 250. In the Mountain subregion, the first grass maximum occurs later in the year (day 175) and the  
199 second grass maximum occurs around day 250 in the late summer. Pollen counts in California are only substantial  
200 during the earlier flowering time ( $C_3$  grasses) and have a similar duration to the Northeast, peaking at around day  
201 135. For the Pacific Northwest, there is one strong early peak of grass pollen in the middle of the summer (day 170)  
202 and a secondary maximum is negligible, although counts below 10 grains  $m^{-3}$  register around days 250-270.

203 Ragweed (*Ambrosia*) pollen is segregated from other grasses and herbs because of the strong allergic response in  
204 humans to this specific species and the unique timing of emissions. Because it is a short-day plant (i.e. its  
205 phenology driven by a shortening photoperiod and cold temperatures (Deen et al. 1998)), ragweed pollen seasons  
206 are generally constrained to the late summer with the exception of the Mountain region where some counts occur in  
207 the spring. Emissions in the Northeast reach a maximum around day 240 at 60 grains  $m^{-3}$  while they occur slightly

208 later in the Southeast, peaking around day 270 with twice the magnitude (120 grains m<sup>-3</sup>). Ragweed pollen in the  
209 Mountain subregion with an expected peak at around day 245, but also an earlier peak at around day 130 with no  
210 confirmed cause. *Ambrosia* is not detected in the station averages for California and the Pacific Northwest, although  
211 some individual sites in these regions record relatively low counts on the order of 10 grains m<sup>-3</sup>.  
212

### 213 **3 Model input data**

#### 214 **3.1 Land cover data**

215 With a goal of developing regional to global pollen emissions, one of the greatest limitations is the description of  
216 vegetation at the appropriate taxonomic level and spatial resolution. While land cover databases specific to species  
217 level are available for some regions, they are not available globally. Alternatively, vegetation land cover in regional  
218 to global models can be represented by classifications based on biophysical characteristics. For climate models, a  
219 common approach to represent land cover is with plant functional types (PFT), and global PFT data is readily  
220 available and used by many regional and global climate models to describe a variety of terrestrial emissions  
221 (Guenther et al. 2006) and biophysical processes in land-atmosphere exchange models. The creation of a pollen  
222 emissions model with PFT categorization would be of use at a broad range of spatial scales and domains while  
223 integrating more readily with climate models. In the pollen emissions model development and evaluation (Sections  
224 4 and 5), we compare two different vegetation descriptions of broadleaf deciduous and evergreen needleleaf trees  
225 including (1) family- or genus-specific land cover and (2) land cover categorized by PFT.

226 The Biogenic Emissions Landuse Database version 3 (BELD) provides vegetation species distributions at 1 km  
227 resolution over the continental United States based on satellite imagery, aerial photography and ground surveys, as  
228 well as other land cover classification data such as geographical boundaries (Kinnee et al. 1997;  
229 <https://www.epa.gov/air-emissions-modeling/biogenic-emissions-landuse-database-version-3-beld3>). The BELD  
230 database includes 230 different tree, shrub and crop taxa across the United States as a fraction of the grid cell area at  
231 either the genus or species level. For family and genus level pollen emissions, the BELD land cover fraction for the  
232 11 dominant pollen-emitting tree taxa identified in Section 2.1 is utilized (Table 1; Figure 3). For species level land  
233 cover data, land cover fraction is calculated as the aggregate of all species within a family or genus.

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

239 Figures 4a-d compare the BELD land cover (summed by PFT) and CLM4 land cover for the two tree PFTs. Region  
240 by region comparison of land cover for all BELD taxa and each tree PFT (from both BELD and CLM4) is given in  
241 Table 2. An important distinction is that CLM4 land cover extends beyond U.S. borders because it is derived from a  
242 global dataset, whereas BELD is constrained to the continental United States. BELD and CLM4 land cover show

243 general agreement on the regional distribution of both tree PFTs. DBF is predominantly in the eastern portion of the  
 244 United States with a gap in the Midwestern corn belt. ENF is present in the Southeast, the Northeast along the U.S.-  
 245 Canadian border, along the Cascade and Coastal mountain ranges and throughout the northern Rockies. A notable  
 246 difference is the CLM4 representation of ENF, which shows a strong, dense band extending from the Sierra Nevadas  
 247 through the Canadian Rockies. The BELD ENF broadly covers the Rocky Mountain Range, yet more diffusely (land  
 248 cover percentage up to 76%), whereas the CLM4 dataset shows sparser and dense ENF land cover (e.g., up to 100%)  
 249 in the same range. For the DBF category, another notable difference is that the strong band of oaks around the  
 250 Central Valley of California, which is evident in BELD but missing from the CLM4 data set. Additionally, the  
 251 CLM4 has far greater densities of DBF along the Appalachian range than BELD. Overall, the CLM4 land cover  
 252 fractions for forest PFTs are higher on average than the summed BELD taxa, about 2 to 10 times as much in each  
 253 region, with the exception of California subregion DBF where CLM4 landcover is about half of that in the BELD  
 254 dataset (Table 2).

255 Grass spatial distributions are given by  $C_3$  (non-arctic) and  $C_4$  grass PFT land cover classes from CLM4 (Figure  
 256 4e,f), which correspond to the observed family-level Poaceae pollen subdivided into  $C_3$  and  $C_4$  categories (described  
 257 in Section 2.2).  $C_3$  coverage is evident across the United States, with broad coverage throughout the Southeast,  
 258 Midwest, and northern Great Plains (Fig. 4e).  $C_4$  coverage is concentrated in the Southeast and Southern Great  
 259 Plains at lower densities (Fig. 4f).

260 Ragweed requires a different land cover treatment, as land cover distributions are not available for ragweed across  
 261 the entire continental United States. Ragweed is known to arise in areas of human disturbances (Forman and  
 262 Alexander 1998; Larson 2003), and is found mainly in disturbed or developed areas such as cities and farms (Katz et  
 263 al. 2014; Clay et al. 2006). *Ambrosia* land cover (Figure 4g) is derived from the urban and crop categories of the  
 264 CLM4 land cover, which are sourced from LandScan 2004 (Jackson et al. 2010) and the CLM4 datasets,  
 265 respectively. The urban data is subdivided by urban intensity, which is determined by population density. We  
 266 assume that ragweed is unlikely to grow in the densest of urban areas (such as city centers), and utilize the lowest  
 267 urban density category that is also the most widespread. Ragweed land cover (plants  $m^{-2}$ ) in urban areas is  
 268 determined by multiplying the average urban ragweed stem density given by Katz et al. (2014) by the urban land  
 269 cover fraction. For crops, the CLM4 subdivides land cover fraction into categories including corn and soybean  
 270 crops, and Clay et al. (2006) provide ragweed stem densities in soybean and corn cropland. Thus, we calculate the  
 271 ragweed land cover in stems  $m^{-2}$  ( $f_{rag}$ ):

$$(1) \quad f_{rag} = \alpha(d_{soy}f_{soy} + d_{corn}f_{corn}) + \beta(d_{urb}f_{urb})$$

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

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

281

### 282 **3.2 Meteorological data for phenology**

283 To develop the emissions model, we use two sources of meteorological data. The first is a high-resolution  
284 meteorological dataset to develop the phenological relationships for the timing of pollen release. Because reliable  
285 measurements are not available at all pollen count stations and there is uncertainty in the siting of these stations  
286 (e.g., they may be in urban areas with highly heterogeneous temperature), we use a gridded observational  
287 meteorological product for consistency across all sites (Maurer et al. 2002). The gridded Maurer dataset interpolates  
288 station data to a 1/8° grid across the continental United States on a daily basis, representing a high spatial resolution  
289 gridded data product where data from each meteorological station has been subject to consistent quality control.  
290 Higher resolution DayMet temperatures (daily 1 km) (Thornton et al. 2014) were used in lieu of Maurer data at  
291 NAB sites where the Maurer dataset did not provide information at the collocated grid cell (Table S1). For offline  
292 emission calculations input into the regional climate model, we use annual-average temperatures computed from  
293 monthly Climate Research Unit (CRU) temperature data (Harris et al. 2014). This data was interpolated from a  
294 0.5°x0.5° grid to the 25 km regional climate model grid used for pollen transport.

295

## 296 **4 PECM model description**

### 297 **4.1 Emission potential**

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

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

300 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  
301 (Section 2.1;  $m^2$  vegetated  $m^{-2}$  total area),  $p_{annual}$  is the daily production factor (grains  $m^{-2} yr^{-1}$ ), and  $\gamma_{phen}$  is the  
302 phenological evolution of pollen emissions that controls the release of pollen (description below). Equation 2 can  
303 apply to either a single taxa or PFT, depending on the prescription of land cover through  $f(x,y)$ . In the simulations  
304 described here, emissions are calculated offline based on this equation and provided as input to a regional climate  
305 model (RCM). This emission potential is later adjusted based on meteorological factors in the RCM where the  
306 pollen grains are transported as aerosol tracers (Section 5.1.1). In the future, Equation 2 can be coupled directly  
307 within the climate model for online calculation of emissions. The phenological and production factors are described  
308 in greater detail below.

309 **4.2 Phenological factor ( $\gamma_{phen}$ )**

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

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

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

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

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

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

317 Linear regressions of observed sDOY and eDOY from individual pollen count stations versus temperature are used  
318 to empirically determine sDOY and eDOY that drive  $\gamma_{phen}$ . An important criteria is the grain count used to determine  
319 the sDOY and eDOY, and we utilize a count threshold adaptable to bimodal emission patterns such as those noted  
320 for *Ulmus* and Poaceae. *Sofiev et al.* (2013) selected dates on which the 5th and 95th percentile of the annual index  
321 (annual sum of pollen counts) were reached, while *Liu et al.* (2016) combined a 5 grains m<sup>-3</sup> threshold with the  
322 additional condition that 2.5% (97.5%) of the annual sum of pollen was reached before the start (end) date. Here,  
323 we implement a pollen count threshold of 5 grains m<sup>-3</sup> and found this was sufficient to reproduce the observed  
324 seasonal cycle. To account for smaller signals that may be due to count errors (e.g., an exceedance of the 5 grains m<sup>-3</sup>  
325 threshold but not followed by an increase in emissions), we used a moving window with a threshold of 25 grains  
326 m<sup>-3</sup> for the sum of pollen counts in the nearest 10 neighboring days; when the sum of the neighbors failed to meet  
327 this threshold, the data point was omitted. In this manner, we calculated the sDOY and eDOY for the full 8-year  
328 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  
329 station for a taxon that was not clearly bimodal, only the first set of dates was retained for the linear regression. For  
330 taxa with an observed bi-modal peak, the second peak was treated as a separate taxon (e.g. early and late *Ulmus*, C<sub>3</sub>  
331 and C<sub>4</sub> Poaceae) with a separate phenology. Once the sDOY and eDOY were determined, outliers in these dates  
332 were determined by bounding the data for each taxon at four times the mean absolute deviation of sDOY and eDOY.  
333 Near surface atmospheric temperature (e.g., 2m height) is an important factor of vegetation phenology. In the  
334 interest of having a regional model of emissions that prognostically calculates the start dates, the previous year  
335 annual average temperature (PYAAT) based on near-surface atmospheric temperature from *Maurer et al.* (2002) and  
336 *Thornton et al.* (2014) (Section 2.2) is the explanatory variable in the linear regressions. For example, for a start  
337 date of February 2, 2007, the PYAAT would be the mean temperature for the year 2006. For *Pinus* and  
338 Cupressaceae, PYAAT is calculated differently from July 1, 2005 - June 30, 2006 because emissions of these  
339 families begin in the early winter (December). Prior studies have shown that the meteorology of the year previous  
340 to the pollen season influences pollen production, especially temperature, suggesting that PYAAT may be a good

341 predictor variable (Menzel and Jochner 2016). While emissions in this study are calculated using offline  
342 meteorological data, this also could be coupled to a dynamic land surface model to predict reasonably accurate  
343 pollen phenological dates.

344 To exemplify this method, Figure 5 shows the phenological dates and regression lines for the *Betula* (birch) genus,  
345 with all 13 modeled taxon shown in Figures S1 and S2. The sDOY and eDOY of the pollen season show a moderate  
346 and considerable trend with temperature for most taxa and PFTs (Table 1; Figures S1 & S2). The linear regression  
347 models for sDOY explain 41% of the variance on average for DBF taxa, 47% on average for ENF taxa, 48% for C<sub>3</sub>  
348 Poaceae, and 8% for *Ambrosia* while having a negligible R<sup>2</sup> for C<sub>4</sub> Poaceae. For eDOY, the linear regression models  
349 explain 21% of the variance on average for DBF taxa, 29% for ENF, 4% for C<sub>3</sub> Poaceae, 32% for C<sub>4</sub> Poaceae, and  
350 37% for *Ambrosia*. All trends except C<sub>4</sub> Poaceae, late elm, and *Ambrosia* are negative, indicating that warmer  
351 previous-year temperatures result in earlier start and end dates. For most tree taxa, the trend of both sDOY and  
352 eDOY are negatively correlated with PYAAT, with a steeper negative slope for sDOY. The correlation for the  
353 duration of the pollen season (eDOY – sDOY) is then positive for all taxa except Cupressaceae. This suggests that  
354 warmer climates have earlier pollen season start and end dates but longer season lengths.

355 Trends for grass in Australasia show that the correlation of the end date of the pollen season with average spring  
356 temperature is positive, while the same relationship for the start date is negative, suggesting also that season start  
357 dates are earlier and season duration increases with warmer climates (Medek et al. 2016). The apparent trend in the  
358 season end date for *Ambrosia* with PYAAT could be due to the increased number of frost-free days, consistent with  
359 global warming, and a strong relationship between frost-free days and changes of ragweed season length (Easterling  
360 2002; Ziska et al. 2011).

361 This agrees with earlier findings that suggest the pollen season will, on average, start earlier with a warmer global  
362 climate and have a longer duration (Confalonieri et al. 2007). The spatiotemporal heterogeneity of climate change  
363 may affect which regions and seasons will be most influenced by climate change (Ziska 2016). In fact, there is  
364 imperfect agreement that earlier start dates and longer seasons will occur unanimously throughout the United States  
365 region, at least for trees (Yue et al. 2015). It is understood that photoperiod and the dormancy-breaking process  
366 controlled by chilling temperatures play a significant role in the phenology of trees (Myking and Heide 1995; Ziska  
367 2016), and it is generally accepted that a plethora of other factors, such as plant age, mortality, and nutrient  
368 availability also affect observed phenological dates (Jochner et al. 2013). However, even without these factors, the  
369 current phenological model is applicable to large regions and provides a clear response of plants to inter-annual  
370 climate variability as well as long-term climate changes. For this first assessment of PECM, we assume that the  
371 pollen production factor ( $p_{\text{annual}}$ ) does not change with time and that the phenological model described above  
372 captures the main features of pollen emissions.

### 373 **4.3 Annual pollen production ( $p_{\text{annual}}$ )**

374 Annual production factors (grains m<sup>-2</sup> year<sup>-1</sup>, where m<sup>-2</sup> refers to vegetated area, or grains stem<sup>-1</sup> year<sup>-1</sup> for ragweed)  
375 for each modeled taxon are provided in Table 1. The annual pollen production factor ( $p_{\text{annual}}$ ) defines the amount of  
376 pollen produced per vegetation biomass per year based on literature values. *Tormo Molina et al.* (1996) report the

377 annual pollen productivity in grains tree<sup>-1</sup> year<sup>-1</sup> measured from three representative trees from several taxa. *Morus*  
378 has no known reference for production factor and was assumed to be 10x10<sup>7</sup> grains m<sup>-2</sup> year<sup>-1</sup>, conservatively at the  
379 low end of the range for other deciduous broadleaf taxa. Other tree taxa and grasses are reported in grains m<sup>-2</sup> year<sup>-1</sup>,  
380 while ragweed is reported in grains stem<sup>-1</sup> year<sup>-1</sup> (Helbig et al. 2004; Jato, Rodríguez-Rajo, and Aira 2007; Hidalgo,  
381 Galán, and Domínguez 1999; Prieto-Baena et al. 2003; Fumanal, Chauvel, and Bretagnolle 2007). To convert the  
382 production factors from *Tormo Molina et al.* (1996) (grains tree<sup>-1</sup> year<sup>-1</sup>), the production factors for each  
383 representative tree are multiplied by the tree crown area, calculated as the circular area of the tree crown diameter  
384 given in Table II of *Tormo Molina et al.* (1996). The resulting individual production factors (grains m<sup>-2</sup> year<sup>-1</sup>) are  
385 then averaged for each taxa.

386 After sensitivity experiments of running pollen emissions in RegCM4, we find that the literature value of  
387  $p_{\text{annual}}$  for Poaceae provides better agreement with observations for C<sub>4</sub> grass when reduced by a factor of 10, thus we  
388 use this value. To obtain the coefficient of daily pollen production over the duration of the phenological curve,  $\gamma_{\text{phen}}$ ,  
389 the integral of the daily pollen production is normalized to  $p_{\text{annual}}$  as demonstrated by Equation 2.

#### 390 4.4 Offline emissions simulations

391 We calculate emissions offline for two versions of PECM that differ in the land cover input data for woody plants.  
392 The first uses the detailed BELD tree database (Figure 3) for tree pollen emissions (hereinafter the “BELD”  
393 simulation), and the second uses globally based PFT data for tree pollen emissions (Figures 4b and 4d) (hereinafter  
394 the “PFT” simulation). For the grass and ragweed taxa, the emissions calculations are identical between the two  
395 simulations as the input land cover is the same for these two categories. While the family and genus level is useful  
396 for the allergen community, the respective taxon land cover databases needed to develop a global, adaptable model  
397 are not always available. While many plant traits are found to vary quite strongly within individual PFTs (Reichstein  
398 et al. 2014), the PFT convention is accepted and remains in use in climate models, particularly because of the lack of  
399 species-level land cover data at large scales. For the PFT version, pollen counts from individual taxa were summed  
400 within each PFT prior to calculation of the phenological regression (Table 1). We exclude the bimodality in *Ulmus*  
401 for the PFT version because it is the only tree taxon that exhibits this behavior, and late *Ulmus* pollen emissions are  
402 relatively small compared to the major DBF season. The production factors for each PFT are calculated as the  
403 unweighted average of the production factors for all the taxa within the PFT (Table 1).

404 Figures 6-11 show the monthly averages of the 2003-2010 emissions potential calculated by the offline models  
405 described in Section 4.1 ( $E_{\text{pot}}$ ; Equation 2). The seasonal cycle can be clearly identified in the emissions potential,  
406 with the onset of pollen emissions beginning in the warmer south and moving northward along the gradient of  
407 annual average temperature. Colder locales such as those at high elevations can interrupt this general trend. Though  
408 pollen seasons generally end later in the colder parts of the domain just as they start later, modeled pollen emission  
409 seasons tend to be shorter at colder locations for most taxa (about 1 day per 1°C, on average). The highest maximum  
410 emissions for DBF occur over the Appalachian range between April and May for both the BELD and PFT versions  
411 (Figures 6 and 7). For ENF, the maximum occurs in April in the American West for the BELD version where  
412 Cupressaceae land cover is dominant, while it is consistent in magnitude between the Southeast and West Coast for

413 the PFT-based version (Figures 8 and 9). The grass PFT maximum emissions occur in June in the northern Rockies  
 414 for C<sub>3</sub> and in September in the South-Central Great Plains for C<sub>4</sub> (Figure 10). Ragweed pollen emissions reach their  
 415 maximum during September throughout the Corn Belt where soybean and corn crops dominate the land surface,  
 416 with local maxima apparent in urban centers (Figure 11).

417

## 418 **5 Emissions implementation and evaluation**

### 419 **5.1 Emissions implementation in a regional climate model**

420 To evaluate PECM, emissions calculated offline are included within a regional climate model to compare simulated  
 421 atmospheric pollen concentrations with ground-based observations from the NAB pollen network. The two  
 422 phenological pollen emissions estimates (BELD and PFT) described above are prescribed as daily emissions, after  
 423 which they are scaled by meteorological factors and undergo atmospheric transport. We use the Regional Climate  
 424 Model version 4 (Giorgi et al. 2012), which is a limited-area climate model that includes a coupled aerosol tracer  
 425 module (Solmon et al. 2006) that readily accommodates pollen tracers (Liu et al. 2016). The pollen tracer transport  
 426 scheme is extended from one to four bins in this study to simulate the four PFTs (DBF, ENF, GRA, and RAG), with  
 427 tracer bin particle effective diameters of 28 μm, 40 μm, 35 μm and 20μm, respectively. Additionally, the temporal  
 428 emissions input is updated to accommodate daily pollen emissions (grains m<sup>-2</sup> day<sup>-1</sup>).

429 RegCM4 is based on the hydrostatic version of the Penn State/NCAR mesoscale model MM5 (Grell et al. 1994) and  
 430 configured for long-term climate simulations. In our RegCM4 configuration, we use the Community Land Model  
 431 version 4.5 (CLM4.5; (Oleson et al. 2010)), the Emanuel cumulus precipitation scheme over land and ocean  
 432 (Emanuel 1991), and the SUBEX resolvable scale precipitation (Pal et al. 2000). The horizontal resolution is 25-km  
 433 with 144x243 grid cells on a Lambert Conformal Projection centered on 39°N, 100°W with parallels at 30°N and  
 434 60°N (Figure 1). The vertical resolution includes 18 vertical sigma levels. Boundary conditions are driven by ERA-  
 435 Interim Reanalysis while sea surface temperatures are prescribed from NOAA Optimum Interpolation SSTs (Dee et  
 436 al. 2011; Smith et al. 2008). Two 8-year simulations of pollen emissions and transport in RegCM4 were conducted  
 437 from 2003-2010 with the BELD and PFT version of the offline emissions model. Six months of spin-up (July-  
 438 December 2002) are run for both simulations that we exclude from the following analysis.

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

$$(5) E_{pollen,i}(x, y, t) = E_{pot,i}(x, y, t) f_w f_r f_h$$

$$(6) f_w = 1.5 - e^{-(u_{10} + u_{conv})/5}$$

444

445

$$(7) \quad f_r = \begin{cases} 1, & pr < pr_{low} \\ \frac{pr_{high} - pr}{pr_{high} - pr_{low}}, & pr_{low} < pr < pr_{high} \\ 0, & pr > pr_{high} \end{cases}$$

$$(8) \quad f_h = \begin{cases} 1, & rh < rh_{low} \\ \frac{rh_{high} - rh}{rh_{high} - rh_{low}}, & rh_{low} < rh < rh_{high} \\ 0, & rh > rh_{high} \end{cases}$$

446 where  $f_w$ ,  $f_r$ , and  $f_h$  are the wind, precipitation and humidity factors, respectively. The meteorological parameters in  
 447 these equations are from online RegCM variables, including  $u_{10}$  and  $u_{conv}$  as the 10-meter horizontal wind speed and  
 448 vertical wind speed, and  $pr$  and  $rh$  are precipitation and relative humidity with low and high thresholds. These  
 449 scaling factors account for the effects of wind, precipitation and humidity on the emission of pollen from flowers  
 450 and cones. The humidity and precipitation factors are piecewise linear functions of the near-surface (10 m) RH and  
 451 total precipitation and range from 0 (high precipitation or humidity) to 1 (no precipitation or low humidity). The  
 452 wind factor ranges from 0.5 to 1.5, as even in calm conditions turbulent motions can trigger pollen release with high  
 453 winds releasing more pollen. These scaled emissions are then transported according to the tracer transport equation  
 454 (Equation 9) of *Solmon et al. (2006)* that includes advection, horizontal and vertical diffusion ( $F_H$  and  $F_V$ ),  
 455 convective transport ( $T_c$ ), as well as wet ( $R_{wls}$  and  $R_{wc}$ , representing large scale and convective precipitation  
 456 removal) and dry deposition ( $D_d$ ) of an individual tracer ( $\chi$ ), represented by  $i = 1$  to 4 for each PFT pollen emission:

$$(9) \quad \frac{\partial \chi^i}{\partial t} = \bar{V} \cdot \nabla \chi^i + F_H^i + F_V^i + T_c^i + S^i - R_{wls}^i - R_{wc}^i - D_d^i$$

## 457 5.1 Model evaluation against observations

458 We evaluate the efficacy of PECM in simulating the timing and magnitude of pollen emissions across the  
 459 continental United States by evaluating RegCM4 tracer concentrations versus observations. We compare the average  
 460 daily simulated near-surface pollen counts and observed, ground-based pollen counts for each of the four modeled  
 461 PFTs (Figure 12). The observed pollen time series in Figure 12 are the spatial average of the average daily pollen  
 462 counts at all pollen counting stations comprising each of the five major U.S. subregions (Section 2.2) and are  
 463 compared with the modeled average daily pollen counts, which averages the individual grid cells that contain the  
 464 pollen counting stations. Interannual variability is assessed using the relative mean absolute deviation for each day  
 465 of the average time series. The inter-annual variability in observed daily pollen counts throughout the year is, on  
 466 average, 81, 78, 78 and 77% of the mean (DBF, ENF, grass and ragweed, respectively), while this variability from  
 467 the simulations is 53% for the BELD version of the DBF model and 61% for the PFT version, 55% and 92% for the  
 468 BELD and PFT versions of the ENF model, 43% for grasses, and 49% for ragweed (Figure 12). This indicates that  
 469 the model is capturing the relative inter-annual variability of the pollen counts between PFTs, but not all of the  
 470 variability in pollen counts from season to season. The unexplained variability in pollen concentrations could be due  
 471 to the lack of sensitivity of annual pollen production factor to the environment, as this may be closely tied with  
 472 precipitation (Duhl et al. 2013) or temperature (Jochner et al. 2013). Additionally the average observed and  
 473 simulated pollen counts are analyzed using box-and-whisker plots to assess the models' representivity of pollen

474 count magnitude in spite of phenology (Figure 13). These metrics are discussed in detail by PFT and U.S. subregion  
475 below.

### 476 **5.2.1 DBF**

477 In the Northeast, the BELD model captures both the observed seasonal timing and the magnitude of DBF pollen  
478 counts (Figure 12a). Observed DBF phenology is also simulated by the PFT-based emissions with even greater  
479 statistical accuracy in reproducing the observed pollen counts, though the BELD model more accurately reproduces  
480 the annual maximum (Figure 13a). The accuracy in this subregion is not surprising, as Northeastern pollen counting  
481 stations contributed the greatest number of data points to the phenological regression analyses. Observed DBF  
482 pollen counts in the Southeast have a large maximum that is greater than the average seasonal maximum of all four  
483 other subregions and all three other PFTs (Figure 12b), which is predominantly from *Quercus*. Neither the BELD  
484 nor PFT version of the simulation recreates this sharp peak, but they do simulate a large majority of the pollen count  
485 distribution (Figure 13b), especially the PFT-based model for which the lower 75% of simulated average pollen  
486 counts agrees well with the lower 75% of observed average pollen counts. The PFT model does not specifically  
487 resolve *Quercus*, and while the BELD model does resolve *Quercus*, it fails to model this maximum. This may be  
488 because the linear regression producing the phenological dates is an average, where a longer season may result from  
489 earlier start dates and/or later end dates that will reduce the maximum of the Gaussian distribution of pollen counts  
490 in the time series. In the Mountain region, there is an observed maximum early in the spring that is not simulated by  
491 either model because the DBF phenology at several cold Mountain sites is exceptionally early, and falls well below  
492 the regression lines (Figures S1, S2). However, both the BELD and PFT model simulate the second Mountain  
493 subregion peak with the correct magnitude. The BELD simulated maximum DBF in California is about 40 days later  
494 than the observed peak, also due to the regionally anomalous phenology in California as compared with the rest of  
495 the U.S., and though the PFT model peaks much closer to the observations, it underestimates DBF pollen counts. In  
496 the Pacific Northwest, the observed pattern is quite similar to the DBF pollen phenology in the Mountain subregion  
497 with only a slightly weaker early spring peak due to low-elevation pollen. The observed phenological pattern (Fig.  
498 12e) and pollen count magnitudes (Fig. 12e) are both more accurately simulated by the BELD model, likely due to  
499 the earlier spring maximum that does not appear in the PFT simulation.

### 500 **5.2.2 ENF**

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

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

### 520 **5.2.3 Grasses**

521 Grass phenology across all subregions for both C<sub>3</sub> and C<sub>4</sub> types is captured by the emissions estimates (Figure 12 k-  
522 o). However, the pollen count magnitude in Northeastern C<sub>3</sub> grass peak is overestimated by about a factor of seven,  
523 even when using the minimum value of the annual production factor in the range estimated by *Prieto-Baena* (2003)  
524 (Figure 12k). The secondary peak, which we attribute to C<sub>4</sub> grasses and is only about half as large, is well-  
525 represented. In the Southeast, the simulated pollen count magnitudes are much closer to observations, while the C<sub>3</sub>  
526 peak is overestimated here by only a factor of 2 and the C<sub>4</sub> peak is within 5 grains m<sup>-3</sup> (Figure 12l). In this region,  
527 the observed duration of the pollen emissions is not fully captured by the simulated grass phenology in the  
528 Southeast, and this is probably due to the non-Gaussian shape of the observed time series. In the Mountain  
529 subregion, the C<sub>3</sub> pollen count is overestimated by the model, but the phenology is represented by a gradual rise in  
530 low emissions beginning in March to match the maximum burst of emissions in June (Figure 12m). C<sub>4</sub> grass pollen  
531 counts are not simulated in the Mountain region due to the relatively low C<sub>4</sub> land cover in the CLM4 dataset (Figure  
532 4f). In California there is a single observed grass peak, which the model attributes to C<sub>3</sub> pollen, and the peak count  
533 in the simulation is about 5 days late and about 2 to 10 times too large (Figure 12n). In the Pacific Northwest, the  
534 average C<sub>3</sub> season is accurately simulated with the exception that the phenology is shifted 20 days earlier than  
535 observed (Figure 12o). A small C<sub>4</sub> peak in the observations at around day 260 is not simulated in this region due to  
536 negligible land cover for C<sub>4</sub> grasses in the CLM4 land cover data (Figure 4f).

### 537 **5.2.4 Ragweed**

538 Simulated ragweed phenology in the Northeast, Southeast, and Mountain subregions follows the observed  
539 phenology of late-summer ragweed very closely, where the peaks of both the simulated and the observed time series  
540 averages occur within a day of each other (Figures 12p-t). Close evaluation of each regional phenological time series  
541 reveals that many of the observed features, like those determined by the rate of increase or decrease of the pollen  
542 count, are reproduced by the model. The magnitude of the modeled ragweed maxima in the Northeast and Mountain  
543 subregions is slightly greater than observed (Figures 12p and 12r), while there is a clear underestimation by a factor

544 of 4 or 5 in the Southeast (Figure 12q). There is a yet unidentified observed spring peak of ragweed pollen at about  
545 day 125 in the Mountain subregion, possibly due to an identification error. The observed average ragweed pollen  
546 counts in California and the Pacific Northwest are negligible, though the simulation predicts them to be similar in  
547 magnitude and timing to the other three subregions (Figure 12s and 12t). These discrepancies may be due to the  
548 land use description developed for ragweed (Section 3.1), which may overestimate the ragweed potential in the  
549 western United States, or potentially the relatively sparse observational stations in these regions may be poorly placed  
550 relative to emissions sources.  
551

## 552 **6 Conclusions**

553 We have developed a climate-flexible pollen emissions model (PECM) for the 13 most prevalent wind-pollinating  
554 taxa in the United States based on observed pollen counts. PECM was adapted to the PFT categorization common to  
555 climate and Earth system models with four major temperate-zone PFTs (DBF, ENF, grasses and ragweed), thus it is  
556 possible to apply this model to larger geographic regions where specific taxon-level data is unavailable. We  
557 evaluated PECM using a regional climate model (RegCM4) to transport emissions and evaluated resulting pollen  
558 counts versus observations. PECM generally captures the observed phenology, and observed surface pollen  
559 concentrations can be simulated within an order of magnitude. While many emissions models to date have focused  
560 on smaller geographical regions with more detailed land cover information and pollen information, this model  
561 represents the first of its kind to simulate multiple taxa over broad spatial areas. This transition to a larger scale does  
562 have its disadvantages, and we define several major sources of uncertainty to consider when scaling up pollen  
563 emissions to the regional or global scale: (1) pollen production factors, (2) climatic sensitivities in phenological  
564 timing, (3) land cover data, and (4) taxa specificity. We discuss each of these uncertainties in greater detail.

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

579 Second, the use of observed relationships between pollen count and temperature to determine the phenological  
580 pollen start and end date also adds uncertainty to our modeling framework. Firstly, we assume stationarity in the  
581 phenological relationships, and this assumption may be violated. Secondly, based on the subregions defined for the  
582 analysis, there appears to be a bias in the linear regressions toward subregions with more available pollen counting  
583 stations, therefore affecting performance differences in these regions. Lastly, even though generally the Gaussian  
584 time series model of the pollen phenology performs well in our analysis, in the PFT representation the Gaussian  
585 absorbs or misses some of the phenological details in the observed pollen seasonality, and in some cases taxa (e.g.  
586 grasses in the Southeast subregion) may not be captured by the existing phenology.

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

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

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

610

## 611 **7 Code and Data Availability**

612 Source code for Pollen Emissions for Climate Models (PECM) is written as FORTRAN90 (\*.f90) and available in  
613 the supplementary material as plain text. Input data is explained in Section 3 of this manuscript.

614

615 **8 Acknowledgements**

616 This work was supported by NSF Grant to AGS 0952659 to ALS. We thank Melissa Zagorski and Yang Li of  
617 University of Michigan for contributions to the emissions model development, and Fabien Solmon and Li Liu of the  
618 International Centre for Theoretical Physics for RegCM support and their prior work. We gratefully acknowledge  
619 the use of the American Academy of Allergy Asthma and Immunology (AAAAI) pollen count data, with individual  
620 station acknowledgments fully provided in Table S1.

621

622 **References**

- 623
- 624 Beggs, Paul J, Branko Šikoparija, and Matt Smith. 2017. “Aerobiology in the International Journal of  
625 Biometeorology, 1957–2017.” *International Journal of Biometeorology*, 1957–2017. Accessed August 8.  
626 doi:10.1007/s00484-017-1374-5.
- 627 Box, George E.P., Gwilym M. Jenkins, and Gregory C. Reinsel. 1994. *Time Series Analysis: Forecasting and*  
628 *Control*. Upper Saddle River, NJ: Prentice Hall. doi:10.1016/j.ijforecast.2004.02.001.
- 629 Cecchi, Lorenzo. 2014. “Introduction.” In *Allergenic Pollen: A Review of the Production, Release, Distribution and*  
630 *Health Impacts*, edited by Mikhail Sofiev and Karl-Christian Bergmann, 1–7. New York, London: Springer  
631 Science+Business Media Dordrecht. doi:10.1007/978-94-007-4881-1.
- 632 Chuine, Isabelle, P. Cour, and D. D. Rousseau. 1999. “Selecting Models to Predict the Timing of Flowering of  
633 Temperate Trees: Implications for Tree Phenology Modelling.” *Plant, Cell and Environment* 22 (1): 1–13.  
634 doi:10.1046/j.1365-3040.1999.00395.x.
- 635 Clay, S A, B Kreutner, D E Clay, C Reese, J Kleinjan, and F Forcella. 2006. “Spatial Distribution, Temporal  
636 Stability, and Yield Loss Estimates for Annual Grasses and Common Ragweed (*Ambrosia Artimisiifolia*) in a  
637 Corn/soybean Production Field over Nine Years.” *Weed Science* 54 (2): 380–90.  
638 [http://www.scopus.com/inward/record.url?eid=2-s2.0-](http://www.scopus.com/inward/record.url?eid=2-s2.0-33750295452&partnerID=40&md5=6931902680ebbd9303c2bd99f3ddc4a7)  
639 [33750295452&partnerID=40&md5=6931902680ebbd9303c2bd99f3ddc4a7](http://www.scopus.com/inward/record.url?eid=2-s2.0-33750295452&partnerID=40&md5=6931902680ebbd9303c2bd99f3ddc4a7).
- 640 Confalonieri, U., B. Menne, R. Akhtar, K.L. Ebi, M. Hauengue, R.S. Kovats, B. Revich, and A. Woodward. 2007.  
641 “Human Health.” In *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working*  
642 *Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by M.L.  
643 Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden, and C.E. Hanson, 391–431. Cambridge, UK:  
644 Cambridge University Press.
- 645 Craine, Joseph M, Elizabeth M Wolkovich, E Gene Towne, and Steven W Kembel. 2011. “Flowering Phenology as  
646 a Functional Trait in a Tallgrass Prairie.”
- 647 Dee, D P, S M Uppala, A J Simmons, P Berrisford, P Poli, S Kobayashi, U Andrae, et al. 2011. “The ERA-Interim  
648 Reanalysis: Configuration and Performance of the Data Assimilation System.” *Quarterly Journal of the Royal*  
649 *Meteorological Society Q. J. R. Meteorol. Soc* 137: 553–97. doi:10.1002/qj.828.
- 650 Deen, William, Tony Hunt, Clarence J Swanton, and William Deen. 1998. “Influence of Temperature , Photoperiod  
651 , and Irradiance on the Phenological Development of Common Ragweed (*Ambrosia Artemisiifolia*).” *Weed*  
652 *Science* 46 (5): 555–60.
- 653 Despres, Viviane R., J. Alex Huffman, Susannah M. Burrows, Corinna Hoose, Aleksandr S. Safatov, Galina Buryak,  
654 Janine Frohlich-Nowoisky, et al. 2012. “Primary Biological Aerosol Particles in the Atmosphere: A Review.”  
655 *Tellus, Series B: Chemical and Physical Meteorology*. doi:10.3402/tellusb.v64i0.15598.
- 656 Duhl, T. R., R. Zhang, a. Guenther, S. H. Chung, M. T. Salam, J. M. House, R. C. Flagan, et al. 2013. “The  
657 Simulator of the Timing and Magnitude of Pollen Season (STaMPS) Model: A Pollen Production Model for  
658 Regional Emission and Transport Modeling.” *Geoscientific Model Development Discussions* 6 (2): 2325–68.

659 doi:10.5194/gmdd-6-2325-2013.

660 Easterling, David R. 2002. “Recent Changes in Frost Days and the Frost-Free Season in the United States.” *Bulletin*  
661 *of the American Meteorological Society* 83: 1327–1332. [http://journals.ametsoc.org/doi/pdf/10.1175/1520-](http://journals.ametsoc.org/doi/pdf/10.1175/1520-0477%282002%29083%3C1327%3ARCIFDA%3E2.3.CO%3B2)  
662 [0477%282002%29083%3C1327%3ARCIFDA%3E2.3.CO%3B2](http://journals.ametsoc.org/doi/pdf/10.1175/1520-0477%282002%29083%3C1327%3ARCIFDA%3E2.3.CO%3B2).

663 Efstathiou, Christos, Sastry Isukapalli, and Panos Georgopoulos. 2011. “A Mechanistic Modeling System for  
664 Estimating Large-Scale Emissions and Transport of Pollen and Co-Allergens.” *Atmospheric Environment* 45:  
665 2260–76. doi:10.1016/j.atmosenv.2010.12.008.

666 Emanuel, Kerry. A. 1991. “A Scheme for Representing Cumulus Convection in Large-Scale Models.” *Journal of*  
667 *the Atmospheric Sciences* 48 (21): 2313–35. <ftp://texmex.mit.edu/pub/emanuel/PAPERS/convect91.pdf>.

668 Emberlin, J., J. Mullins, J. Corden, S. Jones, W. Millington, M. Brooke, and M. Savage. 1999. “Regional Variations  
669 in Grass Pollen Seasons in the UK, Long-Term Trends and Forecast Models.” *Clinical and Experimental*  
670 *Allergy* 29 (3): 347–56. doi:10.1046/j.1365-2222.1999.00369.x.

671 Forman, Richard T T, and Lauren E Alexander. 1998. “ROADS AND THEIR MAJOR ECOLOGICAL EFFECTS.”  
672 *Annu. Rev. Ecol. Syst* 29: 207–31.

673 Fu, Yongshuo H., Matteo Campioli, Gaby Deckmyn, and Ivan A. Janssens. 2012. “The Impact of Winter and Spring  
674 Temperatures on Temperate Tree Budburst Dates: Results from an Experimental Climate Manipulation.”  
675 *PLoS ONE* 7 (10). doi:10.1371/journal.pone.0047324.

676 Fumanal, Boris, Bruno Chauvel, and Francois Bretagnolle. 2007. “ESTIMATION OF POLLEN AND SEED  
677 PRODUCTION OF COMMON RAGWEED IN FRANCE.” *Ann Agric Environ Med* 14: 233–36.  
678 doi:10.1093/annonc/mdw163.

679 Galán, C, García-, H Mozo, L Vázquez, L Ruiz, C Guardia, and Díaz. 2008. “Modeling Olive Crop Yield in  
680 Andalusia, Spain.” *E Agronomy Journal* 100 (1).  
681 <http://search.proquest.com/docview/194514400/fulltextPDF/5D1BBD9891A04649PQ/1?accountid=14667>.

682 García-Mozo, H., C. Galán, J. Belmonte, D. Bermejo, P. Candau, C. Díaz de la Guardia, B. Elvira, et al. 2009.  
683 “Predicting the Start and Peak Dates of the Poaceae Pollen Season in Spain Using Process-Based Models.”  
684 *Agricultural and Forest Meteorology* 149 (2): 256–62. doi:10.1016/j.agrformet.2008.08.013.

685 Georg a, Grell, Jimmy Dudhia, and David R Stauffer. 1994. “A Description of the Fifth-Generation Penn State/NCAR  
686 Mesoscale Model (MM5).” *NCAR Technical Note NCAR/TN-398+STR*, no. December: 121.  
687 doi:10.5065/D60Z716B.

688 Giorgi, F., E. Coppola, F. Solmon, L. Mariotti, M. B. Sylla, X. Bi, N. Elguindi, et al. 2012. “RegCM4: Model  
689 Description and Preliminary Tests over Multiple CORDEX Domains.” *Climate Research* 52 (1): 7–29.  
690 doi:10.3354/cr01018.

691 Guenther, A., T. Karl, P. Harley, C. Wiedinmyer, P. I. Palmer, and C. Geron. 2006. “Estimates of Global Terrestrial  
692 Isoprene Emissions Using MEGAN (Model of Emissions of Gases and Aerosols from Nature).” *Atmospheric*  
693 *Chemistry and Physics* 6: 3181–3210. doi:10.1016/j.cognition.2008.05.007.

694 Harris, I, P D Jones, T J Osborn, and D H Lister. 2014. “Updated High-Resolution Grids of Monthly Climatic  
695 Observations – the CRU TS3 . 10 Dataset” 642 (May 2013): 623–42. doi:10.1002/joc.3711.

696 Helbig, Nora, Bernhard Vogel, Heike Vogel, and Franz Fiedler. 2004. "Numerical Modelling of Pollen Dispersion  
697 on the Regional Scale." *Aerobiologia* 3: 3–19.

698 Hidalgo, Pablo J, Carmen Galán, and Eugenio Domínguez. 1999. "Pollen Production of the Genus Cupressus."  
699 *Grana* 38 (5): 296–300. doi:10.1080/001731300750044519.

700 Hidalgo, Pablo J, Antoine Mangin, Carmen Galán, Odile Hembise, Luis M Vázquez, and Oscar Sanchez. 2002. "An  
701 Automated System for Surveying and Forecasting Olea Pollen Dispersion." *Aerobiologia* 18: 23–31.

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

704 Jackson, Trisha L, Johannes J Feddema, Keith W Oleson, Gordon B Bonan, and John T Bauer. 2010.  
705 "Parameterization of Urban Characteristics for Global Climate Modeling."  
706 doi:10.1080/00045608.2010.497328.

707 Jato, Victoria, F. Javier Rodríguez-Rajo, and M. Jesús Aira. 2007. "Use of Phenological and Pollen-Production Data  
708 for Interpreting Atmospheric Birch Pollen Curves." *Annals of Agricultural and Environmental Medicine* 14  
709 (2): 271–80.

710 Jochner, Susanne, Josef Hofler, Isabelle Beck, Axel Gottlein, Donna Pauler Ankerst, Claudia Traidl-Hoffmann, and  
711 Annette Menzel. 2013. "Nutrient Status: A Missing Factor in Phenological and Pollen Research?" *Journal of*  
712 *Experimental Botany* 64 (7): 2081–92. doi:10.1093/jxb/ert061.

713 Katz, Daniel S W, Benjamin T. Connor Barrie, and Tiffany S. Carey. 2014. "Urban Ragweed Populations in Vacant  
714 Lots: An Ecological Perspective on Management." *Urban Forestry and Urban Greening* 13 (4). Elsevier  
715 GmbH.: 756–60. doi:10.1016/j.ufug.2014.06.001.

716 Kinnee, Ellen, Chris Geron, and Thomas Pierce. 1997. "United States Land Use Inventory For Estimating Biogenic  
717 Ozone Precursor Emissions." *Ecological Applications* 7 (1): 46.

718 Kuparinen, Anna, Tiina Markkanen, Hermann Riikonen, and Timo Vesala. 2007. "Modeling Air-Mediated  
719 Dispersal of Spores, Pollen and Seeds in Forested Areas." doi:10.1016/j.ecolmodel.2007.05.023.

720 Lake, Iain R., Natalia R. Jones, Maureen Agnew, Clare M. Goodess, Filippo Giorgi, Lynda Hamaoui-Laguel,  
721 Mikhail A. Semenov, et al. 2017. "Climate Change and Future Pollen Allergy in Europe." *Environmental*  
722 *Health Perspectives* 125 (3): 385–91.  
723 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5332176/pdf/EHP173.pdf>.

724 Larson, Diane L. 2003. "Native Weeds and Exotic Plants: Relationships to Disturbance in Mixed-Grass Prairie."  
725 *Plant Ecology* 169 (2): 317–33. doi:10.1023/A:1026046810307.

726 Lawrence, Peter J., and Thomas N. Chase. 2007. "Representing a New MODIS Consistent Land Surface in the  
727 Community Land Model (CLM 3.0)." *Journal of Geophysical Research: Biogeosciences* 112 (1).  
728 doi:10.1029/2006JG000168.

729 Lewis, Walter H., Prathibha. Vinay, and Vincent E. Zenger. 1983. *Airborne and Allergenic Pollen of North*  
730 *America*. Baltimore: Johns Hopkins University Press. <https://mirlyn.lib.umich.edu/Record/000779939>.

731 Linkosalo, Tapio, Hanna K Lappalainen, and Pertti Hari. 2008. "A Comparison of Phenological Models of Leaf Bud  
732 Burst and Flowering of Boreal Trees Using Independent Observations." *Tree Physiology* 28 (12): 1873–82.

733 doi:10.1093/treephys/28.12.1873.

734 Liu, Li, Fabien Solmon, Robert Vautard, Lynda Hamaoui-laguel, Csaba Zsolt Torma, and Filippo Giorgi. 2016.  
735 “Ragweed Pollen Production and Dispersion Modelling within a Regional Climate System , Calibration and  
736 Application over Europe,” 2769–86. doi:10.5194/bg-13-2769-2016.

737 Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen. 2002. “A Long-Term Hydrologically  
738 Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and  
739 Extensions.” *Journal of Climate*, 3237–51. doi:http://dx.doi.org/10.1175/1520-  
740 0442(2002)015<3237:ALTHBD>2.0.CO;2.

741 Medek, Danielle E, Paul J. Beggs, Bircan Erbas, Alison K. Jaggard, Bradley C. Campbell, Don Vicendese, Fay H  
742 Johnston, et al. 2016. “Regional and Seasonal Variation in Airborne Grass Pollen Levels between Cities of  
743 Australia and New Zealand.” *Aerobiologia* 32 (2): 289–302. doi:10.1007/s10453-015-9399-x.

744 Menzel, Annette, and Susanne Jochner. 2016. “Impacts of Climate Change on Aeroallergen Production and  
745 Atmospheric Concentration.” In *Impacts of Climate Change on Allergens and Allergic Diseases*, edited by  
746 Paul J. Beggs, 10–28. Cambridge, UK: Cambridge University Press.

747 Moseholm, L., E. R. Weeke, and B. N. Petersen. 1987. “Forecast of Pollen Concentrations of Poaceae (Grasses) in  
748 the Air by Time Series Analysis.” *Pollen et Spores* 29 (2–3): 305–21.

749 Myking, T., and O. M. Heide. 1995. “Dormancy Release and Chilling Requirement of Buds of Latitudinal Ecotypes  
750 of *Betula Pendula* and *B. Pubescens*.” *Tree Physiology* 15 (11). Oxford University Press: 697–704.  
751 doi:10.1093/treephys/15.11.697.

752 Myriokefalitakis, S, G Fanourgakis, and M Kanakidou. 2017. “The Contribution of Bioaerosols to the Organic  
753 Carbon Budget of the Atmosphere.” In *Perspectives on Atmospheric Sciences*, 845–51. doi:10.1007/978-3-  
754 319-35095-0\_121.

755 Oleson, K, D Lawrence, G Bonan, M Flanner, and Kluzek E. 2010. “Technical Description of Version 4.0 of the  
756 Community Land Model (CLM).” *NCAR Technical Note NCAR/TN-478+STR*, 257.

757 Olsson, Cecilia, and Anna Maria Jönsson. 2014. “Process-Based Models Not Always Better than Empirical Models  
758 for Simulating Budburst of Norway Spruce and Birch in Europe.” *Global Change Biology*, 3492–3507.  
759 doi:10.1111/gcb.12593.

760 Pal, Jeremy S, Eric E Small, and Elfatih A B Eltahir. 2000. “Simulation of Regional-Scale Water and Energy  
761 Budgets: Representation of Subgrid Cloud and Precipitation Processes within RegCM.” *JOURNAL OF*  
762 *GEOPHYSICAL RESEARCH* 105594 (27): 579–29. http://geode.colorado.edu/~small/docs/2000JGR.pdf.

763 Prieto-Baena, José C, Pablo J Hidalgo, Carmen Galán, and Eugenio Domínguez. 2003. “Pollen Production in the  
764 Poaceae Family.” *Grana* 42 (3): 153–59.  
765 http://www.tandfonline.com/doi/pdf/10.1080/00173130310011810?needAccess=true.

766 Reichstein, Markus, Michael Bahn, Miguel D. Mahecha, Jens Kattge, and Dennis D. Baldocchi. 2014. “Linking  
767 Plant and Ecosystem Functional Biogeography.” *Proceedings of the National Academy of Sciences* 111 (38):  
768 201216065. doi:10.1073/pnas.1216065111.

769 Scheifinger, Helfried, Jordina Belmonte, Jeroen Buters, Sevcan Celenk, Athanasios Damialis, Chantal Dechamp,

770 Herminia García-Mozo, et al. 2013. “Monitoring, Modelling and Forecasting of the Pollen Season.” In  
771 *Allergenic Pollen*, edited by Mikhail Sofiev and Karl-Christian Bergmann. New York, London: Springer  
772 Science+Business Media Dordrecht. doi:10.1007/978-94-007-4881-1\_4.

773 Schueler, Silvio, and Katharina Heinke Schlünzen. 2006. “Modeling of Oak Pollen Dispersal on the Landscape  
774 Level with a Mesoscale Atmospheric Model.” *Environmental Modeling and Assessment* 11 (3): 179–94.  
775 doi:10.1007/s10666-006-9044-8.

776 Siljamo, Pilvi, Mikhail Sofiev, Elena Filatova, Łukasz Grewling, Siegfried Jäger, Ekaterina Khoreva, Tapio  
777 Linkosalo, et al. 2013. “A Numerical Model of Birch Pollen Emission and Dispersion in the Atmosphere.  
778 Model Evaluation and Sensitivity Analysis.” *International Journal of Biometeorology* 57 (1): 125–36.  
779 doi:10.1007/s00484-012-0539-5.

780 Smith, M., and J. Emberlin. 2005. “Constructing a 7-Day Ahead Forecast Model for Grass Pollen at North London,  
781 United Kingdom.” *Clinical and Experimental Allergy* 35 (10): 1400–1406. doi:10.1111/j.1365-  
782 2222.2005.02349.x.

783 Smith, Thomas M, Richard W Reynolds, Thomas C Peterson, and Jay Lawrimore. 2008. “Improvements to NOAA’s  
784 Historical Merged Land-Ocean Surface Temperature Analysis (1880-2006).” *Journal of Climate* 21 (10):  
785 2283–96. doi:10.1175/2007JCLI2100.1.

786 Sofiev, M., P. Siljamo, H. Ranta, T. Linkosalo, S. Jaeger, A. Rasmussen, A. Rantio-Lehtimäki, E. Severova, and J.  
787 Kukkonen. 2013. “A Numerical Model of Birch Pollen Emission and Dispersion in the Atmosphere.  
788 Description of the Emission Module.” *International Journal of Biometeorology* 57 (1): 45–58.  
789 doi:10.1007/s00484-012-0532-z.

790 Sofiev, M., P. Siljamo, H. Ranta, and A. Rantio-Lehtimäki. 2006. “Towards Numerical Forecasting of Long-Range  
791 Air Transport of Birch Pollen: Theoretical Considerations and a Feasibility Study.” *International Journal of*  
792 *Biometeorology* 50 (6): 392–402. doi:10.1007/s00484-006-0027-x.

793 Sofiev, Mikhail, Jordina Belmonte, Regula Gehrig, Rebeca Izquierdo, Matt Smith, Aslog Dahl, and Pilvi Siljamo.  
794 2014. “Airborne Pollen Transport.” In *Allergenic Pollen: A Review of the Production, Release, Distribution*  
795 *and Health Impacts*, edited by Mikhail Sofiev and Karl-Christian Bergman, 127–59. New York, London:  
796 Springer Science+Business Media Dordrecht. doi:10.1007/978-94-007-4881-1.

797 Sofiev, Mikhail, and Marje Prank. 2016. “Impacts of Climate Change on Aeroallergen Dispersion, Transport, and  
798 Deposition.” In *Impacts of Climate Change on Allergens and Allergic Diseases*, edited by Paul J. Beggs, 50–  
799 73. Cambridge, UK: Cambridge University Press.

800 Solmon, F., F. Giorgi, and C. Liousse. 2006. “Aerosol Modelling for Regional Climate Studies: Application to  
801 Anthropogenic Particles and Evaluation over a European/African Domain.” *Tellus, Series B: Chemical and*  
802 *Physical Meteorology* 58 (1): 51–72. doi:10.1111/j.1600-0889.2005.00155.x.

803 Thornton, P.E., M.M. Thornton, B.W. Mayer, Nate Wilhelm, Yaxing Wei, Ranjeet Devarakonda, and R.B. Cook.  
804 2014. “Daymet: Daily Surface Weather Data on a 1-Km Grid for North America, Version 2. Data Set.” *Oak*  
805 *Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA.*  
806 doi:http://dx.doi.org/10.3334/ORNLDAAAC/1219.

807 Tormo Molina, Rafael, Adolfo Muñoz Rodríguez, Silva Palacios, and Francisco Gallardo Lopes. 1996. "Pollen  
808 Production in Anemophilous Trees." *Grana* 35: 38–46. doi:10.1080/00173139609430499.

809 Weber, Richard W. 2003. "Meteorologic Variables in Aerobiology." *Immunology and Allergy Clinics of North*  
810 *America* 23 (3): 411–22. doi:10.1016/S0889-8561(03)00062-6.

811 Yue, X., N. Unger, T. F. Keenan, X. Zhang, and C. S. Vogel. 2015. "Probing the Past 30-Year Phenology Trend of  
812 US Deciduous Forests." *Biogeosciences* 12 (15): 4693–4709. doi:10.5194/bg-12-4693-2015.

813 Zhang, R., T. Duhl, M. T. Salam, J. M. House, R. C. Flagan, E. L. Avol, F. D. Gilliland, et al. 2014. "Development  
814 of a Regional-Scale Pollen Emission and Transport Modeling Framework for Investigating the Impact of  
815 Climate Change on Allergic Airway Disease." *Biogeosciences* 11 (6): 1461–78. doi:10.5194/bg-11-1461-  
816 2014.

817 Zhang, Yong, Leonard Bielory, Ting Cai, Zhongyuan Mi, and Panos Georgopoulos. 2015. "Predicting Onset and  
818 Duration of Airborne Allergenic Pollen Season in the United States." *Atmospheric Environment* 103. Elsevier  
819 Ltd: 297–306. doi:10.1016/j.atmosenv.2014.12.019.

820 Zhang, Yong, Leonard Bielory, Zhongyuan Mi, Ting Cai, Alan Robock, and Panos Georgopoulos. 2015.  
821 "Allergenic Pollen Season Variations in the Past Two Decades under Changing Climate in the United States."  
822 *Global Change Biology* 21 (4): 1581–89. doi:10.1111/gcb.12755.

823 Ziello, Chiara, Tim H Sparks, Nicole Estrella, Jordina Belmonte, Karl C Bergmann, Edith Bucher, Maria Antonia  
824 Brighetti, et al. 2012. "Changes to Airborne Pollen Counts across Europe." *PLoS ONE* 7 (4).  
825 doi:10.1371/journal.pone.0034076.

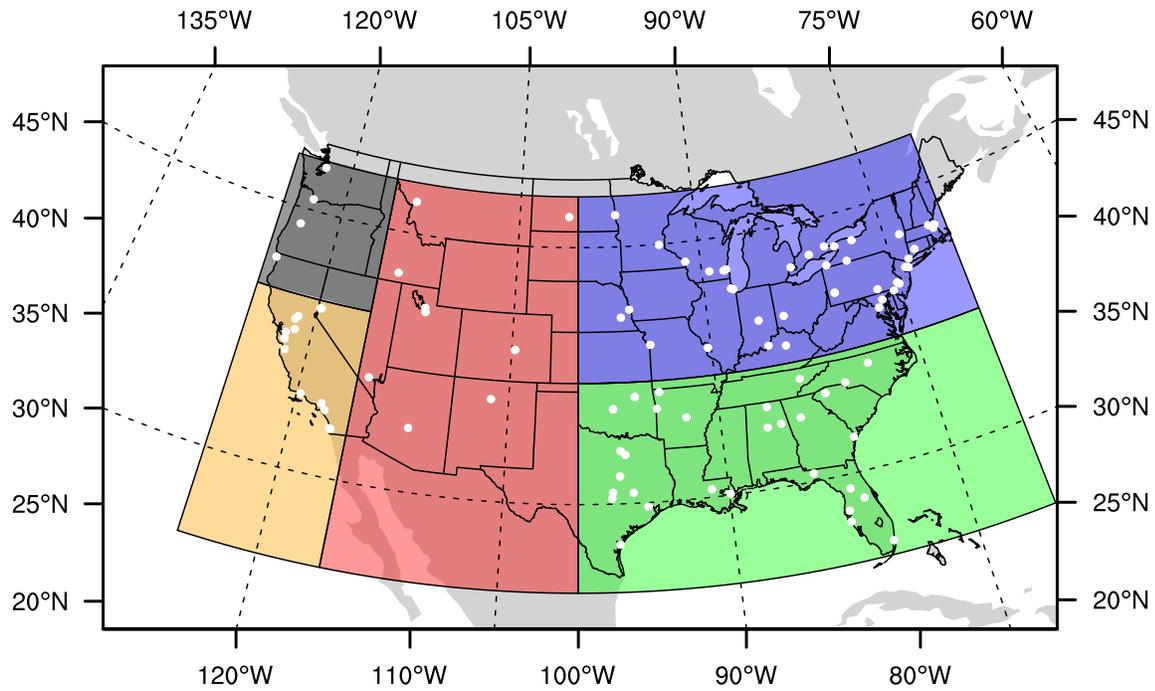
826 Zink, K., a. Pauling, M. W. Rotach, H. Vogel, P. Kaufmann, and B. Clot. 2013. "EMPOL 1.0: A New  
827 Parameterization of Pollen Emission in Numerical Weather Prediction Models." *Geoscientific Model*  
828 *Development* 6 (6): 1961–75. doi:10.5194/gmd-6-1961-2013.

829 Zink, Katrin, Pirmin Kaufmann, Blaise Petitpierre, Olivier Broennimann, Antoine Guisan, Eros Gentilini, Mathias  
830 W Rotach, Mathias W Rotach MathiasRotach, and uibkacat Katrin Zink. 2017. "Numerical Ragweed Pollen  
831 Forecasts Using Different Source Maps: A Comparison for France." *Int J Biometeorol* 61: 23–33.  
832 doi:10.1007/s00484-016-1188-x.

833 Ziska, Lewis H. 2016. "Impacts of Climate Change on Allergen Seasonality." In *Impacts of Climate Change on*  
834 *Allergens and Allergic Diseases*, edited by Paul J. Beggs, 93–112. Cambridge, UK: Cambridge University  
835 Press.

836 Ziska, Lewis, Kim Knowlton, Christine Rogers, Dan Dalan, Nicole Tierney, Mary Ann Elder, Warren Filley, et al.  
837 2011. "Recent Warming by Latitude Associated with Increased Length of Ragweed Pollen Season in Central  
838 North America." *Proceedings of the National Academy of Sciences of the United States of America* 108 (10):  
839 4248–51. doi:10.1073/pnas.1014107108.

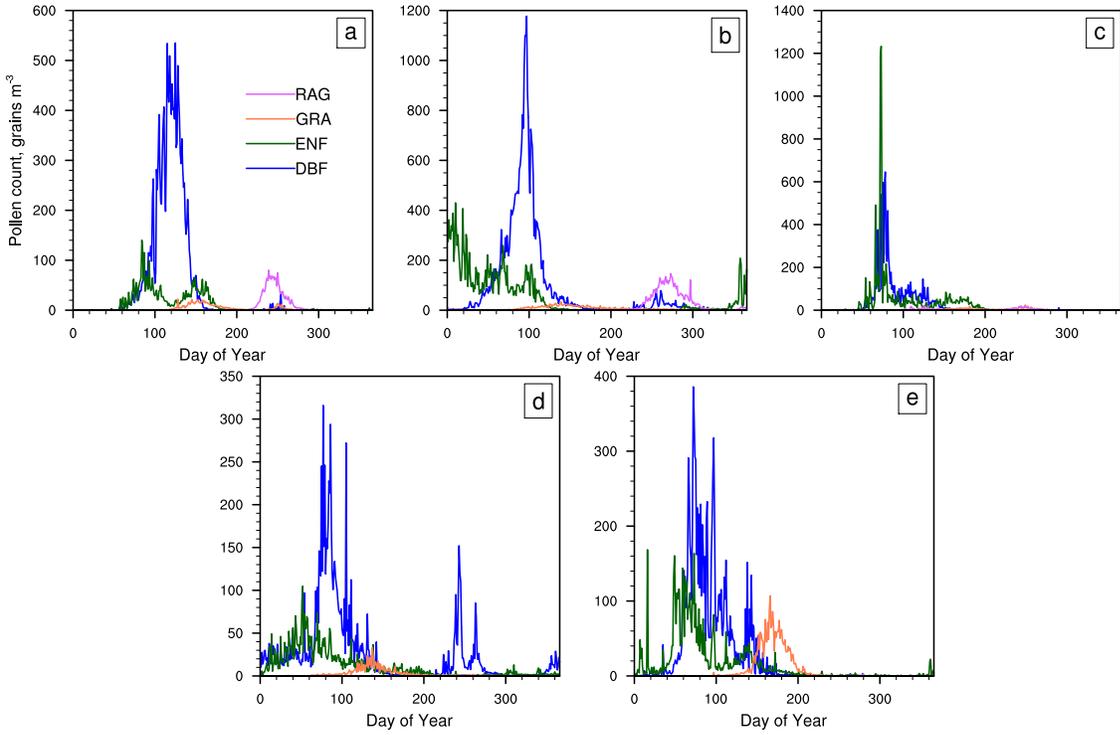
840  
841



842

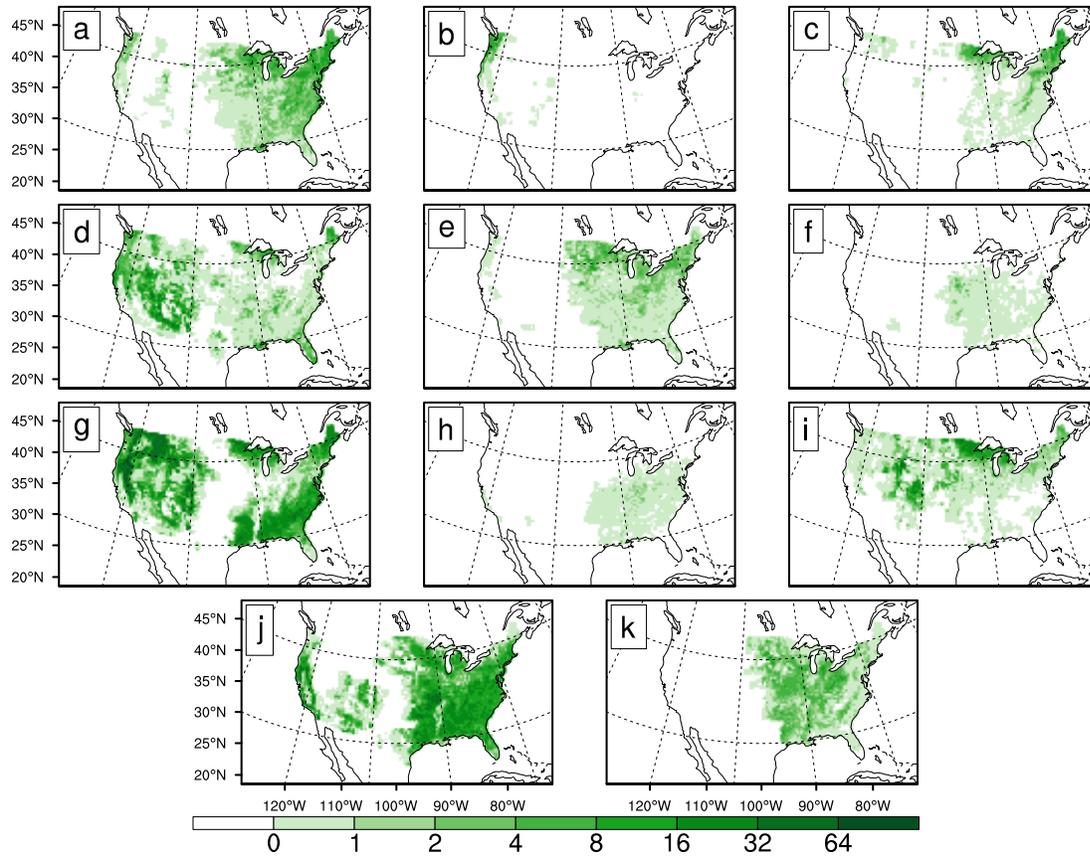
843 **Figure 1.** Locations of AAAAI station locations and geographic regions used in this study: Northeast (NE; 38°-  
 844 48°N, 70°-100°W) in blue, Southeast (SE; 25°-38°N, 70°-100°W) in green, Mountain (MT; 25°-48°N, 100°-116°W)  
 845 in red, California (CA; 25°-40°N, 116°-125°W) in orange, and Pacific Northwest (PNW; 40°-48°N, 116°-125°W) in  
 846 dark grey.

847

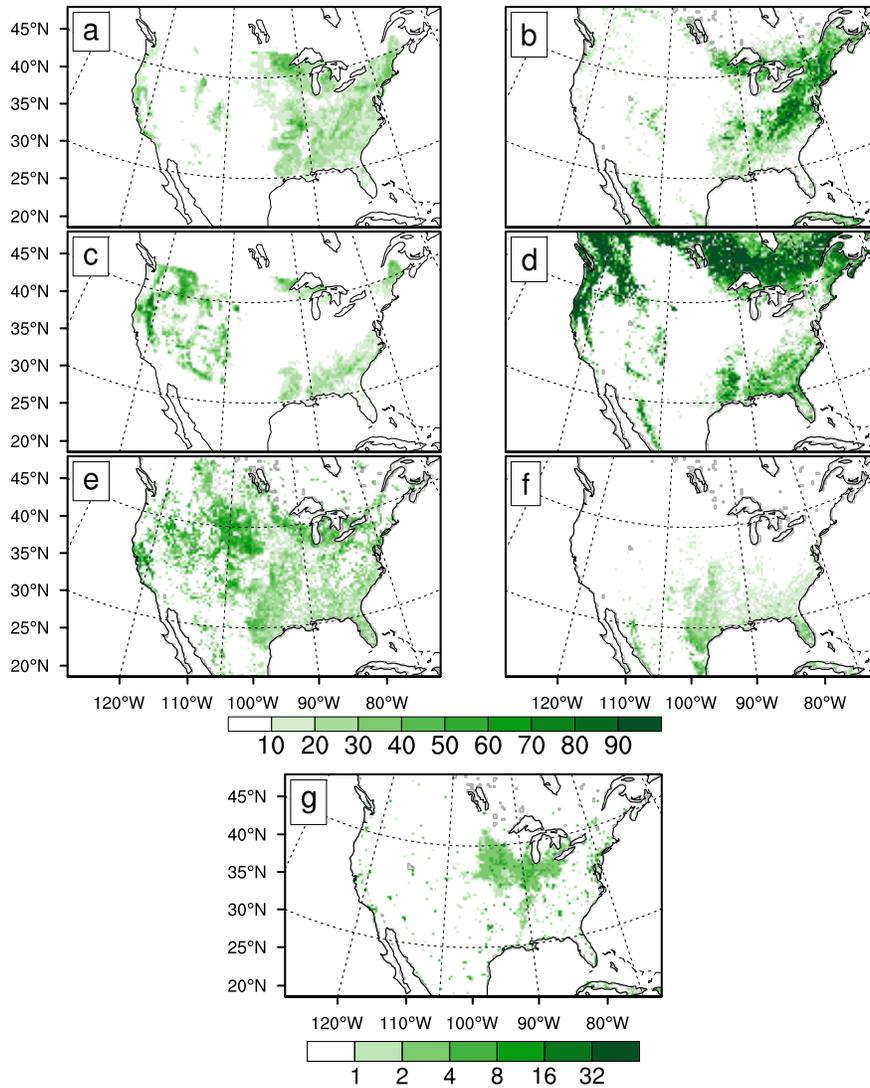


849 **Figure 2.** Daily observed average time series of daily pollen count data (2003-2010) for the four representative  
 850 plant functional types (DBF, ENF, grasses, ragweed) averaged over the five regions in Figure 1: (a) Northeast, (b)  
 851 Southeast, (c) Mountain, (d) California, and (e) Pacific Northwest.

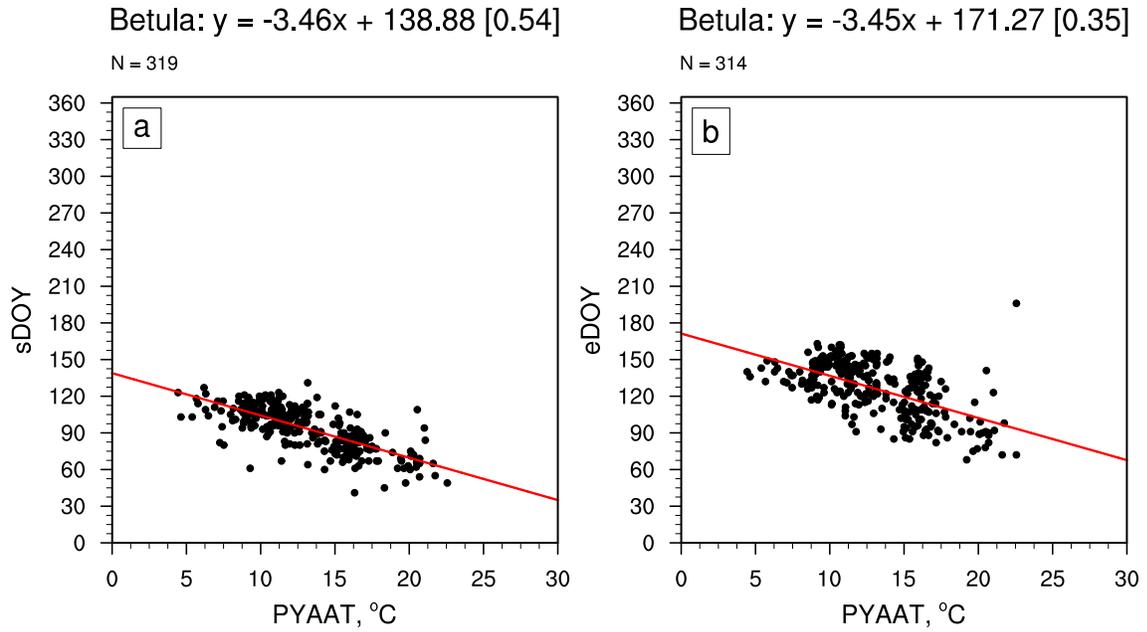
852  
 853  
 854



855  
 856 **Figure 3.** Land cover fraction (% coverage) for 11 tree taxa from the Biogenic Emissions Landuse Database  
 857 (BELD3) regridded to a 25km resolution grid, including: a) *Acer* (maple), b) *Alnus* (alder), c) *Betula* (birch), d)  
 858 Cupressaceae (cedar/juniper), e) *Fraxinus* (ash), f) *Morus* (mulberry), g) Pinaceae (pine), h) *Platanus* (sycamore), i)  
 859 *Populus* (poplar/aspen), j) *Quercus* (oak), k) *Ulmus* (elm).  
 860



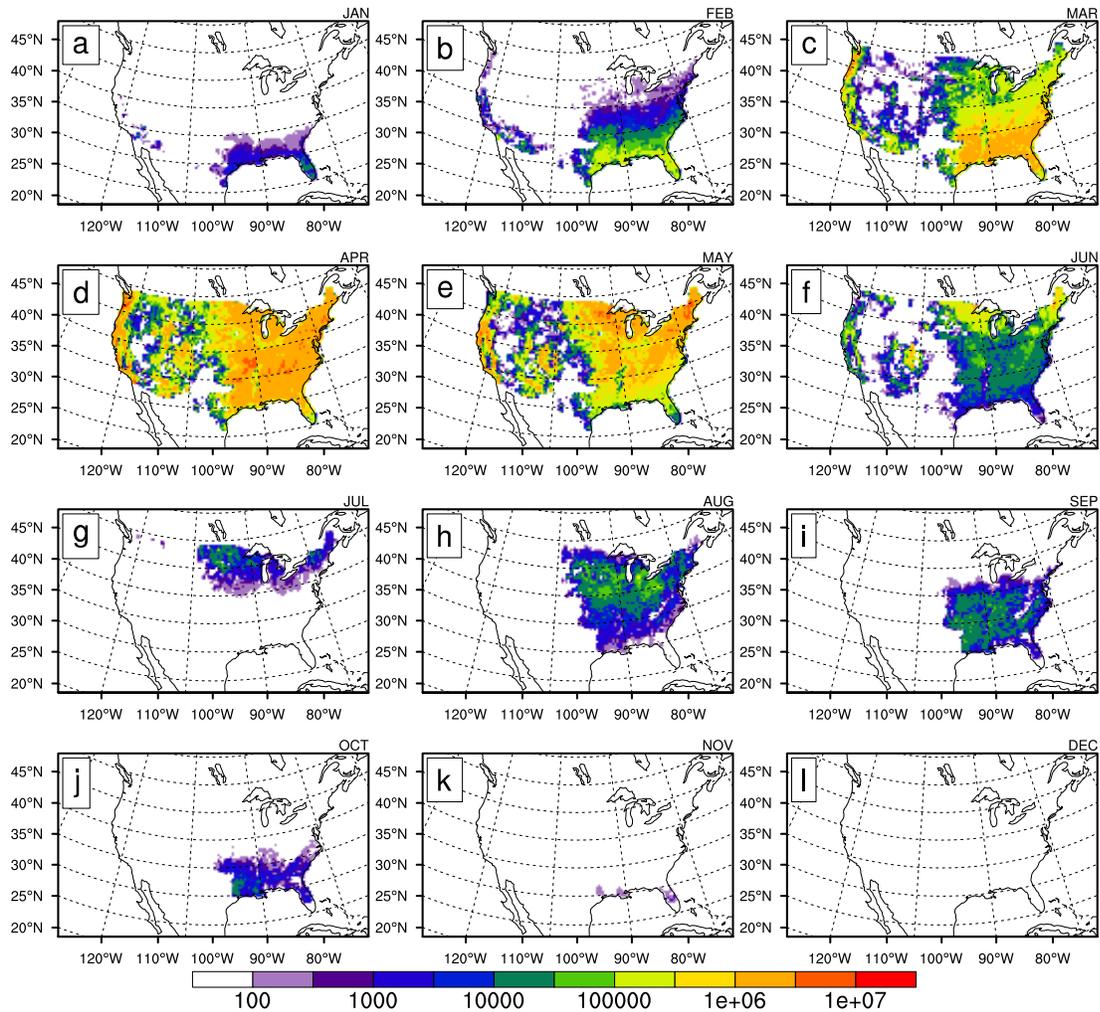
861  
 862 **Figure 4.** BELD3 (a, c) and CLM4 (b, d, e, f) land cover for the four PFT categories that produce pollen emissions,  
 863 including (1) deciduous broadleaf forest for (a) BELD3 and (b) CLM4, (2) evergreen needleleaf forest for (c) BELD3  
 864 and (d) CLM4, (3) grasses, including (e) C3 grasses and (f) C4 grasses, and (g) ragweed, represented by crop and  
 865 urban CLM4 categories.  
 866



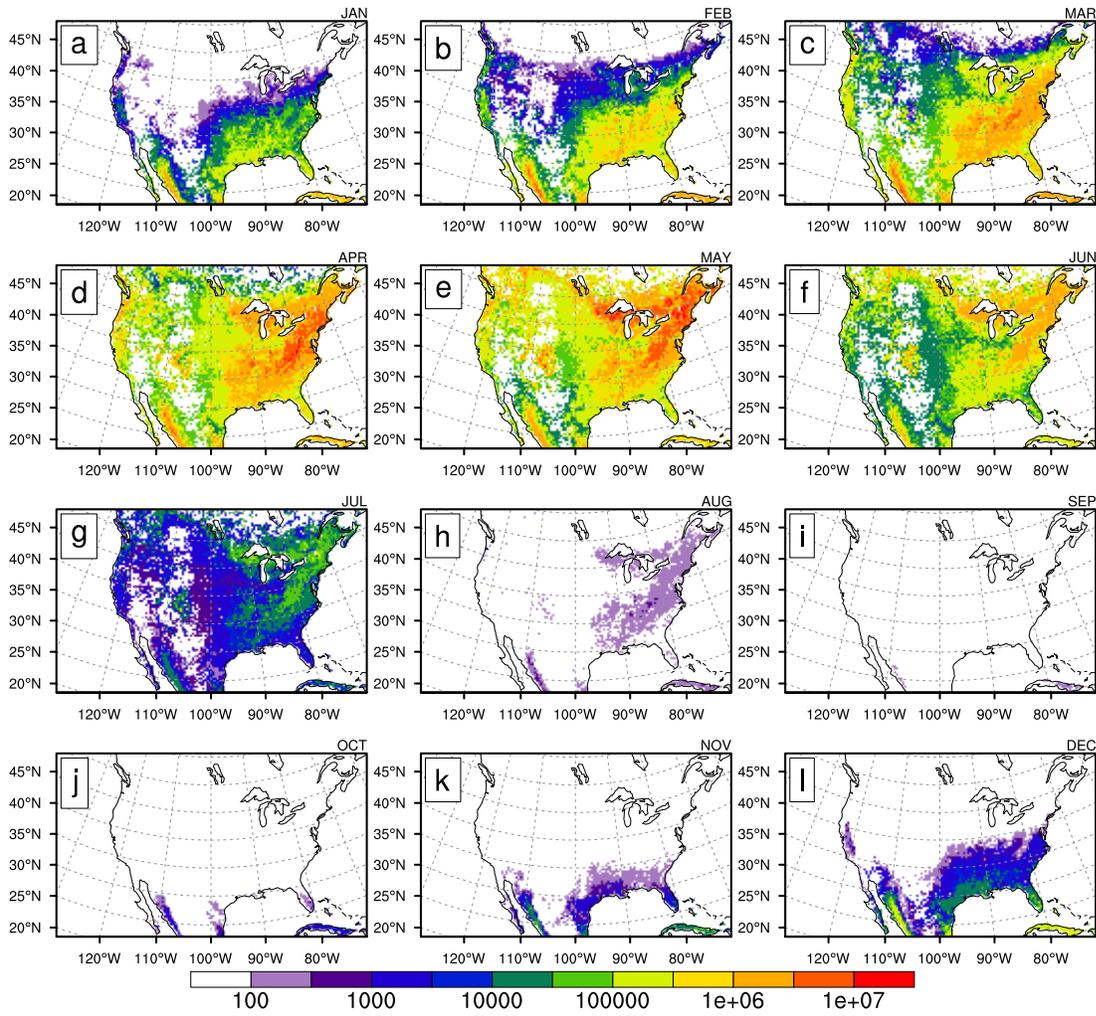
867

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

871



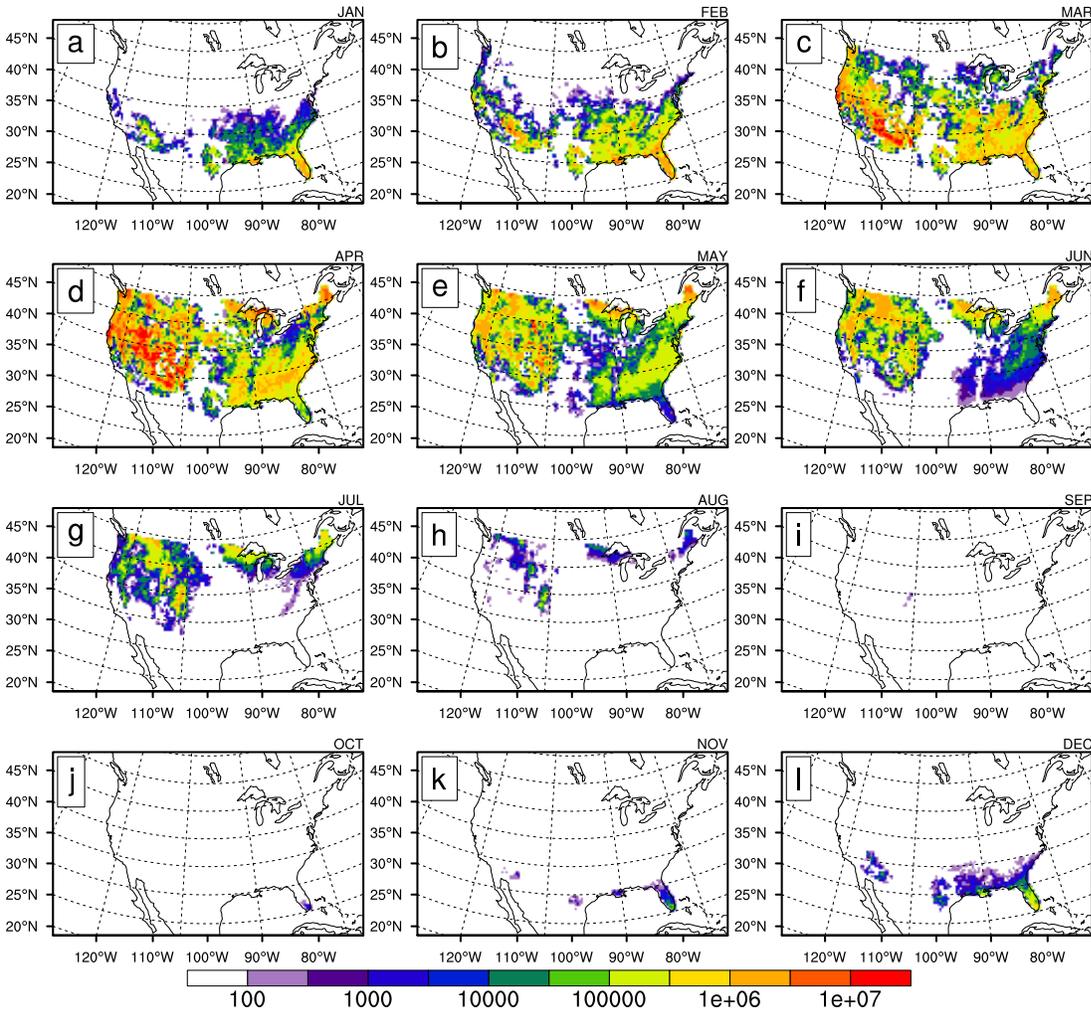
872  
 873 **Figure 6.** Monthly average emissions potential ( $E$ ; Equation 1) for BELD model DBF (2003-2010), in grains  $m^{-2}$   
 874  $day^{-1}$ . a) January, b) February, c) March, d) April, e) May, f) June, g) July, h) August, i) September, j) October, k)  
 875 November, l) December.  
 876



877

878 **Figure 7.** Same as Figure 6, but for PFT model DBF.

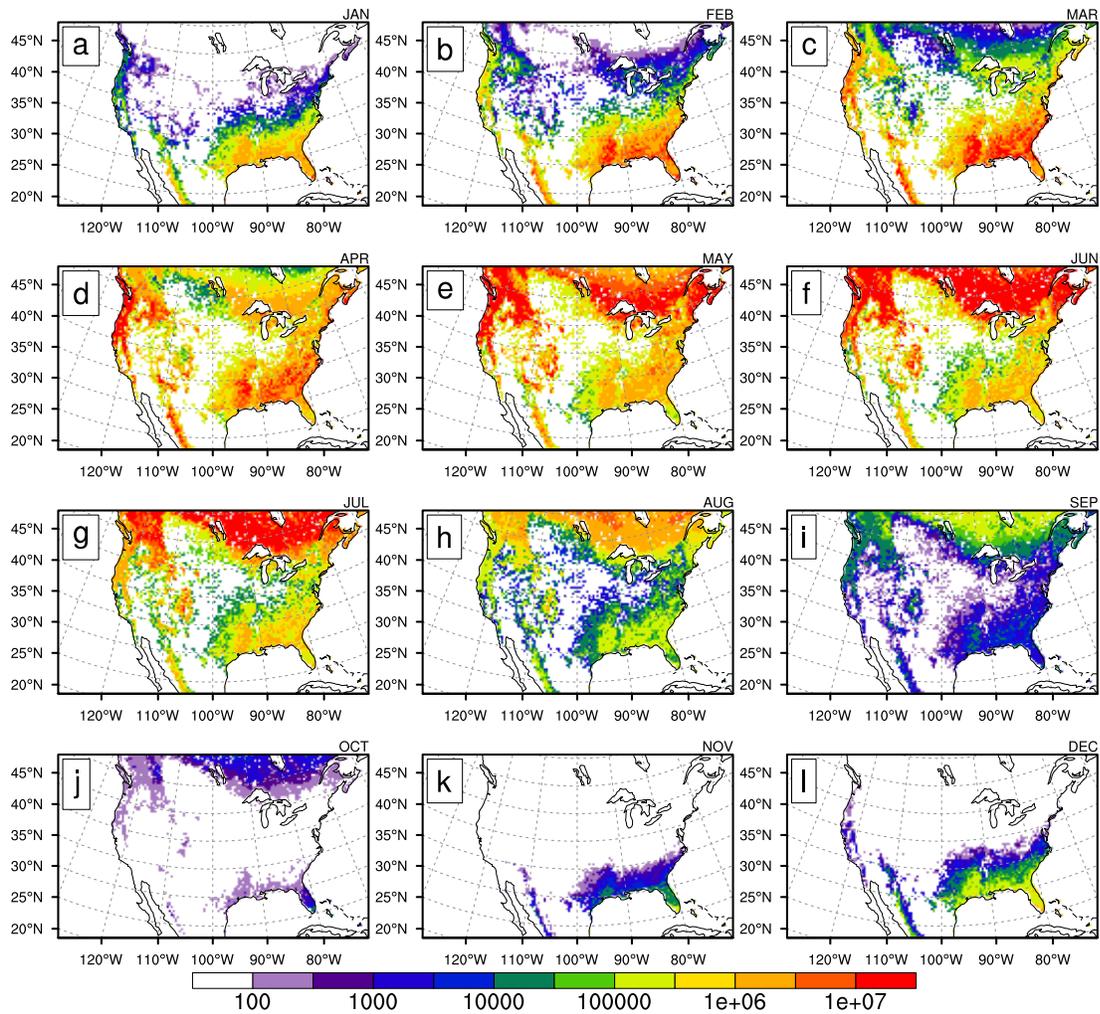
879



880

881 **Figure 8.** Same as Figure 6, but for BELD model ENF.

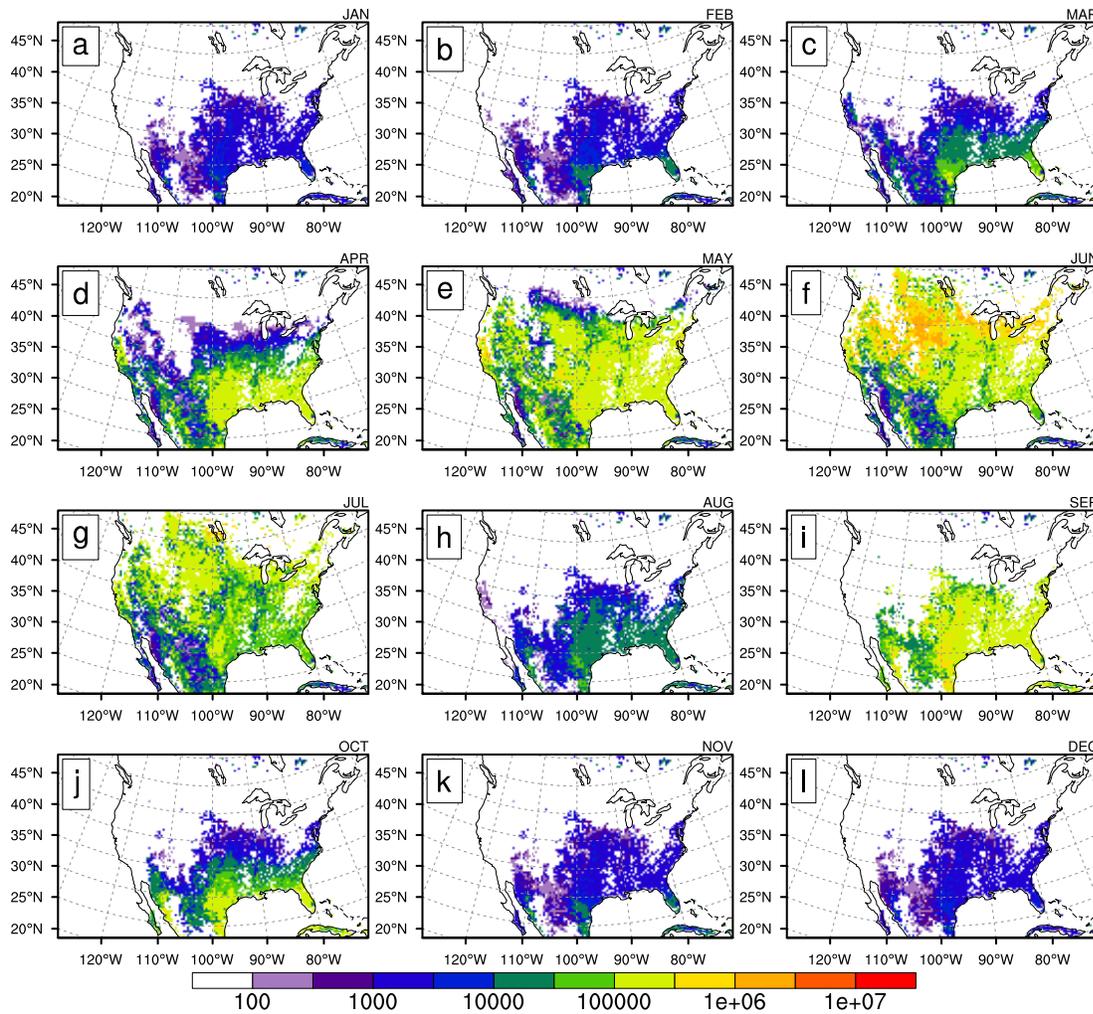
882



883

884 **Figure 9.** Same as Figure 6, but for PFT model ENF.

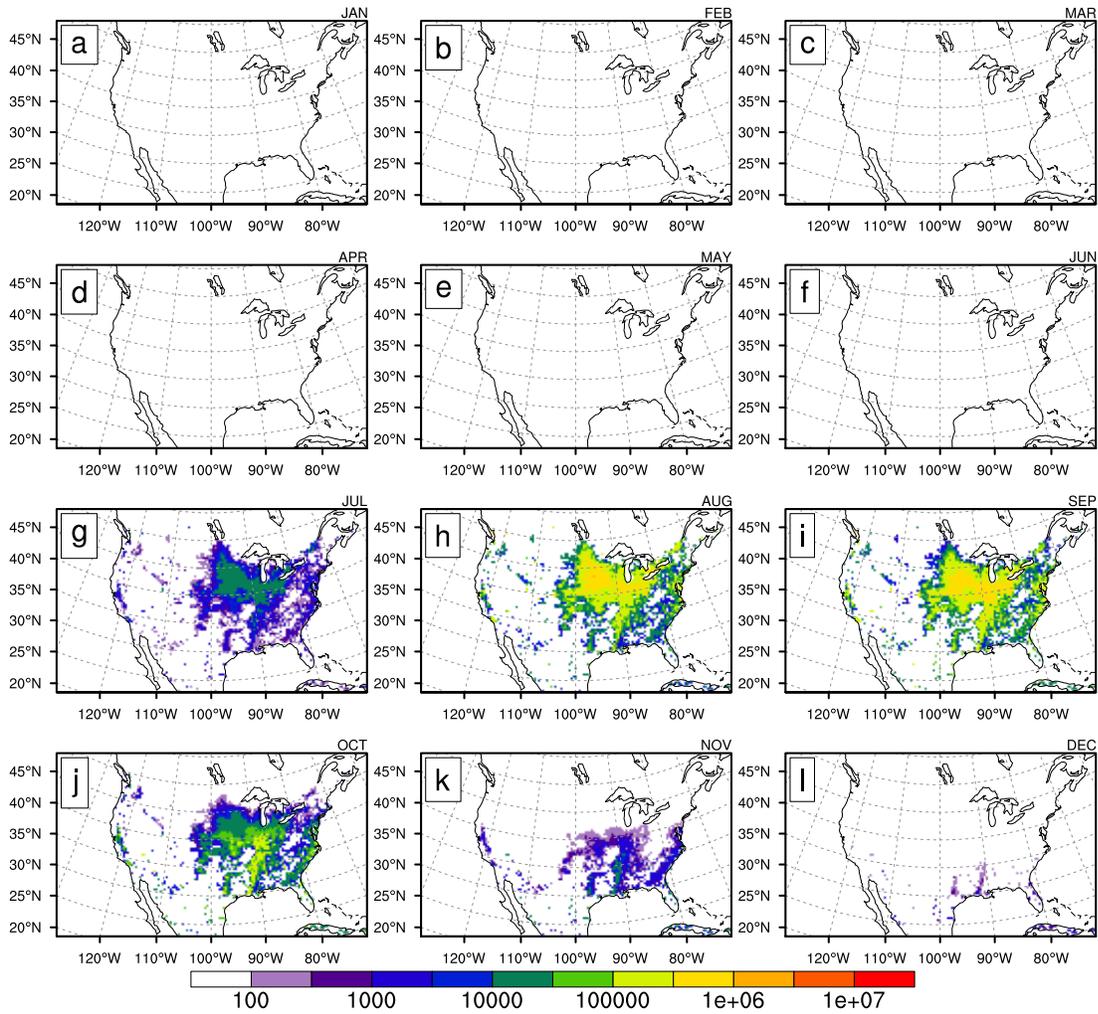
885



886

887 **Figure 10.** Same as Figure 6, but for  $C_3 + C_4$  grass.

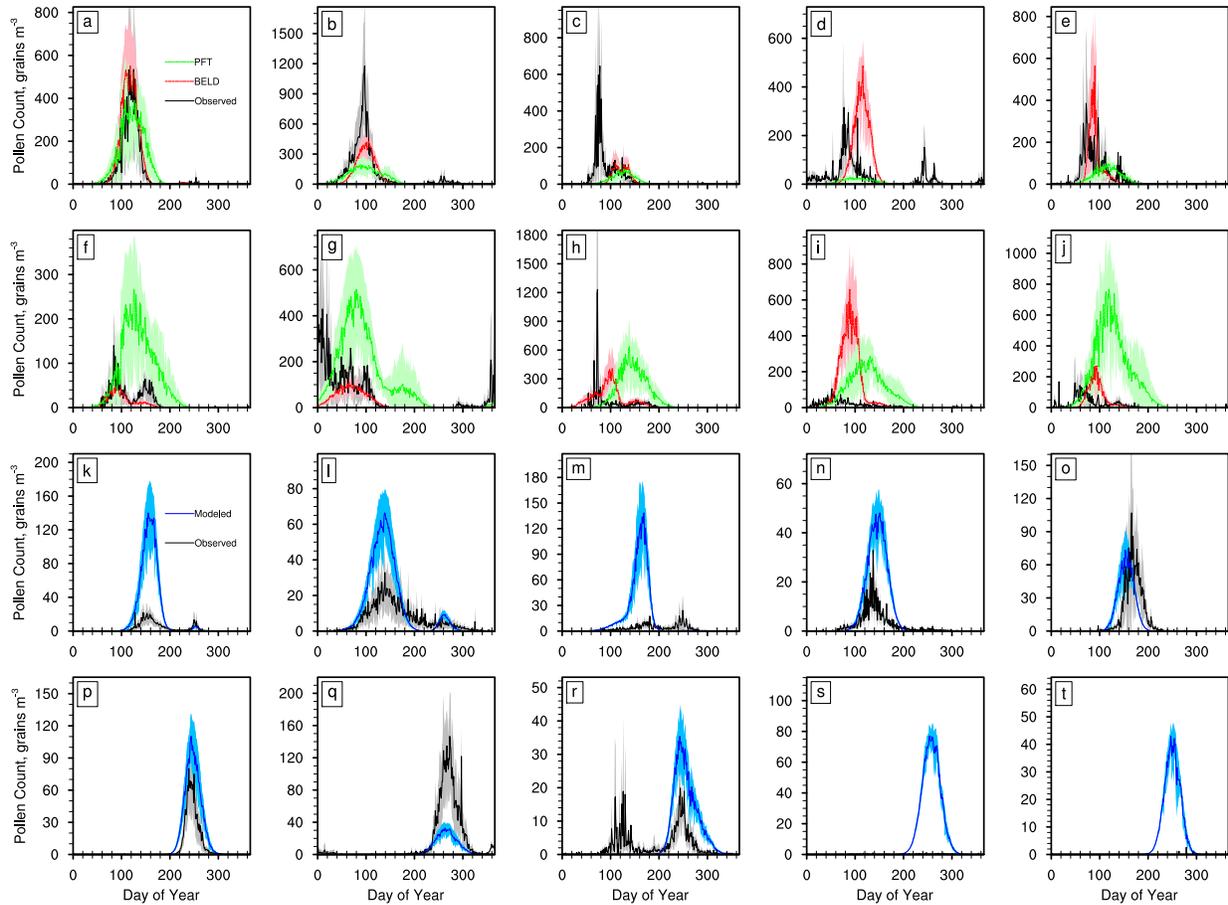
888



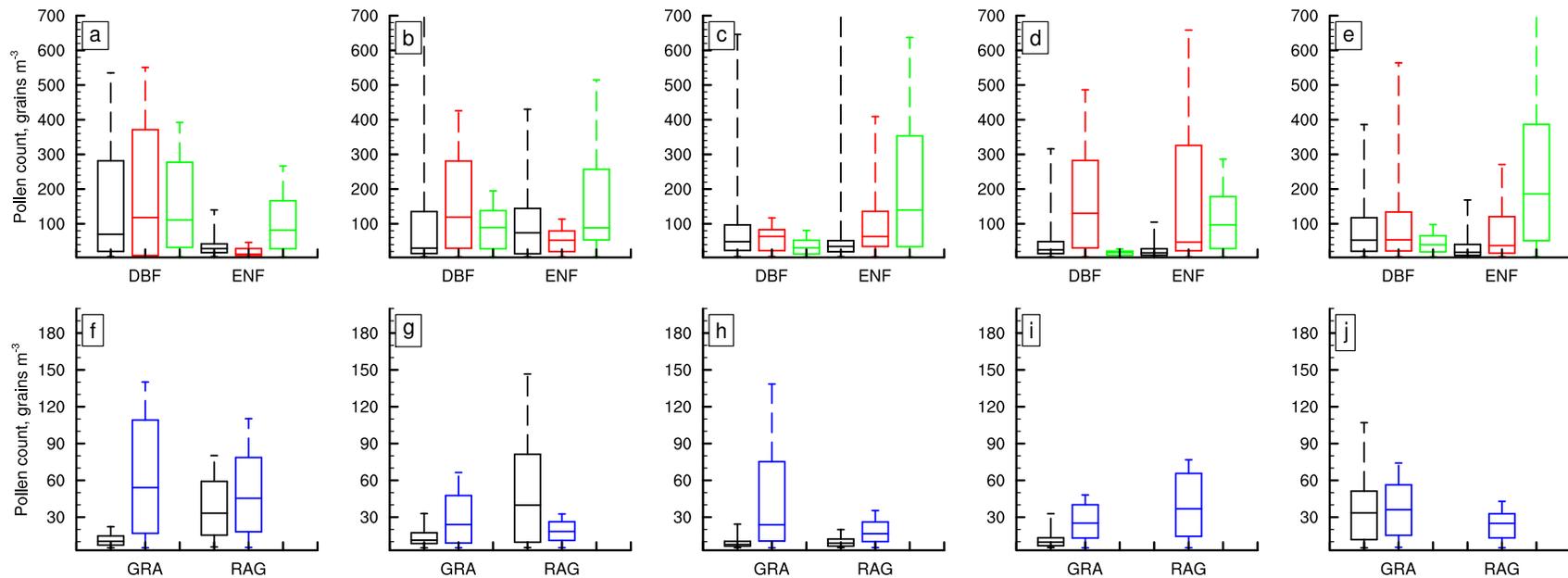
889

890 **Figure 11.** Same as Figure 6, but for ragweed.

891



892  
 893 **Figure 12.** Average (2003-2010) time series of daily pollen counts comparing model and observations for four  
 894 PFTs (a-e, deciduous broadleaf, DBF; f-j, evergreen needleleaf, ENF; k-o, grasses, GRA; p-t, ragweed, RAG) across  
 895 5 U.S. subregions (columns from left to right: Northeast, NE; Southeast, SE; Mountain, MT; California; Pacific  
 896 Northwest, PNW). Shading for the observations and model represents the mean absolute deviation from the  
 897 average for each day of the time series. Note: scale of y-axes varies by region and PFT.  
 898



899

900 **Figure 13.** Box-and-whisker plots showing the statistical spread of the pollen count magnitudes from the regional averages presented in Figure 12. Columns  
 901 from left to right: Northeast, NE (a,f); Southeast, SE (b, g) ; Mountain, MT (c,h); California (d,i); Pacific Northwest, PNW (e,j). DBF and ENF PFTs are shown  
 902 in the top row (a-e) and grass and ragweed PFTs are shown in the bottom row (f-j). Box and whiskers from bottom to top represent the minimum, lower quartile,  
 903 median, upper quartile, and maximum. Maxima that are not visible in panels b, c and e are 1,177 grains m<sup>-3</sup>, 1,233 grains m<sup>-3</sup>, and 766 gains m<sup>-3</sup> respectively. All  
 904 y-axes are the same scale for each row.

Taxon or PFT	P 10 <sup>7</sup> grains m <sup>-2</sup> year <sup>-1</sup>	Reference for P	sDOY (slope/R <sup>2</sup> ) days °C <sup>-1</sup>	eDOY906 (slope/R <sup>2</sup> ) days °C <sup>-1</sup>
Deciduous Broadleaf Forest (DBF)				
<i>Acer</i>	89.1	Tormo Molina et al. 1996	-1.78/0.15	-1.56/0.06
<i>Alnus</i>	210	Helbig et al. 2004	-8.82/0.46	-4.88/0.26
<i>Betula</i>	140	Jato et al. 2007	-3.46/0.54	-3.45/0.35
<i>Fraxinus</i>	45.1	Tormo Molina et al. 1996	-4.69/0.50	-2.92/0.32
<i>Morus</i>	10	N/A	-4.00/0.53	-2.97/0.29
<i>Platanus</i>	121	Tormo Molina et al. 1996	-4.47/0.40	-2.65/0.20
<i>Populus</i>	24.2	Tormo Molina et al. 1996	-2.23/0.24	-0.31/<0.01
<i>Quercus</i>	78	Tormo Molina et al. 1996	-4.09/0.53	-2.03/0.19
<i>Ulmus (early,late)</i>	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		-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		-5.95/0.40	-4.96/0.33
Grasses (GRA)				
Poaceae (C <sub>3</sub> ,C <sub>4</sub> )	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)				
<i>Ambrosia</i>	119 <sup>a</sup>	Fumanal et al. 2007	1.08/0.08	3.42/0.37

907 <sup>a</sup>*Ambrosia* production factor in 10<sup>7</sup> grains plant<sup>-1</sup>.

908 **Table 1:** Production factors (P) and phenological regression coefficients for the start day of year (sDOY) and end  
909 day of year (eDOY) as a function of temperature for the 13 individual pollen-producing taxa. Individual taxa and

910 families are organized into the four PFTs, with the two aggregated tree PFTs denoted as DBF and ENF. Regression  
911 slope (days/°C) and coefficient of determination are provided for both sDOY and eDOY (slope/R<sup>2</sup>).

Land cover class	NE	SE	MT	CA	PNW
<i>Acer</i>	6.79E+04	2.88E+04	1.89E+03	1.97E+02	3.09E+03
<i>Alnus</i>	3.37E+00	1.23E-01	6.49E+01	1.71E+02	9.56E+03
<i>Betula</i>	2.99E+04	2.68E+03	2.78E+02	2.64E+00	4.82E+02
<i>Fraxinus</i>	3.96E+04	1.10E+04	3.14E+03	3.94E+01	2.76E+02
<i>Morus</i>	3.99E+03	2.25E+03	3.89E+01	0.00E+00	0.00E+00
<i>Platanus</i>	3.18E+03	3.38E+03	1.33E+01	1.44E+02	0.00E+00
<i>Populus</i>	5.48E+04	1.23E+03	4.37E+04	1.96E+02	1.55E+03
<i>Quercus</i>	1.30E+05	2.25E+05	2.51E+04	2.82E+04	1.40E+04
<i>Ulmus</i>	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
CLM 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
CLM ENF	1.44E+06	4.26E+05	4.66E+05	4.57E+04	5.34E+05

912 **Table 2:** Total spatial coverage (km<sup>2</sup>) of tree taxa and PFTs from BELD and CLM land cover datasets in the 5 U.S.  
913 subregions (Northeast, NE; Southeast, SE; Mountain, MT; California; Pacific Northwest, PNW). All individual tree  
914 taxa are from the BELD database. BELD DBF and ENF land cover are the sums of the land cover of the taxa  
915 belonging to each PFT.

916