Received: 25 April 2019 | Accepted: 30 April 2019

DOI: 10.1111/2041-210X.13203

APPLICATION

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

AeRobiology: The computational tool for biological data in the air

Jesús Rojo[1,2] | Antonio Picornell[3] | Jose Oteros[2]

[1]Institute of Environmental Sciences (Botany), University of Castilla-La Mancha, Toledo, Spain

[2]Center of Allergy & Environment (ZAUM), Member of the German Center for Lung Research (DZL), Technical University of Munich/Helmholtz Center, Munich, Germany

[3]Department of Plant Biology, University of Malaga, Malaga, Spain

Correspondence
Jesús Rojo
Email: jesus.rojo.ubeda@gmail.com

Funding information
University of Castilla-La Mancha [UCLM]; University of Malaga [UMA]; Zentrum Allergie und Umwelt [ZAUM]; Ministry of Education, Culture and Sport of Spain, Grant/Award Number: FPU15/01668; Helmholtz Zentrum Munich of Germany, Grant/Award Number: PFP II 2018-2020

Handling Editor: Samantha Price

Abstract

1. Aerobiological databases are constantly increasing. Many of them contain long and extensive time series of data which are very difficult and tedious to manage.
2. The development of new real-time automatic sampling devices also requires new tools to reduce time of calculations and data management. In this sense, the *AeRobiology* R package has been implemented to accelerate and facilitate these tasks.
3. This package was structured in three sections based on (a) the checking of the database, (b) calculation of the main aerobiological indexes and (c) visualization of the results.
4. The *AeRobiology* package contains numerous functions which, in conjunction, solve the main general tasks that scientists must assume for the analysis of the biological data.
5. The package is freely distributed under GNU General Public License and can directly be installed from CRAN (http://cran.r-project.org/). The reference manual is available at https://cran.r-project.org/web/packages/AeRobiology. Contact: aerobiology.package@gmail.com.

KEYWORDS
aerobiology, automation, bioaerosols, bioinformatics, computational biology, pollen diversity, R programming

1 | INTRODUCTION

The analysis of the biotic content of the troposphere is very interesting from ecological, agronomic and medical points of view (Oteros et al., 2019; Recio et al., 2018; Romero-Morte, Rojo, Rivero, Fernández-González, & Pérez-Badia, 2018). AeRobiology is the scientific field based on the study of abundance and dynamics of bioaerosols (pollen, spore, bacteria, virus …) in the air. Different devices have been used all over the world for monitoring airborne biological particles, but during the last decades a standardized methodology has been established in Europe based on the use of Hirst-type volumetric traps (Buters et al., 2018; Galán et al., 2014). The foundation of the aerobiological monitoring networks in Europe and other parts of the world during the recent decades has increased the amount of scientific publications related to aerobiological topics (Beggs, Šikoparija, & Smith, 2017).

New aerobiological methods with automatic sampling devices have been developed over the last years and new devices are under improvement (Crouzy, Stella, Konzelmann, Calpini, & Clot, 2016; Kawashima et al., 2017; Oteros et al., 2015). Such novel techniques will require new methods for analysing the huge amount of real-time aerobiological data provided (Ghitarrini et al., 2018). In any case, conventional and real-time methods need faster techniques for analysing the aerobiological data. Programming in an environment as R Software (R Core Team, 2018) is an efficient solution to interpret the results from the raw data (e.g., Big Data analysis, generation of reports or developing forecasting models). In addition, other advantages of using programming for data science are the reproducibility of the results or the robustness of the analysis against human errors.

Some functionalities such as the definition of the pollen season or the replacement of outliers in aerobiological data were previously implemented in R Software by the *pollen* package (Nowosad, 2018). Other packages related to air pollution analysis such as *openair* are helpful for modelling (Carslaw & Ropkins, 2012). *forecast* package is also useful for general statistical methods (e.g., time series analysis) (Hyndman & Khandakar, 2008). However, a new tool was necessary to manage and visualize aerobiological data, including tasks such as checking the aerobiological time series, filling the missing gaps in databases, elaborating the main calculations or multiple visualizations of the results. Following these objectives, the *AeRobiology* package has been developed for R Software (R Core Team, 2018).

2 | STRUCTURE OF THE AEROBIOLOGY PACKAGE

The functions of the package *AeRobiology* are structured in three sections based on their functionality (Table 1): (a) checking the data quality, (b) data analysis and (c) visualization of the results.

2.1 | Checking the data quality

The required data input in the functions of the package consists of a database of biological particles in air structured as a time series, that is, different types of bioaerosol measured along a time sequence. This dataset for the *AeRobiology* package must include a first column with dates and the rest of columns belonging to each bioaerosol type. Presumably, *AeRobiology* package will be applied to manage databases for biological data in the air, despite several general functionalities, such as quality control or interpolation of time series, could also be applied to other environmental data (e.g., meteorological time series).

In a global scientific discipline as aerobiology, it is very important to maintain a quality control that allows the data to be comparable and reproducible. This quality control have been carried out in Europe to check each step of the sampling procedures (Galán et al., 2014; Oteros, Galán, Alcázar, & Domínguez-Vilches, 2013). Moreover, the *quality_control* function from the *AeRobiology* package was implemented in order to check the quality of an historical database before data analysis. The dataset is checked searching for missing data within time series and regarding to certain criteria, for example, number and length of the gaps and position of the gaps during the season. Then, the user may fill the gaps according to different methods implemented in the *interpollen* function: application of the moving average of the daily concentrations; application of different regressions models (linear or spline regressions); interpolation according to the seasonality of the historical database; and interpolation of the missing data of a station by modelling its airborne particulate matter using data from a neighbour station. These functions were implemented for aerobiological tasks but other environmental time series could be checked and completed using this R package (e.g., databases of inorganic air pollutants or meteorological time series).

TABLE 1 List of main functions included on the *AeRobiology* package

Section	Functions	Description
Checking the data quality	*quality_control*	Checking the quality of an historical database of several types of bioaerosol
	interpollen	Filling the gaps of the missing data according to different implemented methods of interpolation
Data analysis	*calculate_ps*	Estimation of the main parameters of the pollen or spore season
	pollen_calendar	Pollen calendar using different commonly used methods from a historical pollen database
	analyse_trend	Trends analysis for the main seasonal indexes of the reproductive season (start-dates, peak-dates, end-dates and amount of bioaerosols)
Visualization of the results	*iplot_abundance*	Plot of the relative abundance of the bioaerosols in the air
	iplot_pheno	Phenological plot based on phenological parameters such as start and end-dates of the reproductive season
	iplot_pollen	Interactive plotting pollen data (one reproductive season)
	iplot_years	Interactive plotting pollen data (one type of bioaerosol)
	plot_ps	Plotting the reproductive season of a single type of bioaerosol
	plot_normsummary	Plotting the amplitude of several reproductive seasons
	plot_summary	Plotting the data during several reproductive seasons
	plot_trend	Calculating and plotting trends from a historical pollen database
	plot_hour	Intradiurnal analysis and plotting of the hourly patterns of pollen concentration
	plot_heathour	Graphical representation of the hourly patterns of pollen concentrations as heatplot

2.2 | Data analysis

The dynamics of the bioaerosols are determined by phenological parameters such as the dates of onset and end of the period of the year when a given particle is registered in the air. Nevertheless, the methods to calculate these parameters are subject to controversy (Bastl, Kmenta, & Berger, 2018) and the use of one or another method depends on the purpose and the type of bioaerosol (Jato et al., 2006). All the actual methods used to estimate the pollen season have been implemented in the *calculate_ps* function, obtaining easily the results about the phenological parameters and pollen intensity measurements (Andersen, 1991; Galán, García-Mozo, Cariñanos, Alcázar, & Domínguez-Vilches, 2001; Pfaar et al., 2017; Ribeiro, Cunha, & Abreu, 2007). Moreover, a new method is proposed for the first time in this package by the authors. According to this new method, the definition of the pollen season is established when the moving average of the aerobiological series reaches a threshold. This method avoids the high variability introduced by daily fluctuations of the concentrations of airborne particles, which can make more difficult the estimation of the phenological dates. Therefore, *AeRobiology* package represents a crucial contribution to the automation of the analysis of a key aerobiological issue such as the definition of the pollen season.

Another important aerobiological task is the calculation of the pollen or spore calendar, which is a graphical representation of the biological spectrum of the atmosphere at a certain location throughout the year. These calendars include information about phenology and intensity of the concentration of different types of bioaerosol in the air. The *pollen_calendar* function allows to calculate the calendar from a historical aerobiological database using the most common methods in the aerobiology field (O'Rourke, 1990; Rojo et al., 2016; Spieksma, 1991; Werchan, Werchan, & Bergmann, 2018).

In addition, the general spectrum of bioaerosols in the air may be analysed by using the *iplot_abundance* function. Thus, the relative abundances (as percentages) of the specific biological particles in the air are calculated. In a long-term temporal context, studies focused on trends of time series are very useful when analysing extensive historical series of environmental data. For this reason, the *analyse_trend* function was included in the *AeRobiology* package. These functions allow to carry out trend analysis of the main seasonal indexes over several years, a crucial analysis to evaluate the influence of climatic variations over the dynamics of the bioaerosols (Recio et al., 2018).

2.3 | Visualization of the results: static and interactive plots

AeRobiology package offers numerous options to visualize results. On the one hand, a proper visualization of the raw aerobiological data may be a great help before analysing the data (*plot_normsummary* and *plot_summary* functions). In addition, the user may carry out a comparison between the airborne daily concentrations of different years and types of bioaerosols (*iplot_pollen* and *iplot_years* functions) or may want to graphically visualize the phenological parameters (*iplot_pheno* function). Intradiurnal analysis and graphical representation of the hourly patterns of pollen concentration may be performed with *plot_hour* and *plot_heathour* functions. Furthermore, *plot_ps* and *pollen_calendar* functions provide different graphical representations. Most of the functions included in the *AeRobiology* package provide some kinds of visualization, either to internally check the results as well as to produce elegant figures that the users may publish. The users may choose between static (Wickham, 2016) and interactive plots (Chang, Cheng, Allaire, Xie, & McPherson, 2018; Sievert, 2018) for improving the interpretation of the results.

3 | EXAMPLE OF APPLICATION

This example illustrates how the *AeRobiology* package is applied for the management and analysis of the pollen database of Munich (period 2010–2015). This database was provided by the research team of Prof. Jeroen Buters (Zentrum Allergie und Umwelt, ZAUM, directed by Prof. Carsten B. Schmidt-Weber).

```
> library (AeRobiology)
> data(munich_pollen)
```

Then, in the Figure 1 the outputs obtained by several functions of the *AeRobiology* package are shown. In the first step, it is very interesting to check the quality of the data (Figure 1.1) and to fill the gaps of the missing data using a spline regression (Figure 1.2):

```
> quality_control(munich_pollen, int.window = 2, perc.miss = 20) # Figure 1.1
> interpollen(munich_pollen, method = "spline", spar = 0.2) # Figure 1.2
```

The users may perform numerous relevant calculations in the field of aerobiology, for example to calculate phenological parameters of the season or generate a calendar that represents the abundance and dynamics of the bioaerosols in the atmosphere. This example shows the calculation of the season (start, peak and end-dates) for the *Betula* pollen season following the criterion proposed by Pfaar et al. (2017) for birch (Figure 1.3). Furthermore, Figure 1.4 shows the output of the calculation of the calendar using a violin plot design (O'Rourke, 1990):

```
> plot_ps(munich_pollen, pollen.type = "Betula", year = 2011, method = "clinical", type = "birch") # Figure 1.3
> calendar.plot <- pollen_calendar(munich_pollen, method = "violinplot") # Figure 1.4
```

Different examples of visualization of the data are shown in Figure 1 such as graphs about relative abundance of the bioaerosols

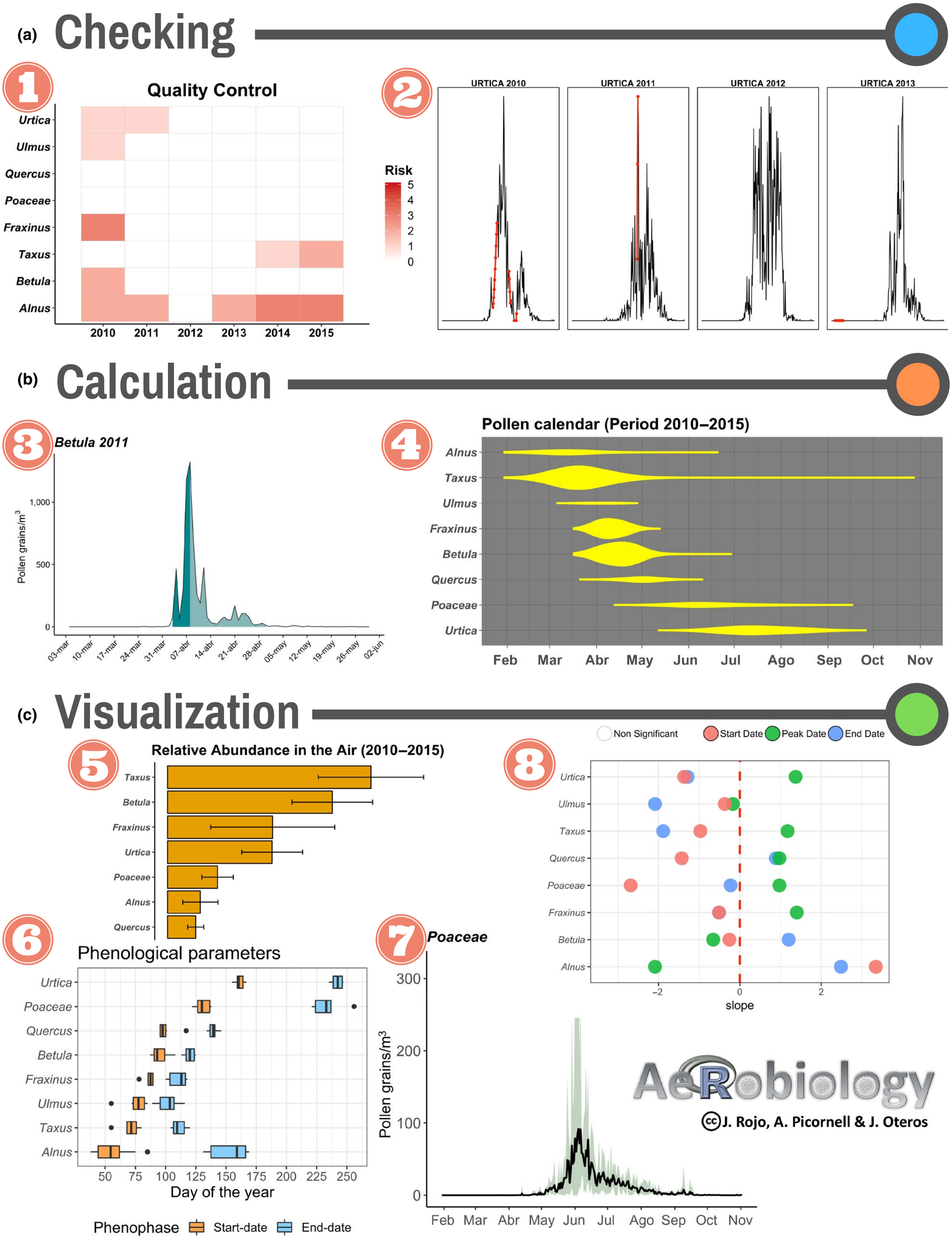

FIGURE 1 Example of the different tasks implemented in the *AeRobiology* package. All these graphs may be obtained by running the script provided in the example of the text: 1 Quality control; 2 Interpolation of the pollen data for filling the missing data; 3 Estimation of the pollen season for *Betula* pollen; 4 Pollen calendar; 5 Relative abundances of the bioaerosols; 6 Main phenological parameters; 7 Graph for the average and amplitude of the Poaceae pollen; 8 Trends of the pollen time series. Pollen data obtained at Munich (Zentrum Allergie und Umwelt, ZAUM) and provided in this package

(Figure 1.5), phenological graphs (Figure 1.6), graphs showing the average daily concentrations and the amplitude of a specific biological particle (Figure 1.7) or graphs about trends of the data (Figure 1.8):

```
> iplot_abundance(munich_pollen, n.types = 7, type.plot = "static") # Figure 1.5
> iplot_pheno(munich_pollen, method = "percentage", perc = 95, type.plot = "static") # Figure 1.6
> plot_normsummary(munich_pollen, pollen = "Poaceae", color.plot = "darkgreen") # Figure 1.7
> analyse_trend(munich_pollen, export.plot = FALSE, export.result = FALSE)[[1]] # Figure 1.8
```

4 | CONCLUSIONS

The *AeRobiology* R package means an important contribution to the scientific field of any aerosol science, especially for aerobiology, since it provides proper solutions for the automatization of calculations and time-consuming procedures of the data analysis. This advance in efficiency and effectiveness for the data analysis is especially relevant in this historical moment where for the first time technological development allows us to automatically get reliable and big databases about the biological quality of the air.

ACKNOWLEDGEMENTS

We specially acknowledge the team of Prof. Jeroen Buters (Christine Weil & Ingrid Weichenmeier) and the Zentrum Allergie und Umwelt (ZAUM, directed by Prof. Carsten B. Schmidt-Weber) for the contribution of the pollen data from Munich. We acknowledge our institutions (University of Castilla-La Mancha [UCLM], University of Malaga [UMA] and ZAUM) for their support. A.P. was supported by a predoctoral grant financed by the Ministry of Education, Culture and Sport of Spain, in the Program for the Promotion of Talent and its Employability (FPU15/01668). J.O. was supported by a Postdoctoral grant of Helmholtz Zentrum Munich PFP II 2018-2020. We are also grateful to Dr. Jakub Nowosad for many helpful suggestions to improve the R package.

AUTHORS' CONTRIBUTIONS

J.R., A.P., and J.O. designed the R package; J.R., A.P., and J.O. contributed to developing and checking the R functions; and J.R., A.P., and J.O. wrote and revised this manuscript.

DATA ACCESSIBILITY

The package is freely distributed under GNU General Public License (GPL) and can directly be installed from CRAN (http://cran.r-project.org/). The reference manual is available at https://cran.r-project.org/web/packages/AeRobiology.

ORCID

Jesús Rojo https://orcid.org/0000-0002-3627-130X

REFERENCES

Andersen, T. B. (1991). A model to predict the beginning of the pollen season. *Grana*, *30*, 269–275. https://doi.org/10.1080/00173139109427810

Bastl, K., Kmenta, M., & Berger, U. E. (2018). Defining pollen seasons: Background and recommendations. *Current Allergy and Asthma Reports*, *18*. https://doi.org/10.1007/s11882-018-0829-z

Beggs, P. J., Šikoparija, B., & Smith, M. (2017). Aerobiology in the International Journal of Biometeorology, 1957–2017. *International Journal of Biometeorology*, *61*(Suppl. 1), 51–58. https://doi.org/10.1007/s00484-017-1374-5

Buters, J. T. M., Antunes, C., Galveias, A., Bergmann, K. C., Thibaudon, M., Galán, C., ... Oteros, J. (2018). Pollen and spore monitoring in the world. *Clinical and Translational Allergy*, *8*. https://doi.org/10.1186/s13601-018-0197-8

Carslaw, D. C., & Ropkins, K. (2012). openair — An R package for air quality data analysis. *Environmental Modelling & Software*, *27–28*, 52–61. https://doi.org/10.1016/j.envsoft.2011.09.008

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2018). *shiny: Web Application Framework for R. R package version 1.1.0*. Retrieved from https://CRAN.R-project.org/package=shiny

Crouzy, B., Stella, M., Konzelmann, T., Calpini, B., & Clot, B. (2016). All-optical automatic pollen identification: Towards an operational system. *Atmospheric Environment*, *140*, 202–212. https://doi.org/10.1016/j.atmosenv.2016.05.062

Galán, C., García-Mozo, H., Cariñanos, P., Alcázar, P., & Domínguez-Vilches, E. (2001). The role of temperature in the onset of the Olea europaea L. pollen season in southwestern Spain. *International Journal of Biometeorology*, *45*(1), 8–12. https://doi.org/10.1007/s004840000081

Galán, C., Smith, M., Thibaudon, M., Frenguelli, G., Oteros, J., Gehrig, R., ...EAS QC Working Group. (2014). Pollen monitoring: Minimum requirements and reproducibility of analysis. *Aerobiologia*, *30*, 385–395. https://doi.org/10.1007/s10453-014-9335-5

Ghitarrini, S., Pierboni, E., Rondini, C., Tedeschini, E., Tovo, G. R., Frenguelli, G., & Albertini, E. (2018). New biomolecular tools for aerobiological monitoring: Identification of major allergenic Poaceae species through fast real-time PCR. *Ecology and Evolution*, *8*, 3996–4010. https://doi.org/10.1002/ece3.3891

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *27*. https://doi.org/10.18637/jss.v027.i03

Jato, V., Rodríguez-Rajo, F. J., Alcázar, P., De Nuntiis, P., Galán, C., & Mandrioli, P. (2006). May the definition of pollen season influence aerobiological results? *Aerobiologia*, *22*(1), 13. https://doi.org/10.1007/s10453-005-9011-x

Kawashima, S., Thibaudon, M., Matsuda, S., Fujita, T., Lemonis, N., Clot, B., & Oliver, G. (2017). Automated pollen monitoring system using laser optics for observing seasonal changes in the concentration of total airborne pollen. *Aerobiologia*, *33*, 351–362. https://doi.org/10.1007/s10453-017-9474-6

Nowosad, J. (2018). *pollen: Analysis of aerobiological data. R package version 0.71.0*. Retrieved from https://CRAN.R-project.org/package=pollen

O'Rourke, M. K. (1990). Comparative pollen calendars from Tucson, Arizona: Durhamvs. Burkard Samplers. *Aerobiologia*, *6*, 136. https://doi.org/10.1007/BF02539105

Oteros, J., Bartusel, E., Alessandrini, F., Núñez, A., Moreno, D. A., Behrendt, H., ... Buters, J. (2019). Artemisia pollen is the main vector for airborne endotoxin. *Journal of Allergy and Clinical Immunology*, *143*, 369–377. e5. https://doi.org/10.1016/j.jaci.2018.05.040

Oteros, J., Galán, C., Alcázar, P., & Domínguez-Vilches, E. (2013). Quality control in bio-monitoring networks, Spanish Aerobiology Network. *Science of the Total Environment*, *443*, 559–565. https://doi.org/10.1016/j.scitotenv.2012.11.040

Oteros, J., Pusch, G., Weichenmeier, I., Heimann, U., Möller, R., Röseler, S., … Buters, J. T. M. (2015). Automatic and online pollen monitoring. *International Archives of Allergy and Immunology*, *167*, 158–166. https://doi.org/10.1159/000436968

Pfaar, O., Bastl, K., Berger, U., Buters, J., Calderon, M. A., Clot, B., … Bergmann, K. C. (2017). Defining pollen exposure times for clinical trials of allergen immunotherapy for pollen-induced rhinoconjunctivitis—An EAACI position paper. *Allergy*, *72*, 713–722. https://doi.org/10.1111/all.13092

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Recio, M., Picornell, A., Trigo, M. M., Gharbi, D., García-Sánchez, J., & Cabezudo, B. (2018). Intensity and temporality of airborne Quercus pollen in the southwest Mediterranean area: Correlation with meteorological and phenoclimatic variables, trends and possible adaptation to climate change. *Agricultural and Forest Meteorology*, *250–251*, 308–318. https://doi.org/10.1016/j.agrformet.2017.11.028

Ribeiro, H., Cunha, M., & Abreu, I. (2007). Definition of main pollen season using a logistic model. *Annals of Agricultural and Environmental Medicine: AAEM*, *14*, 259–264.

Rojo, J., Rapp, A., Lara, B., Sabariego, S., Fernández-González, F., & Pérez-Badia, R. (2016). Characterisation of the airborne pollen spectrum in Guadalajara (central Spain) and estimation of the potential allergy risk. *Environmental Monitoring and Assessment*, *188*, 130. https://doi.org/10.1007/s10661-016-5129-2

Romero-Morte, J., Rojo, J., Rivero, R., Fernández-González, F., & Pérez-Badia, R. (2018). Standardised index for measuring atmospheric grass-pollen emission. *Science of the Total Environment*, *612*, 180–191. https://doi.org/10.1016/j.scitotenv.2017.08.139

Sievert, C. (2018). *plotly for R*. Retrieved from https://plotly-book.cpsievert.me

Spieksma, F. T. M. (1991). Regional European pollen calendars. In G. D'Amato, F. T. M. Spieksma, & S. Bonini (Eds.), *Allergenic pollen and pollinosis in Europe* (pp. 49–65). Oxford, UK: Blackwell Science Publishers.

Werchan, M., Werchan, B., & Bergmann, K.-C. (2018). German pollen calendar 4.0—Update based on 2011–2016 pollen data. *Allergo Journal International*, *27*(3), 69–71. https://doi.org/10.1007/s40629-018-0055-1

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4

How to cite this article: Rojo J, Picornell A, Oteros J. AeRobiology: The computational tool for biological data in the air. *Methods Ecol Evol*. 2019;10:1371–1376. https://doi.org/10.1111/2041-210X.13203