

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Advanced Methods for Spatial Analysis of Bioaerosol Long-Range Transport Processes

*Daniel A. Pickersgill, Hartmut Müller  
and Viviane R. Després*

## Abstract

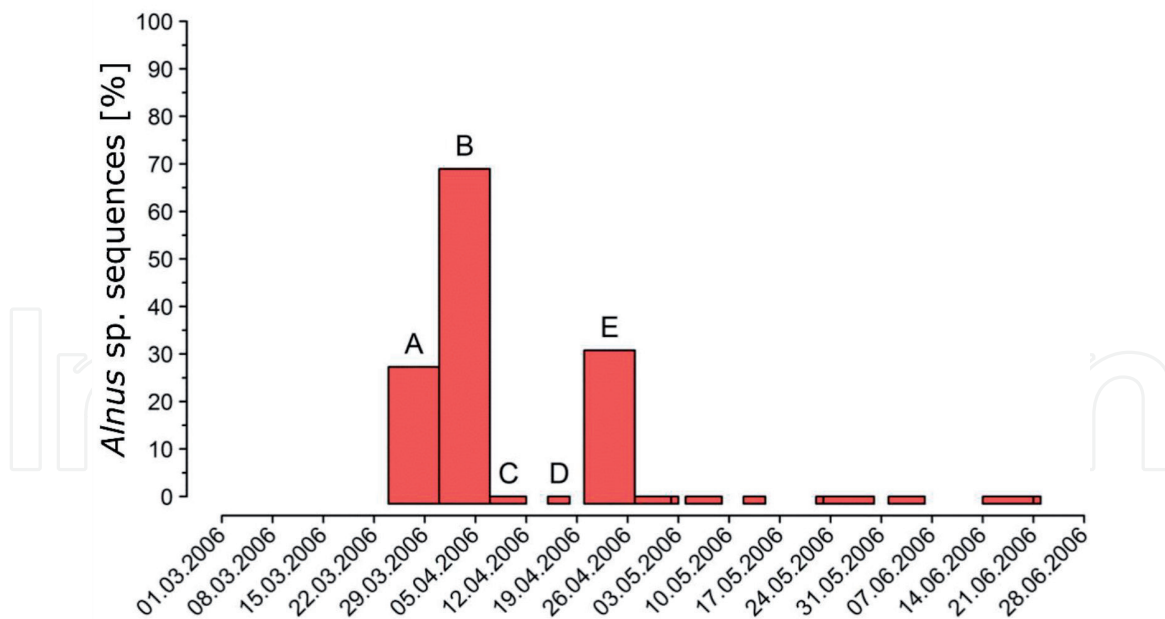
Research on bioaerosol is still in its infancy. The dynamics and, therefore, the effects on atmospheric processes and the biosphere are often underestimated, or have not yet been sufficiently investigated. Atmospheric models such as FLEXPART and HYSPLIT enable researchers to simulate the transport of particles in the atmosphere and provide information on where air-parcels originate from. In the following, we present two methods for combining results of these models with spatial information, e.g., about vegetation. The first method shows how spatial CORINE land cover distribution can be analyzed within the boundaries of HYSPLIT trajectories. In a second method, FLEXPART simulations are used in combination with COSMO rain data and tree maps to generate maps that indicate the potential origin of bioaerosol for selected periods of time.

**Keywords:** bioaerosol, wind trajectories, FLEXPART, GIS, ArcPy

## 1. Introduction

Bioaerosol, more precisely primary biological aerosol particles (PBAP), are particles of biological origin, smaller than 100  $\mu\text{m}$ , that are released into the atmosphere. These include viruses, bacteria, pollen and fungal spores, small and cell fragments, and excrements from organisms [1]. The diameter of PBAP ranges from a few nanometers, such as cell fragments, proteins and viruses, to the upper size boundary seen for many plant pollen [2].

PBAPs emitted into the atmosphere are subject to many physical factors resulting in a considerable influence on the atmospheric residence time. Many meteorological factors such as wind speed, wind direction, convection, temperature and relative humidity influence the residence time which can result in PBAP covering long distances in the atmosphere. In addition, the residence time depends strongly on the aerodynamic diameter of the particles [3]. Removal from the atmosphere can happen by dry and wet deposition. For dry deposition, the PBAPs are removed by sedimentation, while for wet deposition they are washed out by precipitation. PBAP have a direct influence on humans when they act as allergens or pathogens [4, 5]. As an increasing proportion of the population suffers from allergies, the importance of accurate pollen prediction increases. Predictions are currently generated by the Deutscher Wetterdienst DWD (German Weather Service), which issues warnings



**Figure 1.** Proportion of *Alnus* sp. (alder) DNA sequences in the total number of isolated plant sequences collected between March and June 2006 in Mainz, Germany (49°59'31.36"N 8°14'15.22"E). In addition to the percentage (y-axis) the time period of the individual air samples is visualized by the column width (x-axis).

for very large areas if a certain pollen concentration in the air is present. The pollen concentration is determined by the German Pollen Information Service Foundation (<http://www.pollenstiftung.de>) through microscopic analysis. The DWD processes these data using weather models and produces warning maps for pollen. Due to the long processing path and the small number of pollen collecting stations, large inaccuracies occur. A more precise knowledge of the emission and transport processes could lead to considerably improved pollen predictions and thus help patients to be better prepared.

The analysis of PBAP, especially concerning their origin, poses special challenges for science. Atmospheric transport is a highly complex issue. The composition of emissions, strongly depend on the biotope, which is highly variable in space, and in time, dependent on the season. **Figure 1** shows an example of the proportion of *Alnus* sp. (alder) DNA sequences in the total number of isolated plant sequences found on analyzed weekly air sample filters in spring 2006. As can be seen, the proportion decreases from sample B to C from ~70 to 0% and increases from sample D to E from 0 to ~30%. These results raise the question whether such high variances in such a short time frame can be explained by air movement.

In the following sections we will describe methods to identify potential areas of origin of PBAP starting from a firmly defined sampling location and using Lagrangian back-trajectory transport model. By using two-dimensional raster data containing information on potential emission sources we will show methods to identify correlations with observed PBAP data by modelling the potential transport processes of the particles in the atmosphere.

## 2. Data

### 2.1 PBAP data

The aerosols were collected on glass fiber filters for over a year. The data collection station was established at about 20 m height above ground, on the roof of

the old Max Planck Institute for chemistry, located at the campus of the Johannes Gutenberg University Mainz (49°59'31.36"N, 8°14'15.22"E). An air volume of 0.3 m<sup>3</sup> min<sup>-1</sup> was filtered by a high-volume sampler. The operating time was between 1 and 7 days, which was equivalent to an air volume of about 430–3000 m<sup>3</sup>. The DNA was extracted from the air filter samples, the plant DNA isolated and taxonomically identified. The previously unpublished data were kindly provided by Isabell Müller-Germann [6]. In pollen flight periods a large percentage of plant bioaerosol are formed by pollen. In the following, therefore, the simplified term pollen is used, even though other plant particles may have been measured. An exact description of the methodology can be found in Fröhlich-Nowoisky et al. [7], with the modification that plant-specific primers were used.

## 2.2 HYSPLIT

HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) [8] has been used for about 30 years by the atmospheric science community to calculate atmospheric transport, dispersion, chemical transformation and deposition. A major task of the system is the calculation of backward trajectories to determine the origin of air masses. This makes it possible to establish cause-effect relationships. Forward trajectories, on the other hand, are used to predict the propagation of, e.g., volcanic ash or radioactive particles. HYSPLIT uses a hybrid calculation approach and uses both Lagrangian and Eulerian methods [9]. By using the READY system (<http://ready.arl.noaa.gov/index.php>), HYSPLIT calculations can even be performed online. The meteorological basis of the calculations is the globally available "Global Data Assimilation System" (GDAS) provided by the National Oceanic and Atmospheric Administration NOAA with an area resolution of 1 × 1 and a vertical subdivision into 23 layers reaching a height of 26.5 km above ground. The results of a back trajectory calculation are single line features obtained for each start time of the calculation. The line feature contains the four-dimensional information (place and time) on the origins of an air-parcel that was measured at a defined point in time t<sub>0</sub>. The result of the calculation can be generated in zipped Keyhole Markup Language (.kmz) or Shapefile (.shp) format, which eases further processing in Geographic Information Systems.

## 2.3 FLEXPART

FLEXible PARTicle dispersion model (FLEXPART) [10] is a Lagrangian particle dispersion model for calculating the propagation of air masses over long distances. The results are obtained in the multidimensional Network Common Data Form (NetCDF) format which is often used in meteorology [11]. The result files contain spatial information for five air layers and two particle types (tracer). One particle type has an atmospheric half-life of 12 hours, which approximately is equal to the average atmospheric retention time for PBAP of pollen size, when considering dry and wet deposition. For comparison reasons, no half-life parameter was set for the second particle type. It thus represents a so-called air tracer. For every single day of a 10 weeks' period a simulation was performed with a spatial resolution of 10 × 10 km.

The FLEXPART model calculation is based on four-dimensional meteorological raster data sets, which define the level of resolution for the results. The DWD used analysis data from the COSMO-EU model [12] for this purpose (<http://www.dwd.de>). The data resolution is 0.0625° (~7 km) in horizontal direction. In vertical direction, the raster cells are ordered into 40 layers reaching a height of 22.5 km above ground. The raster grid is rotated against the north direction, with the North Pole at 40° latitude and -170° longitude.

In addition, rain areas (see Section 3.2.3) were extracted from the COSMO-EU model. These areas contain information about location, quantity and type of rain (convective or scalar), given in the form of total hourly rainfall accumulated to daily sums in mm.

## **2.4 Tree species maps for European forests**

The “Tree species maps for European forests” published by the European Forest Institute [13] were used as a data source for potential pollen emission sites. These maps were generated using statistical methods such as logistic regression and Kriging and differ according to region and national forest inventory methodology. The resulting data sets represent the only European-wide mapping of tree species to date. The data for each recorded tree species is provided in the form of a GeoTiff raster layer with a resolution of  $1 \times 1$  km per grid cell. The numerical value of each grid cell is the percentage coverage of the cell with the respective tree species.

The tree species maps are available in the spatial reference system ETRS89/ETRS-LAEA (EPSG:3035). For the processing discussed here they were transformed into the geodetic reference system WGS84.

## **2.5 CORINE land cover**

Another used data source is the land use map published by the European Environment Agency’s “CORINE Land Cover” project [14]. Updates to the CORINE Land Cover (CLC) inventory, dating from 1985, are available for 2000, 2006, 2012, and 2018. The 2006 version best fit the time frame of the PBAP data and was therefore used. The map shows the type of land use for the participating 39 EU countries in  $100 \times 100$  m grid cell resolution, subdivided into 44 different land-use classes. The map is available in the spatial reference system ETRS89/ETRS-LAEA (EPSG:3035) and, therefore, was also transformed into WGS84 for processing.

# **3. Methodology**

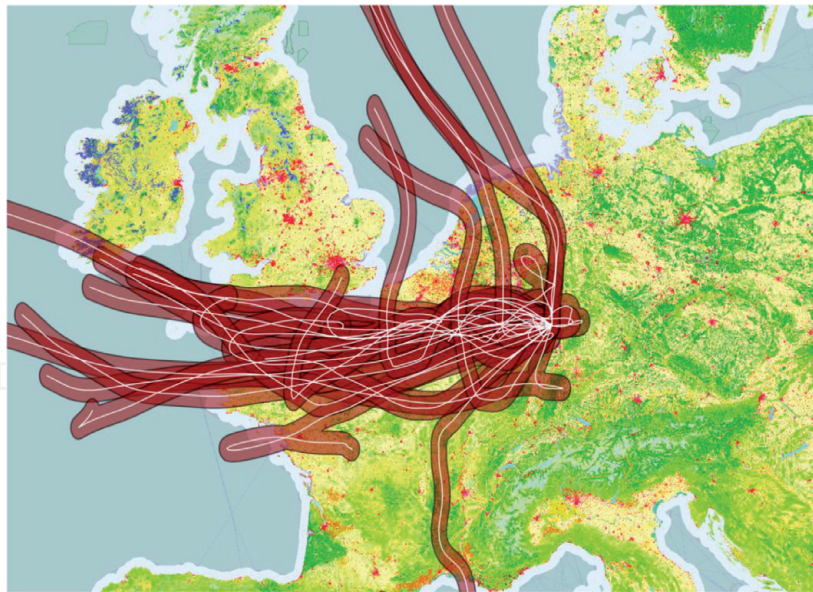
## **3.1 Land use analyses based on trajectories (HYSPLIT)**

Trajectories give information about height, direction and residence time of an air parcel. The used backward trajectories show which ground areas have been overflown before an air parcel reached the measuring point in Mainz. The area composition has a decisive influence on the composition of the collected bioaerosol.

The trajectories calculated with HYSPLIT are only available as line-features and do not provide information that could serve as a basis for area-related calculation. To solve this problem area-buffers were created for the line features. In a second step, these buffers were used to calculate the composition of the underlying CORINE land use areas by polygon overlay (**Figure 2**). The resulting area clips of the data set could then be used to statistically evaluate the pollen values.

## **3.2 Potential maps of pollen origin**

Both FLEXPART and HYSPLIT simulations generate information on the residence time of air parcels arriving at specific points of measurement. However, FLEXPART performs not only line calculations for individual points in time of the residence time as HYSPLIT trajectories do, but can generate grids of air parcel residence time for entire time periods. Therefore, in addition to location and time, one



**Figure 2.**  
*HYSPLIT trajectories (lines) and their buffers superimposed on the CORINE land cover map.*

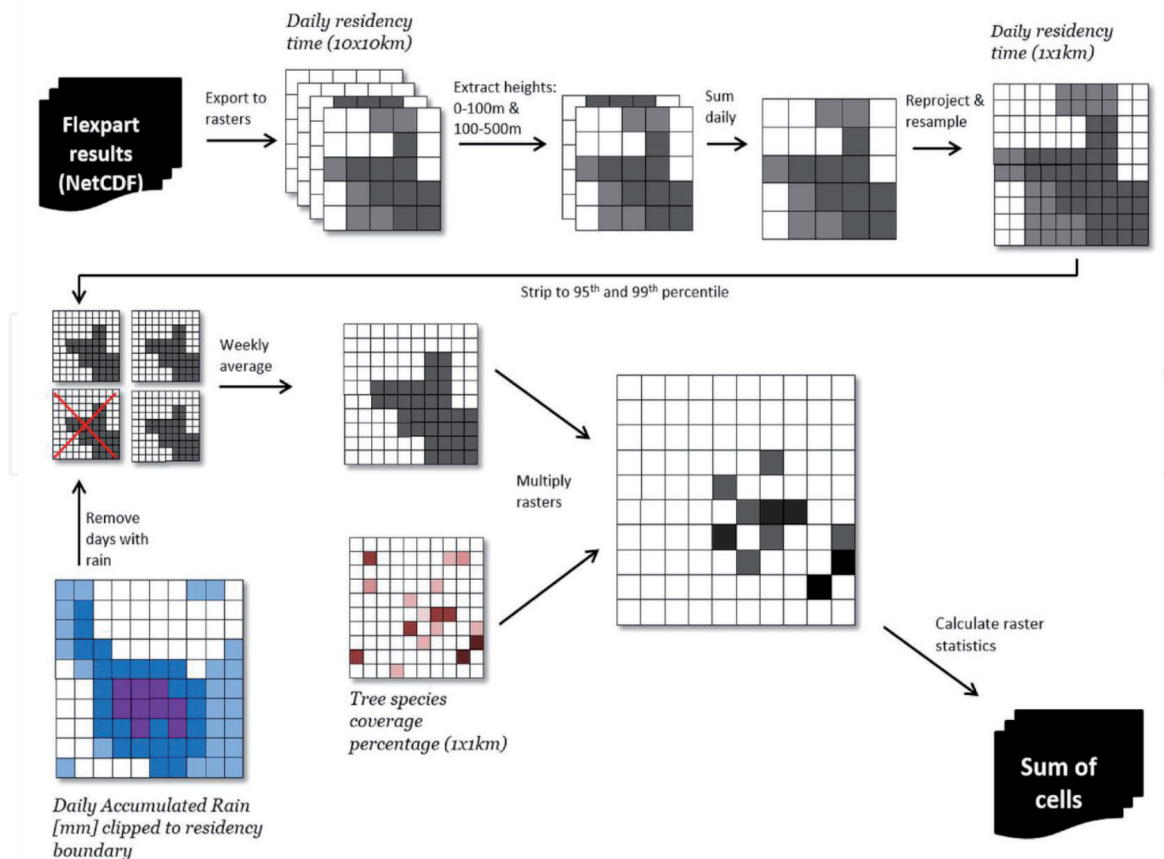
gets the area-related information on how long air parcels have resided in defined areas. Thus, different weights are assigned to areas dependent on the influence on the air parcels that reach the point of measurement.

As discussed in Section 2, the FLEXPART day simulation results for the study periods are available in multidimensional NetCDF format, the format frequently used in meteorology. Due to the high number of possible two-dimensional raster representations the interpretation of the results in view of the transport of air parcels to the point of measurement is very time-consuming. It therefore makes sense to automate the processing steps. Processing steps to be automated include the export of information from NetCDF files into a raster format suited for further processing in a GIS, the subsequent processing of the raster grids including removing days of rain, and the concatenation of this information with tree cover gradients (see Section 2). The aim is to create maps which, by combining the residence time of air parcels with the tree population, lead to new insights into the potential origin of bioaerosol.

The FLEXPART results (see **Figure 3**) are extracted by using ArcPy, a Python library that provides tools for the analysis and conversion of geographical data [15]. Using tools from the “Multidimension toolbox,” the spatial information for all required combinations of time period, height of the air parcels and tracer is extracted and saved as a raster file in GeoTiff format. Defining the required dimensions for NetCDF files with massive content is often a complex task. Therefore, an ArcGIS tool was developed that considerably simplifies the handling of dimensions for NetCDF files and data export [16]. The generated raster layers contain the residence time of the air parcels per day in a resolution of  $10 \times 10$  km.

The individual layers created for each time step must be further processed. For the following steps only the two bottom air mass layers are considered. It is assumed that the so-called atmospheric boundary layer air is a homogeneously mixed in respect to the contained particles [17]. Solar radiation causes the boundary layer to build up during the course of the day. Depending on the conditions, the height of the boundary layer varies, but should normally be higher than 500 m above ground. Consequently, the two bottom air layers (height above ground 0–100 m, and height above ground 100–500 m) are merged into one single layer.

The tree species coverage data are available in a different projection system and at a higher resolution ( $1 \times 1$  km). Therefore, the FLEXPART grids are re-projected



**Figure 3.** Schematic representation of the most relevant processing steps.

and resampled according to the daily residency time. Re-projection eliminates pixel inconsistencies between different projection systems and thus prevents errors in raster operations such as multiplication. Although the resampling does not improve the FLEXPART data resolution, the resampled data can be used for subsequent analysis with tree species coverage data.

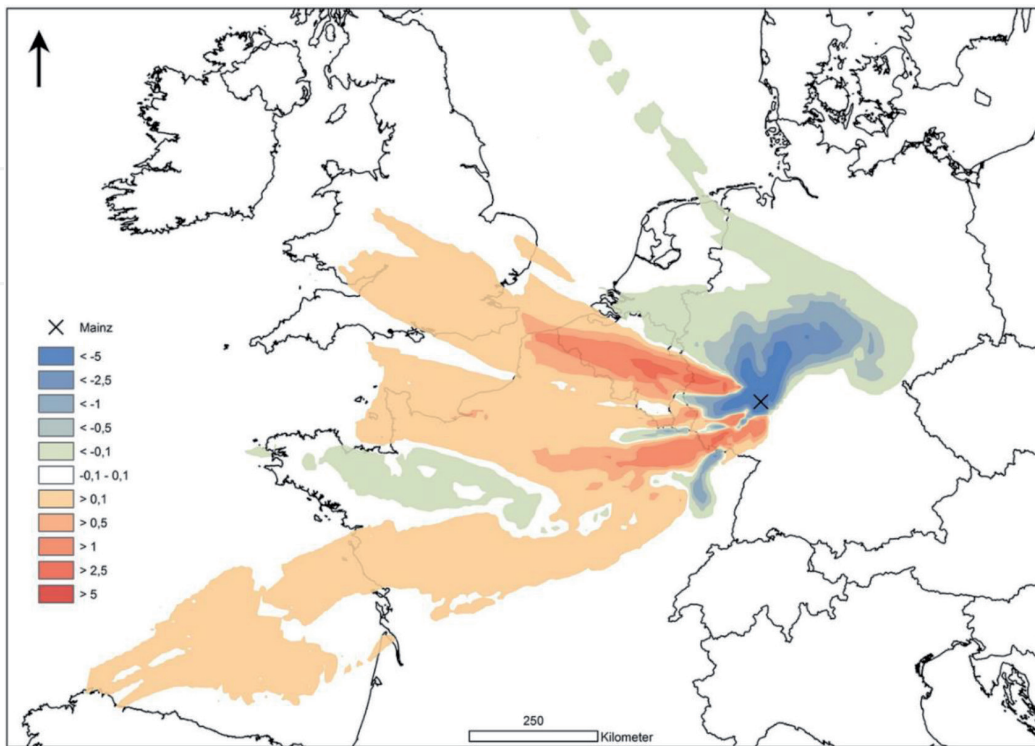
The temporal resolution of the pollen data analysis is usually 7 days; therefore, the daily residence times must also be averaged. Before doing so, days with rain are identified and excluded from the means formation to eliminate the influence of such days on the distribution of pollen in the air due to wet deposition [18].

For this purpose, hour values of rain fields which are available in grid format in the COSMO model are used (see Section 2). The data is provided in COSMO's native model grid and must first be rotated spatially to match the spatial reference system used in this study. In order to avoid the processing of uninvolved areas the rain grids are clipped to the spatial extent of the FLEXPART data sets for the considered periods, day sums are formed for the remaining rain cells. If the value of a day sum exceeds 2 mm per day, the air residency time for this day is removed, i.e., not considered in the week average value calculation.

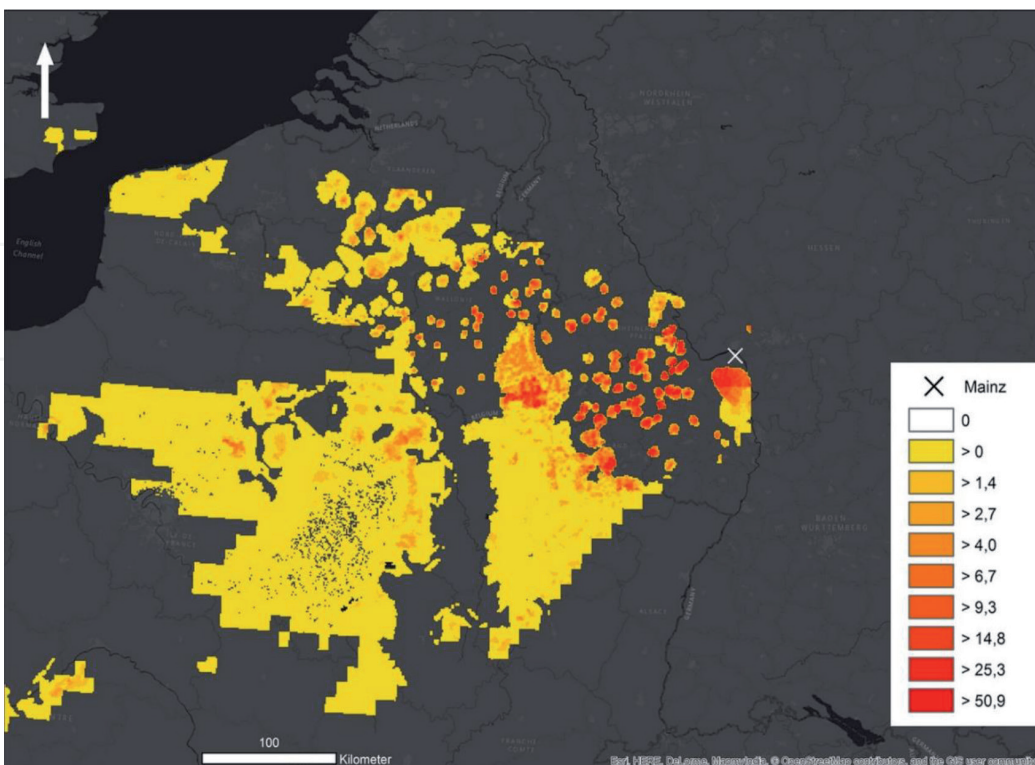
The weekly average is the mean residence times of the remaining days. **Figure 4** shows examples of results for two different weeks. The figure illustrates the difference of these 2 weeks; as already mentioned, only dry days were taken into account.

A first attempt was made to derive land cover statistics for the areas under the air parcels, in analogy to the trajectory method. However, due to the large spatial extent following from the calculation of week average values, the statistics generated were not significant. Therefore, it was not possible to confine areas of potential origin of pollen with sufficient probability. A further attempt to produce significant statistics was made by extracting only areas with 99 and 95 percentile values, but with the persistent problem that all extracted areas had the same weight.

In a second approach, the weekly residence times of the air parcels was multiplied by the percentage of individual tree species coverage such as *Alnus* sp. or *Betula* sp. This approach yields high values for areas characterized by long residency periods and high coverage rates, low values for areas with short residency periods



**Figure 4.**  
Comparison of 2 weeks. The difference of residency time in hours per week is shown. A predominance of the first week is shown with a red color, the second in blue color.



**Figure 5.**  
Potential map. Red: high potential, yellow: low potential, transparent: no potential, because of no tree population or no residency time of an air parcel. Values result from multiplication of tree species coverage in percent and average residency time of air in hours per week.



and low coverage rates. This approach not only suppresses areas with a short residency time of an air parcel, which does not significantly affect the air composition at the point of measurement, but also areas with a long residency time of an air parcel, in areas without population of the respective tree species. The resulting potential maps show tree species populations that could be potential sources of detected pollen at the point of measurement in a given week (see **Figure 5**).

The resulting weekly potential maps can now be visually interpreted. Due to the large number of possible potential grids, however, it is useful to determine a numerical value for the purpose of automated evaluation. Summing up all grid cell values should result in high values for weeks with a high pollen volume.

#### **4. Results and discussion**

The described methods open up new possibilities for the analysis of long-distance transport processes in bioaerosol research. The research question was to establish a quantifiable link between temporally dynamic PBAP data, which are sampled at a spatially static point, and the spatially dynamic but temporally constant GIS raster maps. Both the temporally and the spatially resolved dispersion models function as connecting elements. The observation of near-ground air movements makes it possible to strongly limit the potential emission areas. Flexible calculations with high temporal resolution for each sampling period can be carried out on the basis of raster maps.

In contrast to HYSPLIT trajectories, the residence times of air parcels generated by FLEXPART is not specified at individual points in time, but for entire periods—instead of line features, grids are generated. This makes it possible to assign lower weights to areas with a short residence time of air parcel. In addition, FLEXPART takes the air movements in the near-ground layers of the atmosphere into account in a more realistic manner. Another advantage over HYSPLIT is the higher spatial and temporal resolution of the underlying data set (GDAS vs. COSMO-EU).

By multiplying by the percentage of tree species coverage, potential emission areas can be emphasized or omitted, depending on the occurrence of observed tree species. Consequently, only tree species populations that could be potential sources for pollen deposition within a given week are present in the resulting potential maps. Another advantage is the consideration of rain; it can be assumed that convective and continuous scalar showers efficiently remove aerosols at the point of origin.

The used greatly simplified filtering of rain events is still a source of uncertainty. If the amount of rain on a day exceeded the threshold value, the residence time of all air parcels for that day is not taken into account when calculating the average week value. A fixed threshold value is used and the temporal distribution and the type of precipitation are not considered. A short period of strong convective precipitation with a total of more than 2 mm is equal to a scalar precipitation spread over several hours. In addition, the spatial distribution of precipitation is not taken into account. Although only rain events within significant air parcels are considered, the same weight is assigned to all locations within these areas. Given the large spatial extent of the air masses, locations with substantially different distances to the point of measurement are equally weighted. Likewise, if a precipitation event is locally limited, all non-influenced air parcels for this day are removed.

#### **5. Conclusions and outlook**

Atmospheric convection and transport processes are highly complex issues for which many new insights have been gained in recent years. Hence, highly complex

interactions, such as the transport of bioaerosol in the atmosphere including its physical and biological effects on different areas of the ecosystem, can be studied and evaluated.

The presented methods allow the confinement of potential emission sources of bioaerosol and could aid in assessing the contribution of long-range-transport to the locally measured bioaerosol. This, in a simpler first approach can be used to calculate the relative contribution of different potential emission areas, such as the presented land cover maps. This allows the search for correlations between a specific land-type with the occurrence of specific PBAP. Moreover, by combining the residence time of air-masses with rasterized coverage data, such as the discussed tree species maps, more accurate predictions on the potential contribution of certain areas can be made. This allows the analysis of the atmospheric transport of specific species or groups of species.

However, the developed methods are limited to the extent that it is not yet possible to adequately describe or simulate real conditions in several respects. A clear improvement would result from a trajectory-level consideration of rain events, which effectively remove bioaerosol from the atmosphere. For this purpose, it would be necessary to investigate each air parcel at all time-points to assess whether it is located in a rain event. If an air parcel crosses a rain event, the trajectory could be removed and thus not used for the calculation of the total residency time.

In all respects, the temporal and/or spatial resolution of the PBAP data, meteorological data, therefore the accuracy of model predictions, and the rasterized potential emission source maps, are key for a successful application of the methods. Constantly improving methods of bioaerosol monitoring, more detailed computational models, better computational performances and more accurate geographical data, will lead to more accurate analyses of PBAP transport processes in the atmosphere. Predicting bioaerosol composition and concentrations with a high spatial and temporal resolution would provide an excellent basis for assessing the impacts of bioaerosol on humans, the ecosystems and the climate and ultimately allow the implementation of accurate early-warning systems to minimize negative potential impacts on, e.g., allergy sufferers or agriculture.

## **Acknowledgements**

Special thanks goes to T. Engel, A. Kunz, S. Brill, D. Kunkel, P. Hoor and T. Könemann for the helpful discussions and contributions to this work.

## **Conflict of interest**

The authors declare no conflict of interest.

IntechOpen

### **Author details**

Daniel A. Pickersgill<sup>1</sup>, Hartmut Müller<sup>2\*</sup> and Viviane R. Després<sup>3</sup>

1 Max-Planck-Institute for Chemistry, Mainz, Germany

2 Mainz University of Applied Sciences, Mainz, Germany

3 Johannes-Gutenberg University, Mainz, Germany

\*Address all correspondence to: hartmut.mueller@hs-mainz.de

### **IntechOpen**

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Pöschl U. Atmospheric aerosols: Composition, transformation, climate and health effects. *Angewandte Chemie (International Ed. in English)*. 2005;**44**(46):7520-7540. DOI: 10.1002/anie.200501122
- [2] Després Viviane R, Alex Huffman J, Burrows SM, Hoose C, Safatov AS, Buryak G, et al. Primary biological aerosol particles in the atmosphere. A review. *Tellus B*. 2012;**64**(0):296. DOI: 10.3402/tellusb.v64i0.15598
- [3] Jones AM, Harrison RM. The effects of meteorological factors on atmospheric bioaerosol concentrations—A review. *The Science of the Total Environment*. 2004;**326**(1-3):151-180. DOI: 10.1016/j.scitotenv.2003.11.021
- [4] Kellogg CA, Griffin DW. Aerobiology and the global transport of desert dust. *Trends in Ecology & Evolution*. 2006;**21**(11):638-644. DOI: 10.1016/j.tree.2006.07.004
- [5] Morris CE, Sands DC, Bardin M, Jaenicke R, Vogel B, Leyronas C, et al. Microbiology and atmospheric processes. Research challenges concerning the impact of airborne micro-organisms on the atmosphere and climate. *Biogeosciences*. 2011;**8**(1):17-25. DOI: 10.5194/bg-8-17-2011
- [6] Müller-Germann I, Vogel B, Vogel H, Pauling A, Fröhlich-Nowoisky J, Pöschl U, et al. Quantitative DNA analyses for airborne birch pollen. *PLoS One*. 2015;**10**(10):e0140949. DOI: 10.1371/journal.pone.0140949
- [7] Fröhlich-Nowoisky J, Pickersgill DA, Despres VR, Pöschl U. High diversity of fungi in air particulate matter. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;**106**(31):12814-12819. DOI: 10.1073/pnas.0811003106
- [8] Draxler RR, Rolph GD. HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) NOAA Air Resources Laboratory, Silver Spring. 2003. Available from: <http://www.arl.noaa.gov/ready/hysplit4.html>, Zugriff 10/16
- [9] Stein AF, Draxler RR, Rolph GD, Stunder BJB, Cohen MD, Ngan F. NOAA's HYSPLIT atmospheric transport and dispersion Modeling system. *Bulletin of the American Meteorological Society*. 2015;**96**(12):2059-2077. DOI: 10.1175/BAMS-D-14-00110.1
- [10] Stohl A, Forster C, Frank A, Seibert P, Wotawa G. Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2. *Atmospheric Chemistry and Physics*. 2005;**5**:2461-2474
- [11] Rew R, Davis G, Emmerson S, Davies H, Hartnett E, Heimbigner D, et al. What is NetCDF? 2016. Available from: <http://www.unidata.ucar.edu/software/netcdf/docs/index.html>, Zugriff: 10/16
- [12] Consortium for Small-scale Modelling. General Description of the COSMO-Model. 2011. Available from: <http://www.cosmo-model.org/content/model/general/default.htm>, Zugriff 10/16
- [13] Brus DJ et al. Statistical mapping of tree species over Europe. *European Journal of Forest Research*. 2012;**131**(1):145-157
- [14] CORINE Land Cover. Corine Land Cover 2006 raster data. 2006. Available from: <http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3>
- [15] Environmental Systems Research Institute, Inc. What is ArcPy? 2016.

Available from: <http://pro.arcgis.com/en/pro-app/arcpy/get-started/what-is-arcpy-.htm>, Zugriff 10/16

[16] Paukstadt A. Extraktion meteorologischer Daten aus NetCDF für räumliche Analysen. Geoinformatik und Vermessung, Hochschule Mainz, master thesis; 2015. 110 pages. unpublished

[17] Stull R. An Introduction to Boundary Layer Meteorology. Alphen aan den Rijn, Netherlands: Kluwer Academic Publishers; 1988

[18] Seinfeld, Pandis. Atmospheric Chemistry and Physics. Hoboken, New Jersey, United States: Wiley Interscience; 2006

IntechOpen