Aerobiologia
https://doi.org/10.1007/s10453-022-09774-3

BRIEF REPORT

Observations from the USA National Phenology Network can be leveraged to model airborne pollen

Daniel S. W. Katz · Elizabeth Vogt · Arie Manangan · Claudia L. Brown · Dan Dalan · Kai Zhu · Yiluan Song · Theresa M. Crimmins

Received: 21 July 2022 / Accepted: 28 November 2022

Abstract The USA National Phenology Network (USA-NPN) hosts the largest volunteer-contributed collection of plant phenology observations in the USA. The potential contributions of these spatially and temporally explicit observations of flowers and pollen cones to the field of aerobiology remain largely unexplored. Here, we introduce this freely available dataset and demonstrate its prospective applications for modeling airborne pollen in a case study. Specifically, we compare the timing of 4265 observations of flowering for oak (*Quercus*) trees in the eastern USA to winter–spring temperatures. We then use this relationship to predict the day of peak flowering at 15 pollen monitoring stations in 15 years and compare the predicted day of peak flowering to the peak day of measured pollen ($n=111$ station-years). There was a strong association between winter–spring temperature and the presence of open flowers ($r^2=0.66$, $p<0.0001$) and the predicted peak flowering was strongly correlated with peak airborne pollen concentrations ($r^2=0.81$, $p<0.0001$). These results demonstrate the potential for the USA-NPN's phenological observations to underpin source-based models of airborne pollen. We also highlight opportunities for leveraging and enhancing this near real-time dataset for aerobiological applications.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10453-022-09774-3.

D. S. W. Katz (✉)
School of Integrative Plant Science, Cornell University, Ithaca, NY, USA
e-mail: dankatz@cornell.edu

E. Vogt
Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, USA

A. Manangan · C. L. Brown
Climate and Health Program, Division of Environmental Health Science and Practice, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA, USA

D. Dalan
Allergy & Immunology, MercyOne Waterloo Allergy Care, Waterloo, IA, USA

K. Zhu · Y. Song
Department of Environmental Studies, University of California, Santa Cruz, CA, USA

K. Zhu
Institute for Global Change Biology and School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA

T. M. Crimmins
USA National Phenology Network, School of Natural Resources and Environment, University of Arizona, Tucson, AZ, USA

Published online: 23 December 2022

Keywords Flowers · Pollen cones · Citizen science · Phenological models · Open data

1 Introduction

The timing of pollen release determines airborne pollen concentrations. In situ observations of reproductive activity of wind-pollinated species over space and time can be incorporated into models of airborne pollen concentrations. The USA National Phenology Network (USA-NPN) maintains the most extensive collections of phenology observations in the USA, including observations of flowering and pollen release for hundreds of species. Here, we explore the potential pollen modeling applications of USA-NPN's data.

The USA-NPN was established in 2007 to collect, store, and share plant and animal phenology data and information to support scientific discovery, decision-making, an appreciation for phenology, and equitable engagement within the network. The USA-NPN hosts *Nature's Notebook*, a rigorous platform for monitoring plant and animal phenology (Rosemartin et al., 2014). Since the launch of *Nature's Notebook* in 2009, over 30 million records of phenology have been contributed, both by professional and citizen scientists as well as many partner groups. These data are increasingly used in science, management, and education applications (Crimmins et al., 2022).

The phenology observations curated by the USA-NPN and by other phenology networks such as the Pan European Phenological database (Templ et al., 2018) hold untapped potential to contribute to source-based ('numerical') models of airborne pollen. Such models could supplement pollen monitoring activities conducted by the National Allergy Bureau (NAB) network, whose approximately 80 stations are the foundation of pollen monitoring in the USA. Specifically, pollen models have the potential to provide complete geographic coverage at fine spatial resolutions and are well suited for near-term forecasting. The resulting improvements in pollen forecasts and alerts could have public health applications such as improved allergy medication management or allergen avoidance through behavioral changes. However, the extensive data on source plants that are required to create source-based models of airborne pollen have historically been collected at local scales (e.g., Katz et al., 2019) and regional to continental phenological datasets are seldom leveraged (but see examples like Siljamo et al., 2013). Here, we explore whether observations of flowering status contributed to *Nature's Notebook* could address this gap and be used to model the timing of the pollen season.

2 Methods

Participants in *Nature's Notebook* periodically report leaf, flower, and pollen cone status for individual plants identified to the species level (Denny et al., 2014). A screenshot of the app interface is provided in Fig. 1a. Thousands of *Nature's Notebook* participants across the country report observations at an average frequency of once a week. Participants observe whether reproductive organs are visible; if so, they record the proportion open and can note whether pollen is released. The data are publicly available through an online query tool (www.usanpn.org/data/observational) and an API (Rosemartin et al., 2018). Using the USA-NPN's R package "*rnpn*," we downloaded all observations of flowers or pollen cones from 2009 to 2021 for common allergenic or ecologically important anemophilous species that were in the USA-NPN database. We removed records where multiple observers reported conflicting flowering for the same individual plant on the same day and omitted sites where more than 5% of reports were conflicting; further discussion of data cleaning is provided in the "*rnpn*" package documentation and "data cleaning" vignette.

We specifically focus our case study on *Quercus* because this genus is abundant in North American forests and cities, an important source of allergenic pollen with large public health impacts (Anenberg et al., 2017), and well represented in *Nature's Notebook*. Our objective with this simple case study is to demonstrate the potential of USA-NPN data to support pollen modeling. *Quercus* is a suitable taxon because it tends to exhibit unimodal flowering peaks and relatively minor differences in the timing of flowering between species. We restricted our analysis to the eastern USA (longitudes east of − 90°) as unlike eastern *Quercus,* western *Quercus* is strongly influenced by precipitation (Gerst et al., 2017).

We downloaded average monthly temperature data from PRISM (PRISM Climate Group, 2022) and

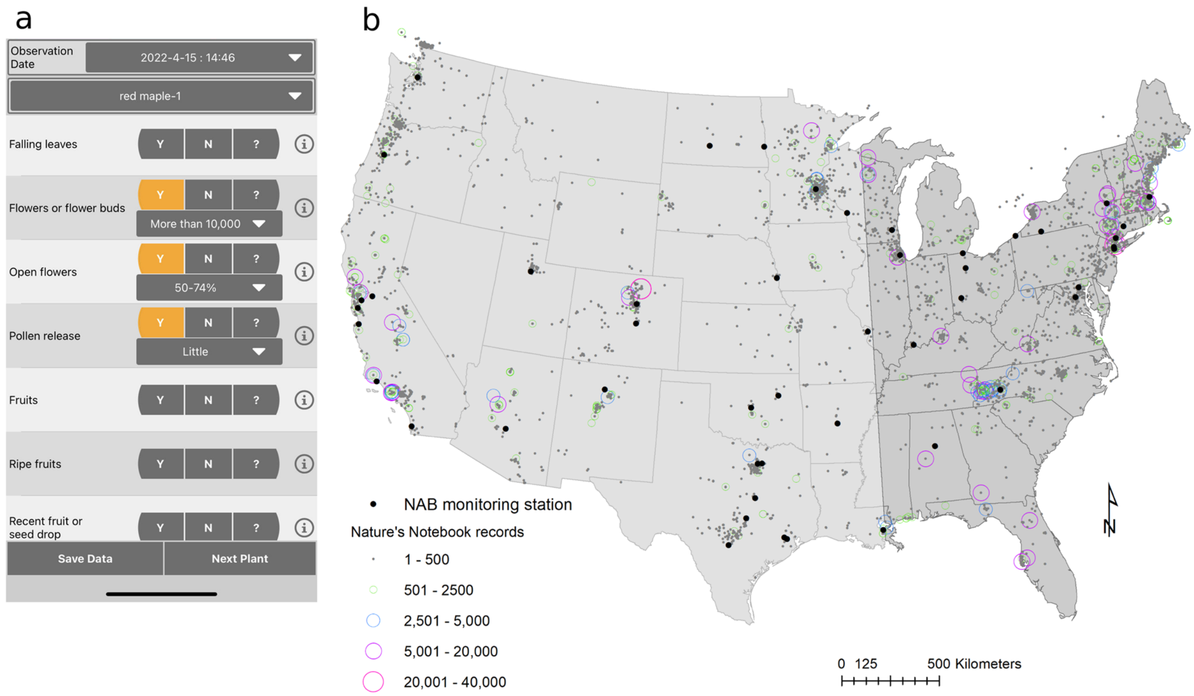

Fig. 1 **a** Observation interface used by *Nature's Notebook* observers to report flowering and pollen release from individual plants **b** flower and pollen status records available from the USA-NPN (2009–2021) in the continental USA and National Allergy Bureau (NAB) certified pollen monitoring stations that provided data for this study. The case study focuses on the eastern US (dark gray)

extracted winter–spring temperature data (defined here as January–April) for the phenology observation coordinates; winter–spring temperatures are strong predictors of *Quercus* flowering phenology (e.g., Katz et al., 2019). Observations that were not biologically plausible (i.e., the 3% of flowering observations reported July first or later) were discarded. We then compared the day of year when open flowers were observed to winter–spring temperature with linear regression across all 18 *Quercus* species for which flowering phenology data were available. This regression predicted peak flowering and we combined it with PRISM data to extrapolate the day of peak flowering across the USA in each year. The predicted day of peak flowering and the accompanying prediction interval was then extracted for each NAB station in each year.

Airborne pollen concentrations were generously provided by 15 NAB stations (American Academy of Allergy Asthma & Immunology, 2022), which used impact samplers (Burkard or Rotorod) on rooftops to measure daily airborne pollen concentrations, generally to the genus level, several times per week. We calculated the peak day of the *Quercus* pollen season for these pollen monitoring stations, resulting in 139 station-years. To calculate the peak day of pollen, we first linearly interpolated missing pollen data (e.g., from stations that did not collect data on weekends) and then calculated a 14-day moving average of observed *Quercus* pollen. The peak day of the pollen season was defined as the calendar day with the highest 14-day moving average of *Quercus* pollen. Moving averages reduced daily stochastic effects caused by day-to-day changes in variables like wind direction, pollen emissions, pollen washout from rain, and differences in which observers reported phenology data. Station-year combinations with low *Quercus* concentrations (where the moving average never exceeded 100 grains/m^3) were excluded from the analysis, resulting in a total of 111 station-years. The relationship between the predicted day of peak flowering and the day of peak airborne pollen was assessed with linear regression and Q-Q plots were used to assess normality. Analyses were conducted in R 4.1.3 (R Core Team 2018), and all code is provided on the author's GitHub page: https://github.com/dankatz/NPN_NAB2/blob/main/NPN_NAB_oak_case_study.R

3 Results

Nearly 1 M observations of flowers or pollen cones in the USA are available for common plants of aerobiological relevance from the USA-NPN for 2009–2021 (Fig. 1b), with the highest densities of observations in the Northeast and Midwest. The number of observations has increased since the program's establishment: in 2010 6,003 records were contributed across the USA, and by 2021, 132,770 were contributed. The most frequently observed tree taxa include: *Acer* (273,762 records of the presence or absence of reproductive structures), *Quercus* (218,715), Pinaceae (54,018), *Populus* (52,808), *Betula* (51,137), and *Juniperus* (22,181). Data for grasses or wind-pollinated herbaceous species are generally sparse but vary by species. (Sample sizes for each species are provided in SI 1.)

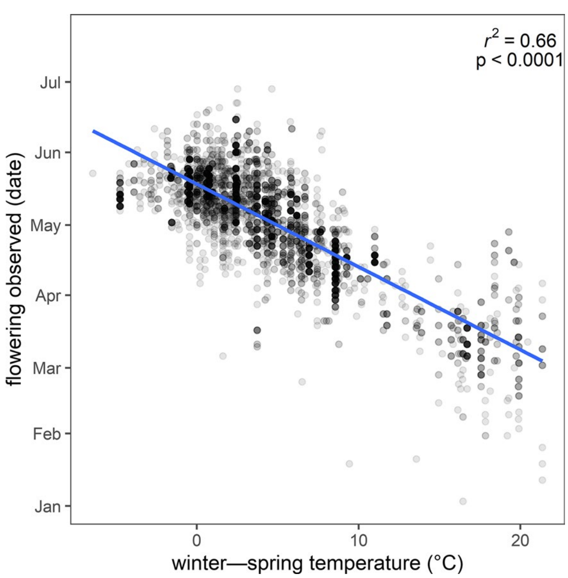

Fig. 2 Association between the day of year when open flowers of *Quercus* species were reported by *Nature's Notebook* observers and winter–spring (Jan–Apr) temperature in the year of observation at the tree's location. Observations are constrained to the eastern USA

In the eastern USA, there were a total of 4265 observations of open *Quercus* flowers, of which most were of *Quercus rubra* L. (1661), *Quercus alba* L. (790), *Quercus montana* Willd. (565), or *Quercus velutina* Lam. (455). *Quercus* flowers in warmer conditions opened substantially earlier than those in colder conditions (Fig. 2). For each degree of temperature increase, flowering time was on average 3.5 days earlier. Geographic patterns of residuals are shown in SI 2. There was substantial variation in the timing of observed flowering even within a particular temperature, and the mean absolute difference between the regression and an observed point was 9.7 days. The modeled timing of peak flowering at the NAB stations based on winter–spring temperature in each year effectively predicted the observed peak day of pollen concentrations (Fig. 3; $r^2 = 0.81$). The general pattern held within sites across years (SI 3) and did not show strong geographic patterns (SI 4). The average absolute difference between predicted peak flowering and peak airborne pollen was 5.8 days.

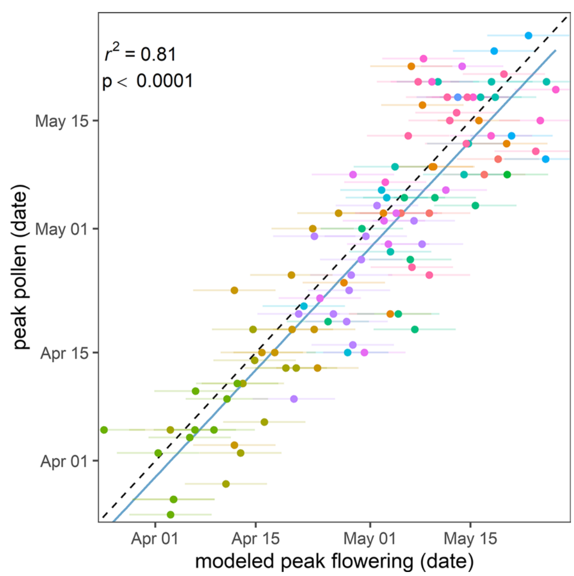

Fig. 3 Association between predicted day of peak flowering and the observed peak day of airborne pollen concentrations. The predicted day of peak flowering is based on the relationship between observations of open *Quercus* flowers contributed by *Nature's Notebook* participants and winter–spring temperature shown in Fig. 2; horizontal error bars show the standard deviation of the prediction interval. Colors represent pollen monitoring stations; a best fit line (blue) and a one:one line (dotted) are included

4 Discussion

4.1 Realizing the USA-NPN's aerobiological potential

Despite the potential utility of the USA-NPN's phenology data for aerobiological applications, we have not previously seen it used for pollen forecasting. One potential barrier is that data are lacking for certain taxa. However, the USA-NPN has demonstrated experience in driving data collection in support of particular research questions and applications (Crimmins et al., 2014). For example, the Pollen Trackers campaign collected 3953 measurements on pollen cone phenology and pollen release for *Juniperus ashei* J. Buchholz in the most recent field season, and these observations are used to create regional pollen models in Texas. New campaigns could collect additional data to support aerobiological investigations for taxa of interest. While volunteer-collected data can be subject to errors in species and phenophase identification as well as infrequent or inconsistent sampling, the approaches used here appear robust to occasional outliers. Moreover, USA-NPN data are available within a day of collection, allowing for continuously updated forecasting (Taylor & White, 2020) and near-term iterative modeling that could improve pollen forecast accuracy.

4.2 Source-based pollen models

Source-based models of airborne pollen have the potential to create pollen forecasts across large geographic regions, including those without pollen monitoring stations. This requires phenological models, which the USA-NPN observations have enabled (e.g., Elmendorf et al., 2019; Gerst et al., 2017), as have other phenological monitoring networks elsewhere (Templ et al., 2018). Basing phenological models on observations of flowering and pollen release instead of on airborne pollen measurements both removes complexity and variability (e.g., from atmospheric dispersion) and allows for predictions that are generalizable beyond the idiosyncratic environments of a particular pollen monitoring station. Moreover, this approach allows models to be created at the species level instead of the genus level, to which airborne pollen is typically identified.

4.3 *Quercus* flowering case study

This case study shows the ease with which phenological observations can be leveraged to create phenological models. The present approach was coarse: We did not optimize the temperature window, include the finest-grained temperature data, apply complex filters to the USA-NPN data, use biologically based phenology modeling approaches, or include the other important processes such as pollen production or dispersion. A comprehensive approach could include modeling phenology for each of the 18 *Quercus* species for which data were available (e.g., SI 5) and then adding their estimates to match pollen data collected at the genus level. Each species could be weighted by local plant composition; however, that remains beyond the scope of this simple example. Even with these caveats, and those associated with airborne pollen measurements (e.g., missing data), this simple model effectively predicted the day of peak *Quercus* pollen within an average of six days. Thus, our case study suggests there is strong potential for USA-NPN data to be used to predict the timing of pollen seasons and ultimately airborne pollen concentrations.

Acknowledgements We would like to thank the American Academy of Allergy, Asthma, & Immunology National Allergy Bureau for providing pollen concentration data and the following individuals in particular for the measurements that made this study possible: D.S. Shulan, G. Robinson, D. Cypcar, S. Fineman, W. Soong, R. Shah, B. Huxtable, P.E. Gallagher, F.H. Lewis, J.A. MacLean, S.E. Kosisky, L. Bielory, G.S. Safadi, and C. Randolph. We also thank the many participants who contribute phenology observations to the *Nature's Notebook* program. The USA-NPN is supported by National Science Foundation grant DBI-2031660. KZ and YS acknowledge National Science Foundation award 2306198.

Authors' contribution DK, EV, and TMC led study conception, design, analysis, and writing, with input and review from all authors.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

References

American academy of allergy asthma & immunology. (2022). National Allergy Bureau. Retrieved 1 January 2022 from https://www.aaaai.org/global/nab-pollen-counts/about-the-nab

Anenberg, S., Weinberger, K. R., Roman, H., Neumann, J., Crimmins, A., Fann, N., et al. (2017). Impacts of oak pollen on allergic asthma in the United States and potential influence of future climate change. *GeoHealth, 1*(3), 80–92. https://doi.org/10.1002/2017GH000055

Crimmins, T. M., Weltzin, J. F., Rosemartin, A. H., Surina, E. M., Marsh, L., & Denny, E. G. (2014). Focused campaign increases activity among participants in nature's notebook, a citizen science project. *Natural Sciences Education, 43*(1), 64–72. https://doi.org/10.4195/nse2013.06.0019

Crimmins, T., Denny, E., Posthumus, E., Rosemartin, A., Croll, R., Montano, M., & Panci, H. (2022). Science and management advancements made possible by the USA national phenology network's nature's notebook platform. *BioScience, 72*(9), 908–920. https://doi.org/10.1093/biosci/biac061

Denny, E. G., Gerst, K. L., Miller-Rushing, A. J., Tierney, G. L., Crimmins, T. M., Enquist, C. A. F., et al. (2014). Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications. *International Journal of Biometeorology, 58*(4), 591–601. https://doi.org/10.1007/s00484-014-0789-5

Elmendorf, S. C., Crimmins, T. M., Gerst, K. L., & Weltzin, J. F. (2019). Time to branch out? Application of hierarchical survival models in plant phenology. *Agricultural and Forest Meteorology, 279*, 107694. https://doi.org/10.1016/j.agrformet.2019.107694

Gerst, K. L., Rossington, N. L., & Mazer, S. J. (2017). Phenological responsiveness to climate differs among four species of *Quercus* in North America. *Journal of Ecology*. https://doi.org/10.1111/1365-2745.12774

Katz, D. S. W., Dzul, A., Kendel, A., & Batterman, S. A. (2019). Effect of intra-urban temperature variation on tree flowering phenology, airborne pollen, and measurement error in epidemiological studies of allergenic pollen. *Science of the Total Environment, 653*, 1213–1222. https://doi.org/10.1016/j.scitotenv.2018.11.020

Oregon state university. (2022). PRISM climate group. Retrieved 3 June 2022 from https://prism.oregonstate.edu

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria. https://doi.org/10.1145/192593.192639

Rosemartin, A. H., Crimmins, T. M., Enquist, C. A. F., Gerst, K. L., Kellermann, J. L., Posthumus, E. E., et al. (2014). Organizing phenological data resources to inform natural resource conservation. *Biological Conservation, 173*, 90–97. https://doi.org/10.1016/j.biocon.2013.07.003

Rosemartin, A., Denny, E.G., Gerst, K.L., Marsh, R.L., Posthumus, E.E., Crimmins, T.M., and Weltzin, J.F. (2018) USA National Phenology Network observational data documentation: U.S. Geological Survey Open-File Report 2018–1060, 24 p., https://doi.org/10.3133/ofr20181060. ISSN: 2331-1258 (online)

Siljamo, P., Sofiev, M., Filatova, E., Grewling, Ł, Jäger, S., Khoreva, E., et al. (2013). A numerical model of birch pollen emission and dispersion in the atmosphere. Model evaluation and sensitivity analysis. *International Journal of Biometeorology, 57*(1), 125–136. https://doi.org/10.1007/s00484-012-0539-5

Taylor, S. D., & White, E. P. (2020). Automated data-intensive forecasting of plant phenology throughout the United States. *Ecological Applications, 30*(1), 1–10. https://doi.org/10.1002/eap.2025

Templ, B., Koch, E., Bolmgren, K., Ungersböck, M., Paul, A., Scheifinger, H., et al. (2018). Pan European phenological database (PEP725): A single point of access for European data. *International Journal of Biometeorology, 62*(6), 1109–1113. https://doi.org/10.1007/s00484-018-1512-8