Science of the Total Environment 900 (2023) 165800

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Designing an automatic pollen monitoring network for direct usage of observations to reconstruct the concentration fields

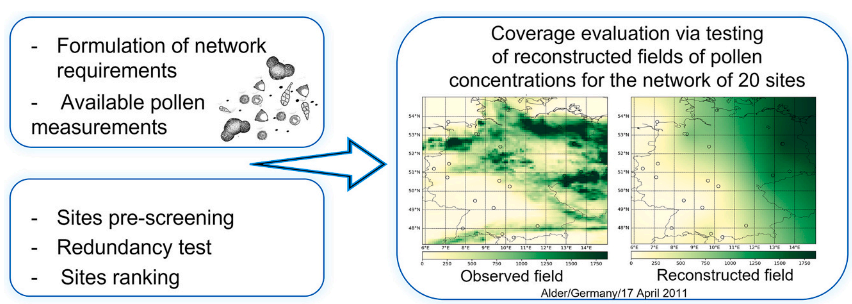
Mikhail Sofiev [a,*], Jeroen Buters [b], Fiona Tummon [c], Yalda Fatahi [a], Olga Sozinova [d], Beverley Adams-Groom [e], Karl Christian Bergmann [f,g], Åslög Dahl [h], Regula Gehrig [c], Stefan Gilge [i], Andrea Kofol Seliger [j], Rostislav Kouznetsov [a], Gian Lieberherr [c], David O'Connor [k], Jose Oteros [l,m], Julia Palamarchuk [a], Helena Ribeiro [n], Barbora Werchan [o], Matthias Werchan [o], Bernard Clot [c]

[a] Finnish Meteorological Institute, Erik Palmenin Aukio 1, 00560 Helsinki, Finland
[b] Center of Allergy & Environment (ZAUM), Member of the German Center for Lung Research (DZL), Technical University and Helmholtz Center, Munich, Germany
[c] Federal Office of Meteorology and Climatology MeteoSwiss, Chemin de l'Aérologie 1, 1530 Payerne, Switzerland
[d] Faculty of Geography and Earth Sciences, University of Latvia, Rainis bvld 19, Riga LV-1586, Latvia
[e] University of Worcester, UK
[f] Institute of Allergology, Charité – Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
[g] Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Allergology and Immunology, Berlin, Germany
[h] Department of Biology and Environmental Sciences, University of Gothenburg, Box 461, S-405 30 Gothenburg, Sweden
[i] Deutscher Wetterdienst, Berlin, Germany
[j] National Laboratory of Health, Environment and Food, Slovenia
[k] School of Chemical Sciences, Dublin City University, Ireland
[l] Department of Botany, Ecology and Plant Physiology, Agrifood Campus of International Excellence CeiA3, University of Cordoba, Rabanales Campus, Celestino Mutis Building, E-14071 Córdoba, Spain
[m] Andalusian Inter-University Institute for Earth System IISTA, University of Cordoba, Spain
[n] Faculty of Sciences, University of Porto and Earth Sciences Institute (ICT), Pole of the Faculty of Sciences, University of Porto, Portugal
[o] German Pollen Information Service Foundation (PID), Berlin, Germany

HIGHLIGHTS

- A problem of a network design for automatic pollen atmospheric monitoring in Europe is considered
- The paper analyses the features of the network oriented to a direct usage of observations for downstream services
- Factors influencing locations and the number of monitoring sites are considered from theoretical and practical standpoints
- The existing pollen network in Germany is assessed and a new network for automatic monitors is constructed on its basis

GRAPHICAL ABSTRACT



* Corresponding author.
E-mail address: Mikhail.sofiev@fmi.fi (M. Sofiev).

https://doi.org/10.1016/j.scitotenv.2023.165800
Received 4 January 2023; Received in revised form 1 June 2023; Accepted 24 July 2023
Available online 17 August 2023
0048-9697/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

ARTICLE INFO

Guest Editor: Pavlos Kassomenos

Keywords:
Automatic pollen monitoring
Network design
Network analysis
Station representativeness

ABSTRACT

We consider several approaches to a design of a regional-to-continent-scale automatic pollen monitoring network in Europe. Practical challenges related to the arrangement of such a network limit the range of possible solutions. A hierarchical network is discussed, highlighting the necessity of a few reference sites that follow an extended observations protocol and have corresponding capabilities.

Several theoretically rigorous approaches to a network design have been developed so far. However, before starting the process, a network purpose, a criterion of its performance, and a concept of the data usage should be formalized. For atmospheric composition monitoring, developments follow one of the two concepts: a network for direct representation of concentration fields and a network for model-based data assimilation, inverse problem solution, and forecasting. The current paper demonstrates the first approach, whereas the inverse problems are considered in a follow-up paper.

We discuss the approaches for the network design from theoretical and practical standpoints, formulate criteria for the network optimality, and consider practical constraints for an automatic pollen network. An application of the methodology is demonstrated for a prominent example of Germany's pollen monitoring network. The multi-step method includes (i) the network representativeness and (ii) redundancy evaluation followed by (iii) fidelity evaluation and improvement using synthetic data.

1. Introduction

Airborne pollen and fungal spores are routinely measured across the world for a variety of applications, including allergy diagnosis and treatment (Amato et al., 1998; Cecchi et al., 2010; D'Amato et al., 2007; de Weger et al., 2013; Sofiev and Bergmann, 2013), crop forecasting and plant disease control (Ben Dhiab et al., 2017; Rodríguez-Rajo et al., 2010), as well as biodiversity monitoring and understanding the impacts of climate change (Chmielewski and Rötzer, 2002; Galán et al., 2016, 2005; Newnham et al., 2013; Smith et al., 2014; Ziello et al., 2012). In Europe, most of the monitoring sites are part of regional or national networks run by a range of different organisations, from research or medical institutions to national weather services, most of whom also participate in the European Aeroallergen Network EAN (https://www.ean-net.org/ (Rybnicek and Jaeger, 2001; Berger et al., 2011; Buters et al., 2018)). To date, practically all networks use volumetric pollen and spore traps of Hirst-type (Hirst, 1952), for which a European standard has been developed (CEN 16868:2019).

The availability of instruments that can identify pollen automatically is leading to a paradigm shift in the bioaerosol monitoring across Europe. Regional and national automatic monitoring networks have been established in Bavaria, Germany (Oteros et al., 2020), Croatia and Serbia (Tešendić et al., 2020), and Switzerland (Crouzy et al., in prep). Apart from these networks, there are individual sites scattered around the continent, and their number is growing. To coordinate these activities, an AutoPollen Programme was established by European association of meteorological institutes EUMETNET in 2018, to facilitate the development of a real-time pollen monitoring network across Europe (Clot et al., 2020). It will be a hybrid network, incorporating monitoring sites of regional and national networks from across the continent and using various measurement systems. The programme aims to provide high-quality information on pollen concentrations in a standardised way. User-tailored products will be made freely available to the public and specific target audiences, such as medical practitioners, researchers, environmental agencies, agriculture, and forest industries.

An overview of the monitoring technologies suitable for (near-) real-time pollen monitoring has been provided by (Buters et al., 2021), whereas guidelines and protocols for carrying out these observations have been described by Tummon et al. (2022). A practical question related to the establishment of a Europe-wide automatic pollen network is the optimal distribution of the monitoring sites across the continent, as well as the number of the sites. Given the relatively high initial costs, the number of stations will likely be limited in the coming years, so the return on the investment will strongly depend on representativeness of these sites.

The objectives of this paper are: (i) to bring together the rigorous generic principles and practically used technique of network design; (ii) to suggest a practically applicable procedure for designing multi-purpose multi-taxa pollen monitoring networks at regional and continental levels; and (iii) to demonstrate its application for Germany (European scale will be addressed in a follow-up manuscript).

The paper is organized as follows. The next section outlines the general principles of the network design, developed theoretical and practical approaches, and their applications to pollen (eventually, spores and other bioaerosols) network. A formal problem statement and specific technologies for its solution are presented in Sections 4 and 5, respectively. As an example of the technology application, the current pollen monitoring network of Germany is analysed and configurations for an automated network are proposed in Section 6. The pros- and contras- of the suggested approach and its future challenges are discussed in Section 7.

2. Approaches to the network design

2.1. Network design from a theoretical standpoint

The network design problem is a specific case of the problem of optimal design of an experiment: to choose a setup and procedure for collecting data that will maximize the pre-selected criteria of statistical representativeness of the collected data for a given problem (Berliner et al., 1999). Specific procedures and workable algorithms are far from trivial and fully dependent on the problem at hand (Berliner et al., 1999; Bocquet et al., 2015). Arguably, the most-important step in formulating the network design problem is to formalize the purpose of the network and the corresponding representativeness criterion, which needs to be optimized by the site selection procedure.

For atmospheric composition monitoring, the network design has typically been formulated via one of two concepts (Bocquet et al., 2015; Koohkan et al., 2012; Lauvaux et al., 2012; Ramaker et al., 2003; Reza Koohkan and Bocquet, 2012).

2.1.1. Network for direct reconstruction of concentration fields from point measurements

The first concept maximizes the network representativeness for a direct evaluation of the concentrations over the target area (e.g., a country), primarily via spatial inter−/extrapolation to fill gaps between the stations. Such a network does not necessarily require models for the operations but can be combined with classical model data assimilation when measurements are used for adjusting the modelled concentrations in vicinity of the stations. Such a combination is able to create acceptable analysis fields but is inefficient in improving model forecasts (Elbern et al., 2007; Vira and Sofiev, 2012; Sofiev, 2019). A somewhat more efficient alternative is the model-measurement data fusion (Sofiev et al., 2017), which can be a suitable tool for interpolation of the measured data between the monitoring sites and for improving the model forecasts.

2.1.2. Network for applications together with atmospheric composition models

The second concept is based on source inversion problems: observations of the concentrations are assimilated into a model to refine the pollen emission estimates, both in time and space. Thereafter, concentrations over the region are computed by the model – and these model predictions are used as the input for downstream services. Such a network works together with models. This combination provides estimation of the pollen sources, consistent concentration patterns, and simultaneously improves the model forecasts (Elbern et al., 2007; Vira and Sofiev, 2012; Sofiev, 2019).

The current paper demonstrates the application of the first concept: construction of a pollen observational network for the direct usage of observations (a de facto current standard in aerobiology). A follow-up paper will concentrate on the inverse problems.

An example of a network designed for the direct observation-based representation of the concentration fields is the nuclear accidents alert system in France (Abida et al., 2008), where a "near-optimal" set of stations was pre-selected from 507 possible locations using the Simulated Annealing statistical method (Kirkpatrick et al., 1983). The evaluation of possible networks was performed via spatial interpolations of the observations, i.e., the network was built so that the concentrations anywhere in France could be restored using observations alone via interpolation between the stations. With the simple interpolation instead of dispersion modelling, the evaluation of fidelity of each network configuration was computationally cheap, whereas the Simulated Annealing method limited the number of configurations to test. Apart from a practical example, the study of Abida et al. (2008) delivered a warning: networks for direct usage of observations must be dense.

All published studies highlight that a rigorous solution of any real-life network design problem is not feasible from a computational standpoint.

The most popular simplification has been a pre-selection of site locations followed by an evaluation of the changes in fidelity when adding or removing sites (Abida et al., 2008; Bocquet et al., 2015; Lauvaux et al., 2012). But even then, the problem remains demanding: the number of stations and their locations are the parameters to identify from combinatorics, and each combination needs to be assessed. Therefore, the number of inverse problems to solve is of the order of magnitude of $\binom{N}{n}$, where N is the total number of possible station locations (e.g., 507 for France alone in the Abida et al., 2008 work), and n is the number of stations that can be distributed, let say, 100. The Onion Peeling method (Kirkpatrick et al., 1983) is one of the ways to limit the computations. Other methods suggested in the literature are based on multiple-objective planning analysis: (i) statistical methods - spatial correlation coefficients (Elkamel et al., 2008), PCA and cluster analysis (Wang et al., 2018), stepwise genetic algorithms (Hao and Xie, 2018; Li et al., 2019); (ii) holistic approach - Fuzzy Analytic Hierarchy process for multi-purpose tasks (Mofarrah and Tahir, 2011; Mofarrah and Husain, 2010), heuristic optimization algorithms (Elkamel et al., 2008), scale-related regionalized optimization (Abida et al., 2008; Lauvaux et al., 2012), etc.

2.2. Pollen network design from a practical standpoint

The problem of designing an optimal pollen-monitoring network is affected by additional, and often not formalized factors. They originate from the differences between applications for pollen monitoring: medical and environmental users are likely to be interested in different species and locations. Technical limitations, logistical feasibility, basic infrastructure needs, funding schemes, and administrative concerns, etc. put additional constraints on possible designs.

2.2.1. A European network or a set of national networks in Europe?

Today, automatic pollen monitoring networks across Europe are being constructed at the regional or national level, wherever financial resources and political will make it possible (Chappuis et al., 2020; Crouzy et al., 2016; Oteros et al., 2020; Šaulienė et al., 2019; Tešendić et al., 2020). Due to very diverse internal procedures and conditions, as well as different devices selected for operations, only a limited coordination between these projects has been achieved. The current progress has been largely due to harmonization and communication efforts within the AutoPollen framework (see Annex). Therefore, in addition to designing and planning a European-wide optimal network, it is necessary to consider regional and national scales, where a comparatively rigorous approach has been developed and used, for example, in the design of the ePIN network in Bavaria, Germany (Oteros et al., 2019) or in the reevaluation of the design of the SwissPollen Network in Swizerland (Lieberherr et al. in prep).

2.2.2. Stakeholder-driven constraints

A pollen-monitoring network designed for medical applications should provide the best data for as large fraction of the population as possible. Since interpolations, inverse problems, and models have their uncertainties, locating the stations in or near large cities is a tempting way to increase the accuracy and obtain user-relevant (i.e., for clinical needs) information directly from the observations. A limitation is the representativeness of individual stations: it falls fast with increasing distance, i.e., each city would have to have at least one station. But even this may not be enough: intra-city differences are high, driven by park zones, microclimate, configuration of streets, etc. (Werchan et al., 2018). This problem is also related to the temporal resolution of the data (section 7.3).

Pollen emission from plants describes vegetation phenology better than the airborne concentrations do. Therefore, the network constructed for the best emission-inversion criterion would be more suitable for the environmental, forestry and agriculture stakeholders. Sub-urban and rural sites usually have better spatial representativeness than the city- or street-level locations, but they also face this problem: local relief, proximity to water bodies, forests, or agricultural areas also create inhomogeneity in the pollen distribution patterns. The spatial scale of variability is much wider than in cities, but so are the areas to cover. Therefore, these applications will also require the network density corresponding to the needed spatial and temporal resolution.

Many applications, such as monitoring the environment's state and the population's changing exposure to allergenic pollen in a changing climate, require multi-decadal assessments. The EAN network has been generating a rich set of long-term observations (Fig. 1), which should be maintained with the new network, despite these sites might be sub-optimal for other tasks.

These considerations bring about three major reductions in the list of potential locations of the sites:

- all existing pollen monitoring sites with long time series of high quality should be included in the automatic network; "long" here means at least three complete solar cycles, i.e. 33 years (Gray et al., 2010; Nagovitsyn and Kuleshova, 2012; WMO, 2017), Fig. 1.
- Capital cities or/and major cities, including their suburbs (e.g., over one million of inhabitants) should be considered as priority locations for additional stations after the long-time series sites are counted.
- infrastructure availability and logistical feasibility (see Buters et al., 2021 for requirements to the station locations) makes existing air quality or meteorological monitoring sites attractive for setting up the pollen monitoring. However, requirements of pollen monitors can differ from those of other applications.

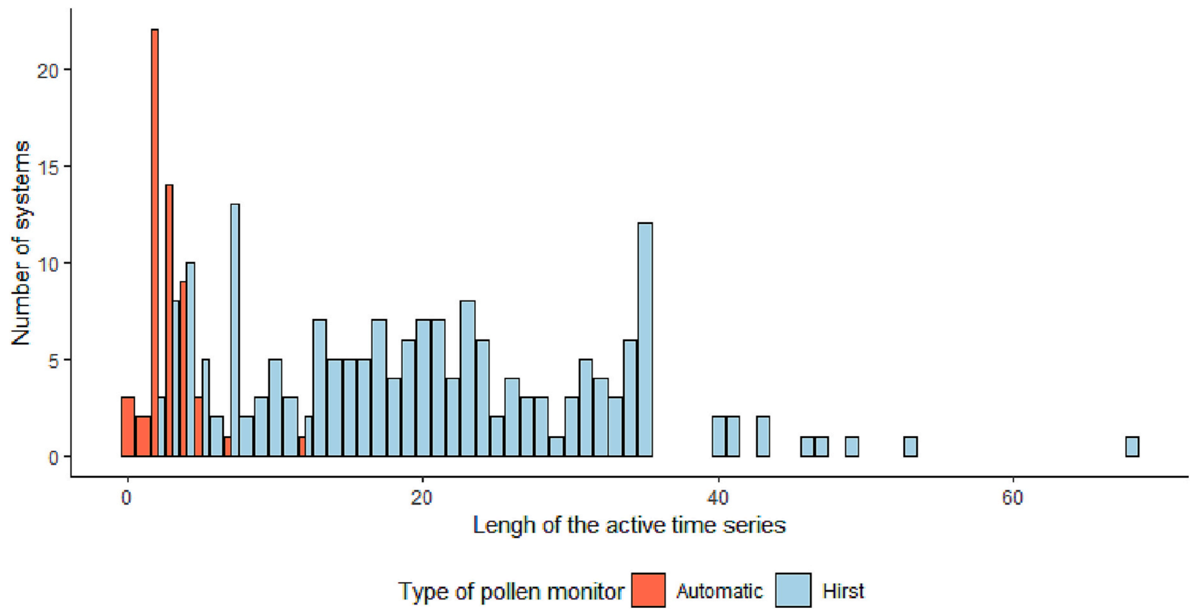


Fig. 1. Histogram of the length (years) of the time series from the European pollen and spores observations. Blue bars present European Aeroallergen Network (EAN) data with manual Hirst-type traps (see more from the EAN Web portal https://www.ean-net.org/en.html). Red bars show the length of timeseries of automatic networks. Totally 239 sites, which (i) were operational in 2022, (ii) have known starting date. Source: World pollen map database (Buters et al., 2018), the new URL: https://www.zaum-online.de/pollen/pollen-monitoring-map-of-the-world.html visited 29.12.2022.

The above requirements can be already overwhelming with regard to the number of sites, thus leaving no devices for other locations. In general, networks for direct usage of the observations can utilise as many stations as one can afford (Abida et al., 2008). The denser the network the more details it can see, and the higher its temporal and spatial resolution can be. The limitations are purely financial. Therefore, below we consider the total number of the sites as an external parameter and concentrate on optimizing the locations of these sites.

Besides selecting the geographically optimal location (macro scale), the local environment at the micro scale is equally important (Galán et al., 2014; Oteros et al., 2019). For instance, a roof must satisfy a number of requirements: height of the station (10–15 m above ground, (Rojo et al., 2019)), no air flow obstructing objects in the near surrounding, position of the instrument inlet (at least 1.5 m above the roof) and minimal distance from the edges, are the factors that may disqualify an otherwise optimal location. These requirements also depend on the purpose of the monitoring. For example, stations established for agricultural applications could sample particles directly at the crop level. In urban conditions, a network reflecting the variability of pollen distribution in a city may also require street-level locations (Hjort et al., 2015; Werchan et al., 2018).

2.2.3. Urban bias of European aeroallergen network

Over 80 % of the 639 EAN stations (about half of which are currently active) have been located in cities with population density exceeding 100 inhabitants per km^2 (Fig. 2). Only 29 sites represent regions with a population density of fewer than 10–50 people per km^2, and only about 15 of the rural stations are currently operational. This urban/rural disbalance has two reasons: (i) the Hirst trap needs a regular, usually weekly, change of the sticky tape and other maintenance, (ii) the urban pollen-sensitive population is the main user of the data. Consequently, the data have an urban bias, sometimes with a disproportionally large representation of pollen from ornamental plants, high anthropogenic stress on local flora, urban heat island effects (i.e., higher urban temperature than in the surrounding countryside), polluted and often artificial soil, etc.

Since the contribution of local vegetation to the air concentrations is significant, such bias is likely to cause substantial alterations in timing of the pollen season, the pollen features and allergenicity (Behrendt and Becker, 2001; Bosch-Cano et al., 2011; Katz et al., 2019).

2.3. Hierarchical pollen network

Relatively high cost of automatic devices, their continuous development, and the necessity to homogenise their data with historical manual time series suggest that a hierarchical monitoring network can be an optimal solution.

A hierarchical network would include a few **reference sites**, with extensive infrastructure to host high-end observations, experimental campaigns, tests, calibrations, and intercomparison campaigns. These sites should be representative for various conditions to ensure that the tested monitors can recognise a wide range of different pollen types and operate under diverse climatic conditions.

Baseline stations equipped with automatic monitors would constitute the bulk of the network. They will also include high-quality stations with long-term measurements of Hirst type. Those will be used to establish transfer functions between the Hirst-type measurements and the new time series from the automated monitors.

There is an option of having many **low-cost sites** fitted with cheaper sensors, e.g., for measuring total-pollen, if a use case for such data can be found or some specific easily recognizable taxa will be of a particular interest. The “Hanakosan” network in Japan is based on such low-cost sensors (Kawashima et al., 2017) exploiting the feature of the Japanese pollen season with one easily-identifiable pollen type seemingly dominating the allergy pattern. Due to complexity of the European pollen composition, such sensors, while tested (Tummon et al., 2021), were not adapted over the continent and this study does not consider this option.

2.3.1. Reference sites and measurements

A limited number of well-selected reference sites will aim to provide the best estimate of pollen (and, eventually, other biological aerosols) abundance and its associated uncertainty. Apart from routine monitoring and various campaigns, these sites will provide metrological references, in particular, ensuring traceability of the new measurements. Traceability, the key metrology concept behind the reference

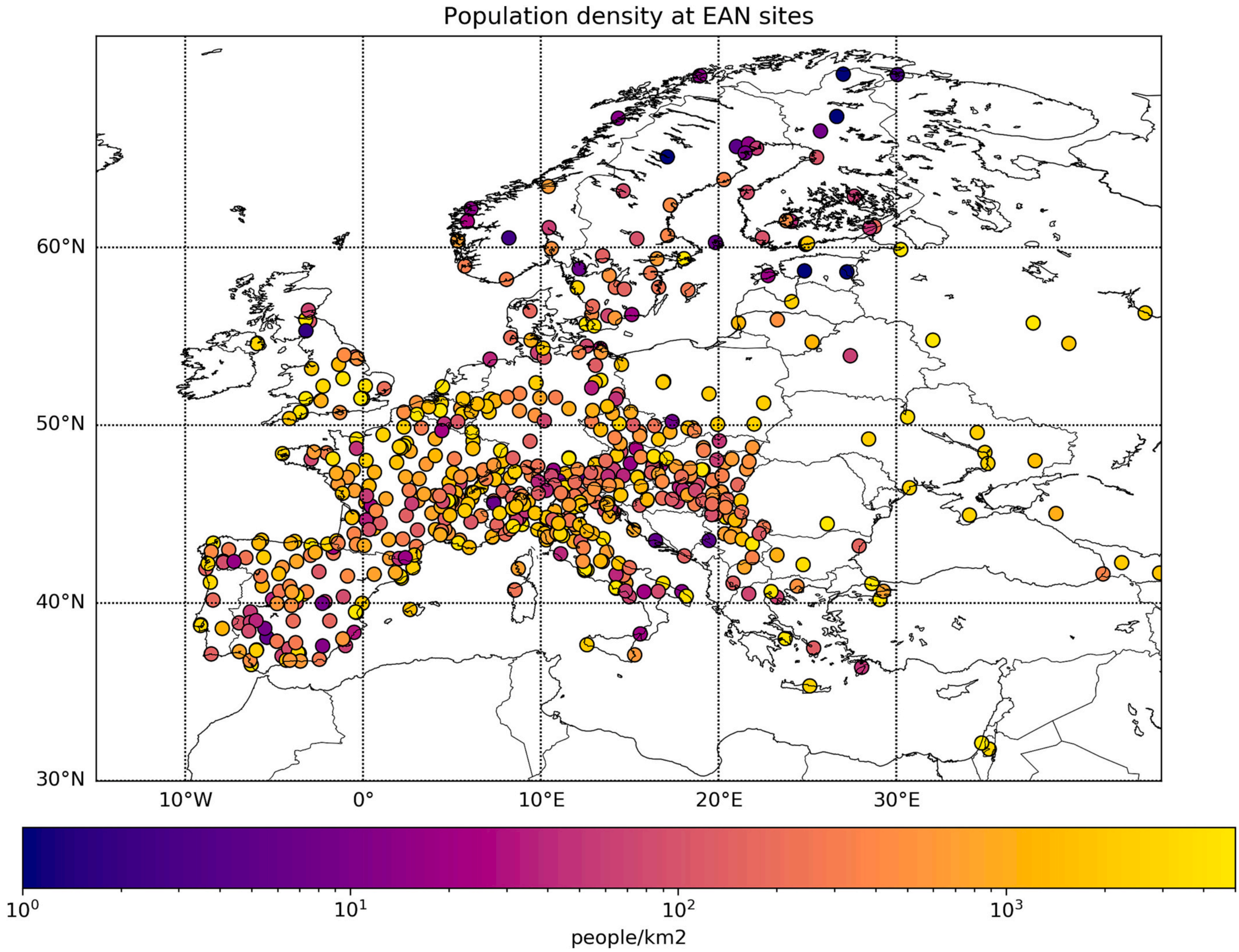


Fig. 2. Population density (2.5″ grid, ~3 × 4.6 km resolution) at the locations of the 639 active and historical sites of the European Aeroallergen Network. The middle of the colour bar corresponds to a city with ~20,000 inhabitants. Only violet dots characterize the rural conditions on the map. Source: Gridded Population of the World, version 4 (GPWv4, (CIESIN, 2016)).

measurements, refers to the possibility of tracing a measurement value and its uncertainty back to a reference value (typically made using a defined measurement procedure that involves national or international standards) through a documented chain of calibrations quantifying the components of the measurement error. Establishing these components will be one of the tasks for the reference sites.

Given the challenges related to observing bioaerosols, a multitude of instruments at the reference sites will allow for establishing a multi-instrument consensus about the actual concentrations of biogenic particles in the air.

2.3.2. Criteria for reference station selection

Beyond meeting the requirements for standard AutoPollen sites (Tummon et al., 2022), the reference stations should also meet as many of the following requirements as possible.

2.3.2.1. Coverage of main European bioclimatic zones. The AutoPollen network needs to cover a variety of bioclimatic zones and biogeographical regions (ETC/BD, 2017), so that the measurement systems can be tested over the whole range of different environments and pollen taxa.

2.3.2.2. Well-identifiable pollen season. The pollen season needs to be well-identifiable, with concentrations high enough for the relevant taxa in the specific bioclimatic zone (Pfaar et al., 2016). It is also preferable, albeit not mandatory, to have multi-annual records from Hirst-type traps available for each reference station.

2.3.2.3. Multiple automatic instruments at the site. At least one automatic monitor and a Hirst-type trap should be installed at each reference site. Both instruments should be well-calibrated and regularly checked. Furthermore, a weather station should be located at or close to the site, providing observations of temperature, precipitation, wind direction and speed, solar radiation, as well as humidity with at least an hourly resolution.

2.3.2.4. Instrument intercomparison. Space and infrastructure should be available to install additional instruments at the same location for the intercomparison experiments. Knowledgeable personnel should also be available to provide technical support on-site.

2.3.2.5. Managing instrumentation changes. Changes in the instrumentation, the instrument location, the surrounding micro-environment, the operating procedures, and the data processing algorithms can and will occur during the lifetime of the network. Certain changes are punctual (e.g., algorithm changes), whereas others are more gradual (e.g., urbanization around a measurement site). Unnecessary changes should be avoided, whereas inevitable ones should be monitored, documented and analysed to ensure the integrity of long-term records (Aguilar et al., 2003; WMO, 2020).

2.3.2.6. Complementary monitoring. The reference sites should be located close to an automated meteorological station and preferably also close to an air quality monitoring station, or at least close to a PM_{10} monitoring station (e.g., World Meteorological Organization Global Atmospheric Watch, WMO GAW, https://community.wmo.int/activity-areas/gaw, visited 20.12.2022, or Co-operative programme for monitoring and evaluation of the long-range transmission of air pollutants in Europe, EMEP, http://emep.int, visited 20.12.2022). Information from these stations is important to better understand the influence of weather or high non-pollen aerosol concentrations on the reference measurements.

2.3.2.7. Reacting to environmental changes. Selecting the reference sites, one has to account for on-going climate change, reforestation processes in Europe, degradation of some regional environments, biodiversity trends, changes in agriculture practices and planted species, etc. The reference sites should be placed in locations that allow them to notice these changes at early stages, so that the main network can be timely adjusted accounting for the new sources and species.

2.3.3. Selection of bioclimatic zones

The reference sites are expensive, therefore establishing them in each of the 11 European bioclimatic zones (ETC/BD, 2017) will hardly be feasible. Luckily, comparing these zones with the maps of pollen occurrence (Skjøth et al., 2013), it becomes apparent that several zones have similar pollen profiles, i.e. can be represented with one site. At the same time, in some bioclimatic zones one can have regions that differ substantially in pollen timing and composition, i.e., a “similar-pollen zone” might be also smaller than a bioclimatic zone.

At the top level, Atlantic/Continental/Boreal/Alpine bioclimatic zones can be differentiated from Mediterranean. These two large clusters strongly differ in pollen and fungal spores composition and timing (Sofiev and Bergmann, 2013). At the next level, different abundance of individual pollen taxa in Atlantic/Continental/Boreal/Alpine cluster can be taken into account. Notably, there is less diversity in the Boreal and Alpine regions, which also warm faster than the rest of Europe, with potentially large impact on regional flora. A specific and important invasive plant is *Ambrosia*, which has the main distribution areas in the Pannonian zone in South-Eastern Europe and in steppe zone – those might require a dedicated high-end monitoring station.

Maps of tree habitats in Europe can further help to locate the most important pollinating trees. Examples include the European Atlas of forest tree species (https://forest.jrc.ec.europa.eu/en/european-atlas, visited 20.12.2022), the European Forest Genetic Resources Programme (http://www.euforgen.org/species, visited 20.12.2022), or the Euro-Med PlantBase, a resource to visualise the different plant distributions across the Mediterranean (http://ww2.bgbm.org/EuroPlusMed/query.asp, visited 20.12.2022).

2.3.4. Selection of reference site locations

Representativity of any monitoring site is always important but for reference sites it is particularly vital to have its observations representative for at least regional scale (a few kilometres for hourly, tens of kilometres for daily values). The representativity of the site can be affected by numerous factors, including local topography, proximity of strong sources of pollen, large water bodies within few km distance (because of local circulations, such as sea breeze), etc. The roles of such factors should be estimated, documented, and minimized.

Summarising the above considerations, we propose as a program-minimum to establish at least three reference sites, at least one in each of the three greater bioclimatic zones. At least one reference site should be in the Atlantic/Continental/Boreal region, in a location where preferably all the relevant pollen and fungal spore taxa of this region are present in sufficiently large amounts. A second reference site should be in the Mediterranean area, covering the main Mediterranean taxa. For *Ambrosia* and *Artemisia*, a third reference site should be selected in the Pannonian region. Consideration should be given to sites representing the southern and northern conditions as well.

2.4. Practical examples of regional baseline networks

Locations of baseline stations can be based on various selection procedures. As a practical example, the ePIN network followed a two-step approach, which established a dense network over Bavaria and subsequently used cluster analysis to group the redundant sites into clusters and obtain the new site locations (Oteros et al., 2019).

The SwissPollen network was redesigned based on the local pollen load at the existing Swiss sites. A measurement site was considered representative for all points in space that show similar pollen load for the pollen taxa in the COSMO-ART dispersion model. With this, the spatial coverage of all individual measurement points was evaluated. The approach was developed during the automation of the Swiss pollen monitoring network: the existing network was evaluated using COSMO-ART predictions, and then, in an iterative process, the identified weaknesses in population and territorial coverage were addressed (Lieberherr et al. in prep.).

3. Basic equations and tools for designing the network for direct usage of observations

3.1. Target parameter for optimization: concentration field

Observed time series of airborne pollen concentrations $C_o(x_n, y_n, t)$ across a network S_N consisting of N stations $\{s_n(x_n, y_n), n = 1..N\}$ will be used to directly represent concentration distributions over the considered area $C(x, y, t)$. Here, n is the index of the observation site, x and y are horizontal coordinates (e.g., longitude and latitude), and t is time.

Restoring concentration fields directly from observations without a dispersion model implies spatial inter- and extrapolations performed for each time moment t. Concentration $C(x, y, t)$ at an arbitrary point (x, y) and time moment t is then a function $F(\ C_o(x_n, y_n, t)\ )$ of values observed at the stations at the same time moment. The specific form of the function is largely up to a developer's choice. It can be, for instance, taken as a linear combination of values observed at the stations:

$$C_{int}(x, y, t) = F(\ C_o(x_n, y_n, t)\ ) = \sum_{n=1}^{N} \alpha_n(x, y)\, C_o(x_n, y_n, t) \tag{1}$$

Here coefficients α_n are represented as a map of weights describing the assumed relation between the concentration at each station and at the point of interest (x, y). There are many ways to determine α_n, which can be deterministic (e.g., interpolation splines or radial-based functions) or geostatistical (e.g., kriging) (Böhner and Bechtel, 2018). The analysis below is based on the radial-based function RBF (Broomhead and Lowe, 1988), which have numerous practical applications and for which an efficient implementation exists in the SciPy Interpolate library of Python 3. Another possible option could be bicubic 2D interpolation B-spline ((Prautzsch et al., 2002), see also “spline interpolation” in the online Encyclopaedia of Mathematics of European Mathematical Society, https://encyclopediaofmath.org/wiki/Main_Page).

3.2. Quantifying the network quality

The ability of a network to reproduce the concentration field over the entire area of interest can be quantified via RMSE (Square Root of the Mean Square Error) between the "true" concentration field $C(x,y,t)$ and $C_{int}(x,y,t)$ reconstructed via interpolation of the network observations. Time-resolving RMSE, *J(t)*, is simply:

$$J^2(t) = \frac{1}{A}\iint_{\Omega} \left(C(x,y,t) - C_{int}(x,y,t)\right)^2 dx\, dy \tag{2}$$

Here Ω is the region of interest, 2D, with the area A.

To obtain the final single value quantifying the network skills, integration over time (e.g., a full season) is needed, but it can be somewhat misleading because the process is not stationary. In many cases, one may be interested in a maximum error during the season of the considered taxon rather than in a year-round average:

$$J_{mean} = \frac{1}{T}\int_T J(t)dt, J_{max} = \max_T(J(t)) \tag{3}$$

The Eq. (2) requires knowledge of the "true" concentration field C, which can be estimated via two approaches.

The first approach used for the ePIN network development included one year of operations of a dense network with station separations <30 km homogeneously spread over Bavaria (Oteros et al., 2019). Assuming that this network is representative for the "true" concentration field, the integral in (2) turns into a sum over this dense network, J_{DN}:

$$J_{DN}^2(t) = \frac{1}{M_{DN}}\sum_{i=1}^{M_{DN}} \left(C_{DN}(x_{DN}(i), y_{DN}(i), t) - C_{int}(x_{DN}(i), y_{DN}(i), t)\right)^2 \tag{4}$$

Here, $x_{DN}(i), y_{DN}(i)$ are coordinates of the sites of the dense network, M_{DN} is the number of stations constituting it, and C_{DN} is the concentration observed at its stations.

The dense-network approach, being tempting because of usage of actual observations, suffers from two drawbacks: (i) high cost and long time required for construction of the dense network and its operation over an extended period of time (one year, as in the case of ePIN, is an absolute minimum), (ii) an assumption that this dense network is, indeed, representative and the switch from the integral (2) over the whole region to the sum over this network (4) does not add large uncertainty.

The second approach is to construct a numerical experiment using a pollen dispersion model. With such a model, one can simulate the pollen distribution C_{mdl} over the target region, thus creating a synthetic "true" distribution pattern. This pattern is sampled at the station locations generating synthetic observations, which will be the model predictions at the corresponding grid cells and for the corresponding times. The problem then becomes discrete, and the network quality criterion J_d can be written as follows:

$$J_d^2(k) = \frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y} \left(C_{mdl}(i,j,k) - C_{int}(i,j,k)\right)^2 \tag{5}$$

$$J_{d\_mean} = \frac{1}{K}\sum_{k=1}^{K} J_d(k), J_{d\_max} = \max_K(J_d(k))$$

Here i,j are grid indices, k is time index, $n_x n_y$ are horizontal grid sizes, and K is the number of times in the model output. Note that error summation now goes over the whole domain, thus being a discrete representation of the integral (2).

The J criteria are expressed in [pollen/m3] and need a reference scale to compare with. There may be several scales suitable for this purpose. Below, we shall use a root mean square model concentration. Setting the observations C_{int} to zero in (2)–(5) and adding a surrogate to the measurement uncertainty to avoid division by zero, one obtains normalizers W, W_{DN}, W_d for the above criteria:

(a) Norm for the general case (2) $$W^2(t) = D^2 + \frac{1}{A}\iint_{\Omega} \left(C(x,y,t)\right)^2 dx\, dy$$

(b) Norm for the dense network case (3) $W_{DN}^2(t)$

$$= D^2 + \frac{1}{M_{DN}}\sum_{i=1}^{M_{DN}} \left(C_{DN}(x_{DN}(i), y_{DN}(i), t)\right)^2 \tag{6}$$

(c) Norm for the discrete problem case (5) $W_d^2(k)$

$$= D^2 + \frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y} \left(C_{mdl}(i,j,k)\right)^2$$

Here D is measurement uncertainty, a constant, which becomes relevant for low concentrations when the second term approaches zero. The value of D should be close to the detection limit, so we take 10 pollen m^{-3} as its rough estimate.

Normalising J, J_{DN}, J_d with W, W_{DN}, W_d, respectively, one obtains convenient dimensionless parameters J/W, J_{DN}/W_{DN}, J_d/W_d. They are equal to 0 for a perfect observational network and reach 1 when observations become essentially useless (the error is as large as in case of no observations at all). It can also exceed 1 for the cases when extrapolation of observations produces so large errors that zero-concentration hypothesis becomes preferable.

The normalized scores of the eqs. (4)–(6) will be used in the analysis below, whereas the simulated pollen concentrations will be taken from the SILAM model.

3.3. SILAM atmospheric composition model

Model simulations presented in this paper have been performed with the System for Integrated modeLling of Atmospheric coMposition (SILAM; http://silam.fmi.fi, visited 25.10.2022). It is an offline global-to-local-scale chemistry transport model developed for evaluating atmospheric composition and air quality (Sofiev et al., 2015b), emergency decision support applications (Sofiev et al., 2006), source inversion problems (Sofiev, 2019; Vira and Sofiev, 2012), and analysis of observations (Meinander et al., 2020; Tarasova et al., 2007). Among the variety of physical and chemical transformation modules of SILAM, this study uses the pollen source modules (Prank et al., 2013; Siljamo et al., 2012; Sofiev et al., 2012, 2015a, 2017) as well as the basic dispersion and deposition modules (Kouznetsov and Sofiev, 2012; Sofiev, 2002; Sofiev et al., 2010).

SILAM has been evaluated in a variety of regional and global studies showing robust performance (Brasseur et al., 2019; Huijnen et al., 2010; Kouznetsov et al., 2020; Petersen et al., 2019; Sofiev et al., 2015a, 2020; Xian et al., 2019).

The model currently includes 12 types of bioaerosols: pollen of alder, birch, grass, olive, 5 groups of mugwort species (their pollen is indistinguishable from Hirst-type slides but the flowering time and the distribution in Europe are drastically different), and ragweed, as well as hazel and aphids (small wind-transported insects) (Sofiev et al., 2012, 2015a, 2017; Prank et al., 2013; Sofiev, 2016, 2019; Siljamo et al., 2012; Ritenberga et al., 2017).

3.4. Footprint-based analysis

A comparatively inexpensive way to evaluate a network's coverage is based on footprint computations. They can be made through adjoint dispersion modelling, which is a part of variational data assimilation cycles but also widely used for analysis of observational campaigns (Vira and Sofiev, 2012; Bocquet et al., 2015; Meinander et al., 2020; Veriankaitė et al., 2010; Saarikoski et al., 2007). Formally speaking, adjoint computations go beyond the "observations-only" concept, but comparatively low costs and substantial added value make them useful even for this case.

A surface footprint of an observation is an area from where the sources will contribute to the observation if they emit pollen during the corresponding time. The footprint absolute value is proportional to the fraction of emitted amount reaching the station. Footprints, unlike backward trajectories, explicitly account for deposition processes, such as sedimentation and washout by precipitation. This method does not say whether the areas covered by the footprint did emit pollen at a specific time. However, no areas outside the footprint can affect the monitor (assuming a perfect model), which makes the footprint a handy tool for identifying regions not, or poorly, covered by a particular monitoring network.

Footprints directly relate to the fraction of air reaching the monitoring device. Formally, the convolution of emission ξ, and the footprint absolute value φ^* over space and time is equal to the mean concentration C at the monitor location during the period over which the observation is made:

$$\sum_{\tau}\sum_{I,J,K}\varphi^*(i_o,j_o,t_o,i,j,k,t)\ \xi(i,j,k,t) = C_o(i_o,j_o,t_o) \tag{7}$$

Here, $C_o(i_o,j_o,t_o)$ is the concentration observed at a given height near the surface at the site location (i_o,j_o) at a discrete time interval t_0. The location and timing of the observations are used as input for the adjoint calculations of the footprint.

In our analysis, the absolute values of the footprint are of little importance, the message comes from the difference of these values at different locations. The higher the footprint value is, the larger the contribution from the sources in the corresponding area to that specific observation. One can also consider a cut-off threshold for the footprint value, below which it is set to zero. It would remove areas with low contribution to the observations and reduce the total footprint area. For the cut footprint, the integral (7) will be lower than the observed value. One can require it to be, e.g., 90 % of the observed concentration, thus removing all regions contributing <10 % of the observations.

It should be stressed that footprints have no connection to the probability of the source to contribute to the observation, despite the opposite "silent assumption" is quite widespread. Indeed, as seen from (7), all sources covered by the non-zero footprint will contribute to the observation. It is only the strength of their contribution that is controlled by the footprint value.

4. Network design based on an existing proxy

Designing a network using an existing "proxy" network as a starting point reduces the problem of selection of the station locations to finding an optimal combination of the existing sites. Additionally, the time series of the proxy network provide useful information on its characteristics in the past. These additional pieces, albeit not rigorous from mathematical standpoint, allow for a simple pre-screening of potential solutions and can significantly simplify the computations. The proxy network is usually not dense, i.e., one cannot assume that it provides exhaustive information on the concentration field.

This section suggests pre-screening steps, which identify the main strengths and weaknesses of the proxy network and show the directions of its improvement. The screening addresses general features of the proxy network and can therefore be combined with any of the above network quality criteria.

4.1. Removing redundant sites in the existing network

The first pre-screening step assesses the existing sites, identifies and removes the redundant ones. To do this, one requires several years of observations (ideally, it should be one full solar cycle, 11 years, but even 2–3 years might be sufficient for the screening purposes) from all sites. Various methods can be applied to identify the stations reporting consistently similar values (redundant sites). This includes cluster analysis (Gehrig, 2019; Oteros et al., 2019; Rodríguez-Fernández et al., 2022), correlation analysis, etc.

A potential problem of this step is that it does not distinguish between informative sites delivering new information and outliers that appear to provide "unique" information for technical reasons or a non-representative site location. It is therefore important to manually identify and remove all outliers, specifically for each taxon, in the initial proxy network before starting the analysis.

Removal of the redundant stations retains only sites that each contributed with substantially different time series in the past. The network of non-redundant stations may suffer from a lack of coverage: (i) the initial proxy network itself may have insufficient coverage, (ii) removal of the redundant sites may leave holes, which will no longer be covered. The coverage should be verified and, most probably, improved by adding extra sites. Since the number of non-redundant stations depends on a similarity threshold applied in the clustering, it should be set so that the number of retained stations is smaller than the planned size of the final network.

4.2. Checking the coverage of the non-redundant network

The second pre-screening step is to assess the network coverage, e.g., by computing and summing-up cumulative season-long footprints for all sites. For pollen monitoring, two additional considerations simplify the problem. Pollen emission (unlike concentrations) at night is low, i.e., the area covered by a footprint at night is less relevant than that covered during day. One can apply a pollen release model (e.g., (Sofiev et al., 2012; Zink et al., 2013)) as an additional scaling factor or even use a fixed diurnal profile, e.g. the one shown in Fig. 3. Since only a rough estimate is needed, such a simple profile may be sufficient for screening purposes.

The second simplification is that the footprints are only computed for the flowering period and not an entire year, which decreases the computational costs and removes the impact of meteorological situations during extraneous seasons.

These simplifications are quite well grounded for pollen but should not be applied if a network also targets, e.g., fungal spores that can be emitted at night and over most of the year.

The computed footprints for all stations, when summed-up, should cover all potential source areas. If some areas are not covered, extra stations should be added. Their locations can be determined by running the pollen model in prediction (forward) mode using only the non-covered areas as sources. The simulated concentration fields will show the regions affected by these sources.

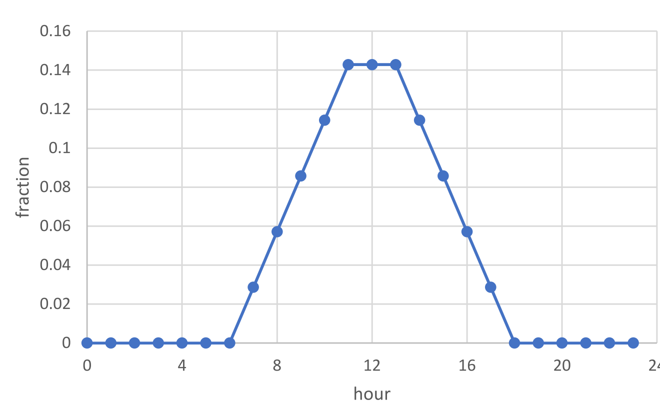


Fig. 3. A simplified synthetic diurnal cycle of relative pollen emission as a function of local solar time.

4.3. Finalising the proxy-based network for direct usage of observations

Having the screening steps completed, the final step in the proxy-based network design is a stepwise evaluation of the network quality criteria (6) while adding sites at places with insufficient coverage until the cost function reaches a pre-defined quality threshold or the number of sites reaches the maximum limit.

Choosing the form of the cost functions (6), one should keep in mind that the proxy network is, in most cases, much sparser than a would-be dense network for the region. For instance, the pollen network in Bavaria had 5 stations (two operational by the start of ePIN), the ePIN dense network had 24 locations, and the final one has 8 automatic sites. It is therefore recommended to use synthetic model-based approach for evaluation of the network skills, i.e., discrete grid-based cost function I_d (5) and the corresponding normalized criterion (6c).

Selection of the network quality threshold is not a rigorous step. The normalized cost function should certainly be <1 (the no-observations case) but further refinement is needed. As one of possible approaches, the L-curve formalism (Hansen, 1992) would identify the number of stations above which the error only weakly decreases with additional sites.

5. Example: constructing a national automatic pollen monitoring network for Germany based on an existing proxy

This section presents an example analysis of the German Hirst-type pollen monitoring network for the needs of automatic monitoring. It includes the pre-screening steps followed by the formal assessment of quality of the non-redundant network and analysis of means of improving its skills.

The pre-screening is performed using observations from years 2010–2013 for eight different pollen and spore taxa: *Alnus* (in tables and figures referred to as ALNU), *Alternaria* (ALTE), *Ambrosia* (AMBR), *Artemisia* (ARTE), *Betula* (BETU), *Cladosporium* (CLAD), *Corylus* (CORY), and *Poaceae* (POAC). These pollen types were selected following the recommendations of the COST Action EUPOL (Sofiev and Bergmann, 2013), with some types not included because of their low abundance in Germany (e.g., *Olea*).

5.1. Redundancy test

The redundancy test was started from ranking the sites with regard to their importance for health and climate-related applications. We built a metric that combines the population density around the site (within a radius of 35 km, a rough estimate of a size of a large city agglomerate) and the total length of available observations from 1974 to 2014. The sites were ranked according to each criterion, and the ranks for each site were summed up to obtain its final standing (see Table 1 for the ten most important stations, the entire table is shown in the Annex Table A.1). This approach and its thresholds have no rigorous justification, but they ensure that a station within a densely populated area is considered more important than a site located far from cities. Conversely, stations with very short datasets are considered less important than those with longer time series.

A correlation analysis was applied to daily pollen concentrations, separately for each year, with the data completeness of at least 20 % of the year (73 days, including days with zero concentration). This threshold is quite loose, but (i) pollen observations are practically never done throughout a full year, (ii) it allowed for incorporation of more stations in the analysis with little impact on the robustness of the results: one does not need, e.g., an entire grass season to note that the stations are (not) correlated, 2.5 months is enough for that. Sensitivity tests were carried out using thresholds of 10 % and 30 % data availability, with a similar outcome.

The Pearson correlation coefficient (*r*) was calculated for each pair of stations and pollen type to assess the similarity between the sites. A redundancy threshold of 0.85 was used since this value was previously found to be optimal when applying a similar analysis for the design of the ePIN network (Oteros et al., 2019). A minimum correlation coefficient of 0.85 was also determined as a threshold separating clusters of pollen stations for a mean of seven allergenic taxa of the Swiss pollen network (Gehrig, 2019). It also corresponds to the accuracy of the Hirst-type observations, with $r^2 \cong 0.7$ corresponding to ~30 % of uncertainty associated with the manual observations (Adamov et al., 2021; Oteros et al., 2017). As a result, sites with mutual correlations higher than 0.85 were considered to duplicate each other.

The stations were grouped into clusters around the most-important stations that correlated $r > 0.85$ with all other sites in the cluster. The cluster was then reduced, retaining only the highest-ranked station (Table 1 and Table A.1). The procedure was applied iteratively, i.e., only one cluster was reduced at a time, after which clustering was repeated.

For example, for ragweed in 2010, the highest-ranked site is DEMUNC (Table 1), which does not correlate with any other site in the network. In contrast, the station DEBOCH, at rank 2 (Table 1) correlates with several stations, including DEBONN($r = 0.89$), DEFLEN(0.90), DEMAGD(0.86), DEMUST(0.86), DENEST(0.89), DEPOTS(0.88), DEROST(0.91), DEVECH(0.87), and DEWANG(0.86). Therefore, these stations were all removed from the list, and the next round of pair-wise correlation clustering was performed. The loop was repeated until no station series correlated >0.85 to each other. The outcome of the procedure for 2010 is shown in Table A.2.

The clustering reduction procedure was performed independently for each taxon (8 taxa) and year (4 years). For each year, the lists of non-redundant sites for each of the eight taxa were compared and the stations ranked according to the number of lists where they were present. These ranks were averaged over each of the 4 years, ending up with 40 stations that were relevant for >2 taxa over the 2010–2013 period (Table 2, Fig. 4). Interestingly, only two sites were relevant for all 8 taxa for all 4 years, while 20 were important for 6 taxa. To evaluate the scores, a manual "common-sense-based" ranking was made selecting 20

Table 1
The 10 top-rank German stations sorted by the population density within a radius of ~35 km from each site and the length of available data from 1974 to 2014.

	dx=0.333333						
Site	Population count 2005	Year obs. 1970-2014	Days observed 1970-2014	Rank population	Rank years	Sum of ranks	Final rank
DEMUNC	1360090	22	6507	6	8	14	1
DEBOCH	1568150	21	4614	3	12	15	2
DEBERL	1872870	20	5040	2	20	22	3
DEHANN	386712	26	4320	28	3	31	4
DEERLA	688240	21	3607	18	14	32	5
DEDRES	446392	21	6105	25	10	35	6
DEFREI	352737	23	6888	30	5	35	7
DEGERL	1074500	19	2396	11	25	36	8
DEMOEN	685401	20	5165	19	19	38	9
DEDELM	224470	33	9743	39	1	40	10

Table 2
Mean ranking of the sites that provided data for all years in the period 2010–2013. The a priori ranking follows those of Table 1 and Table A.1.

Final site rank	Site code	Mean Nbr of taxa for which site is informative	A-priori rank: Table A.1	Final site rank	Site code	Mean Nbr of taxa for which site is informative	A-priori rank: Table A.1
1	DEMUNC	8	1	21	DEHAGE	5.75	12
2	DEDELM	8	10	22	DESOES	5.5	55
3	DENORD	7.25	53	23	DEKOEN	5.5	20
4	DEKIEL	7.25	44	24	DEHEID	5.5	16
5	DEBONN	7.25	13	25	DEGOET	5.5	24
6	DEOBER	7	58	26	DEROST	5.25	21
7	DEMOEN	7	9	27	DEGARZ	5.25	67
8	DEMANN	7	32	28	DEAUKR	5.25	47
9	DEBOCH	7	2	29	DEREIN	5	25
10	DECHEM	6.75	18	30	DETREU	4.75	66
11	DEWANG	6.5	31	31	DELIPP	4.75	15
12	DEMUST	6.5	46	32	DENEUS	4.5	70
13	DEDRES	6.5	6	33	DEVECH	4.25	52
14	DEWEST	6.25	36	34	DEGREI	4.25	50
15	DELOEW	6.25	17	35	DENEST	4	41
16	DEFULD	6.25	38	36	DEJENA	4	27
17	DEHANN	6	4	37	DEDREB	4	79
18	DEFREI	6	7	38	DEFLEN	3.75	40
19	DEBERL	6	3	39	DEPRER	3.25	72
20	DEBER1	6	11	40	DEPOTS	2.25	14

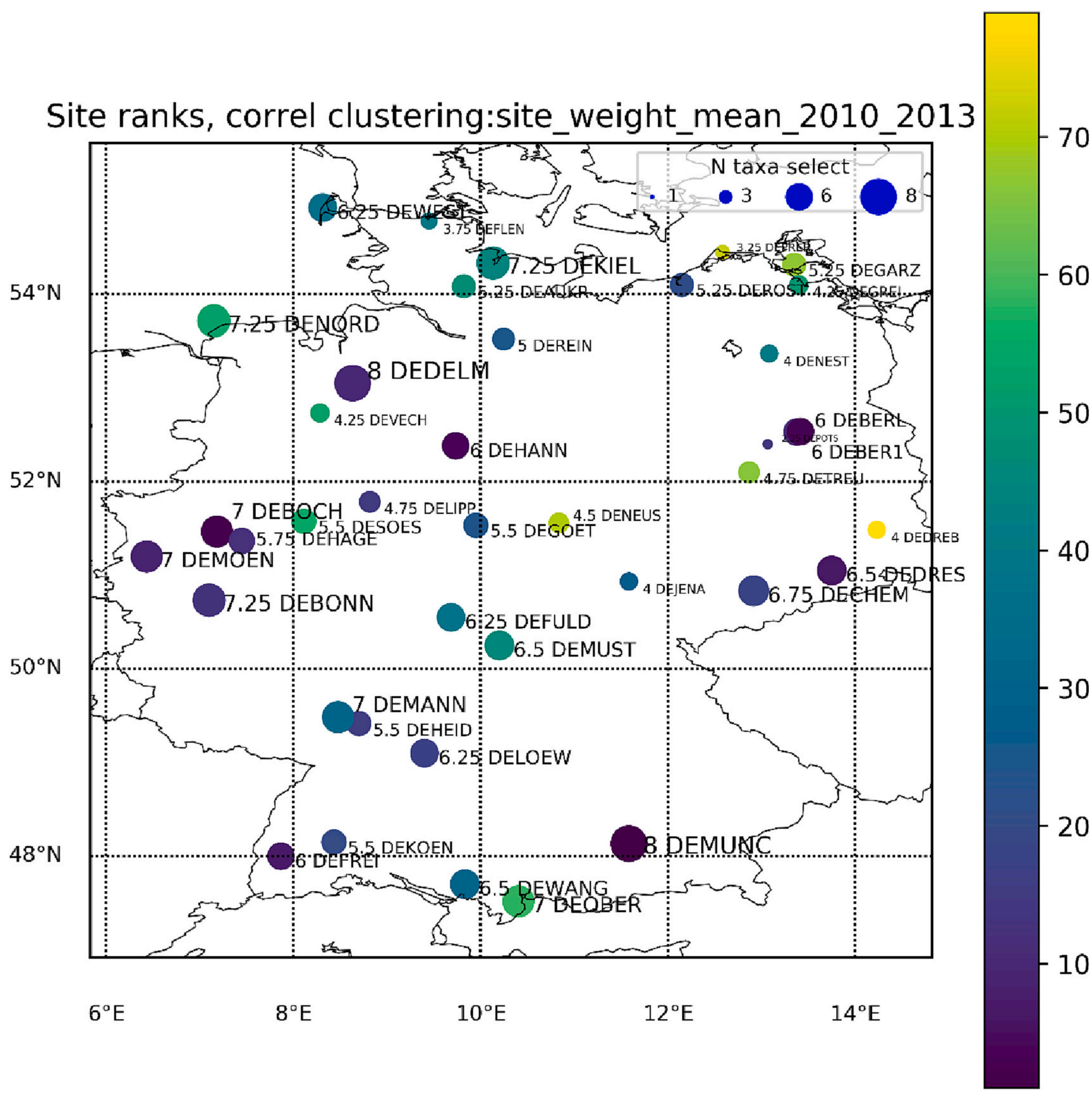


Fig. 4. Map of the final station rankings (Table 2) averaged over the period 2010–2013. The size of the dot and font of the site name are proportional to the number of taxa for which each site is important (Table 2). The colour coding follows the a-priori importance (Table 1, population and length of time series – the smaller rank denoted by dark violet, is for the most-important sites).

sites. It turned out that these 20 included most of the top 10 sites determined using the formal method, i.e., the user-defined and formally computed importance of the sites generally agree with each other.

5.1.1. Outliers: unique sources of information or poorly established observations?

Several sites appeared as outliers for one or several species. Arguably, the most-specific outlier time series are shown by DEHANN, which correlated with the rest of the network only for grasses and birch. Several other stations manifested strongly different data for one or more taxa. An interesting picture was seen for ragweed: about half of the stations correlate very well with each other, whereas the other half do not correlate at all with any other site. No clear pattern of correlations was found, e.g., there was no evidence of *Ambrosia* being present in the south and not in the north.

The root causes of poor correlation between the closely located sites or high correlation between distant stations, need to be studied separately in each case. The above analysis assumed that the data in the EAN database for Germany did not contain significant errors. As such, any two sites differing from each other were considered informative regardless the distance between them. This explains a tendency in Fig. 4, where several pairs of sites were retained, despite they are located close to each other. Common sense suggests that one of them is redundant, but their time series are different enough to retain both in the network. If future quality assurance test proves that some time series are indeed of low quality, the redundancy tests need to be repeated with reduced datasets.

5.1.2. Robustness of the site ranking

The above procedure includes several assumptions and arbitrarily constants, which can influence the final site ranking. A series of additional tests have been performed to estimate and, wherever possible, reduce, the uncertainty of the final ranking.

Variability was found to be substantial from one year to another. To reduce it, data from four sequential years, 2010–2013, were used and the rankings obtained individually for each year were averaged. Considering longer periods was not a reasonable possibility given the small number of sites with longer datasets.

Pollen observations made predominantly over the season and sometimes with substantial periods of missing data. The impact of incomplete time series was investigated by testing three completeness thresholds: 10 %, 20 %, and 30 %, i.e., the necessity of having 36, 73, and 109 days of data each year. The final ranking for all thresholds was essentially the same (not shown).

To further evaluate the robustness of the selection procedure, the key step was modified: the reduction of the correlated clusters was changed to the reduction of the largest cluster rather than the cluster containing the most important site. The differences turned out to be negligible. The proposed method can therefore be considered sufficiently robust.

5.2. Network coverage

For this step, the footprints of the 40 informative German sites (Table 2) were computed and summed over two months in spring, covering the extended birch season. The procedure of the analysis was the following:

1. for each day 15.03.2014–15.05.2014, the footprint was calculated for the previous three days, including the day of the observation. The footprint map was saved at an hourly resolution;
2. each hourly footprint was scaled with an assumed diurnal cycle of the birch pollen release (Fig. 3), i.e., the areas covered by the footprint during the night were removed and the diurnal sections were scaled up;
3. the scaled hourly footprints were summed for each station and each day of observation, thus obtaining the total footprint per day and site (the result of this step is shown in Fig. 5);
4. the station-day footprints were summed in different combinations:
   - o for all observation times of each station: footprints of individual stations ($\Phi_{station}$)
   - o for all observation times and all stations: footprint of the whole network (Φ_{total})
   - o for all observation times and all stations, each footprint was weighted with the observed pollen concentrations (if available) ($\Phi_{weighted}$ highlighted the footprints corresponding to the highest birch pollen loads in Germany in 2014, i.e., the areas that were the most productive).

Sum of the individual station footprints over the 2014 season, for both uniformly scaled (Φ_{total}) and weighted footprints ($\Phi_{weighted}$), shows that the network coverage in 2014 was not homogeneous in Germany (Fig. 6). A few areas (dark blobs) are covered excessively (see also the redundancy tests above), whereas some others (light-colour areas) have considerably lower footprint density. The coverage of sources outside Germany was also lower, except for the western part of Poland, whose birch sources affect the Eastern German sites and show up in the footprint map regularly. Large forests in Russia do not affect Germany on a regular basis, therefore the footprint density there is low. One can expect, however, that the existing sites in Eastern Germany will be able to observe episodic plumes if they reach the region.

The weighted footprint map has a different pattern than the uniform one. As can be expected, it correlates with the map of birch productivity,

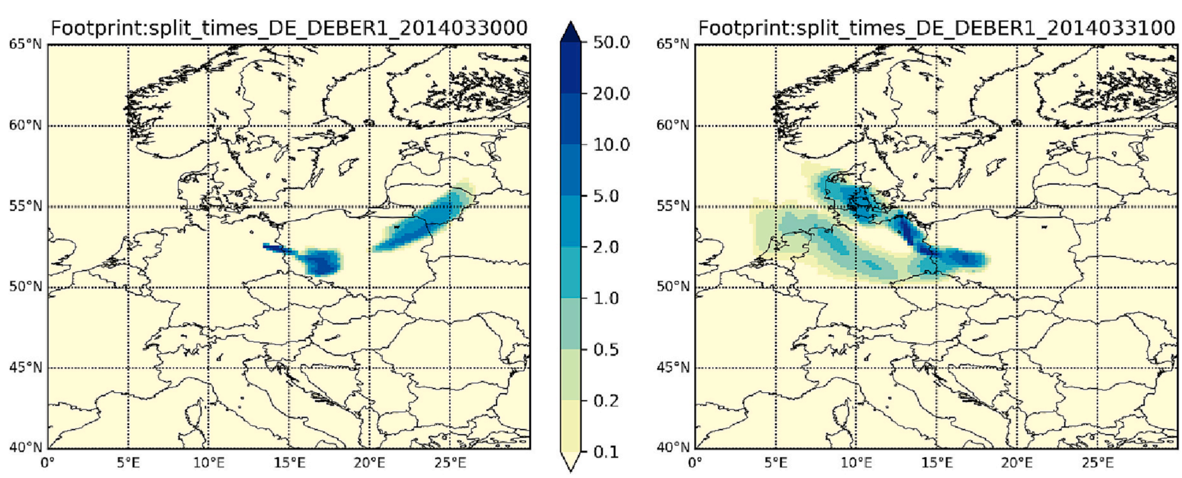


Fig. 5. Footprints for two sequential days for the Berlin station (30.03–31.03.2014). Units are relative.

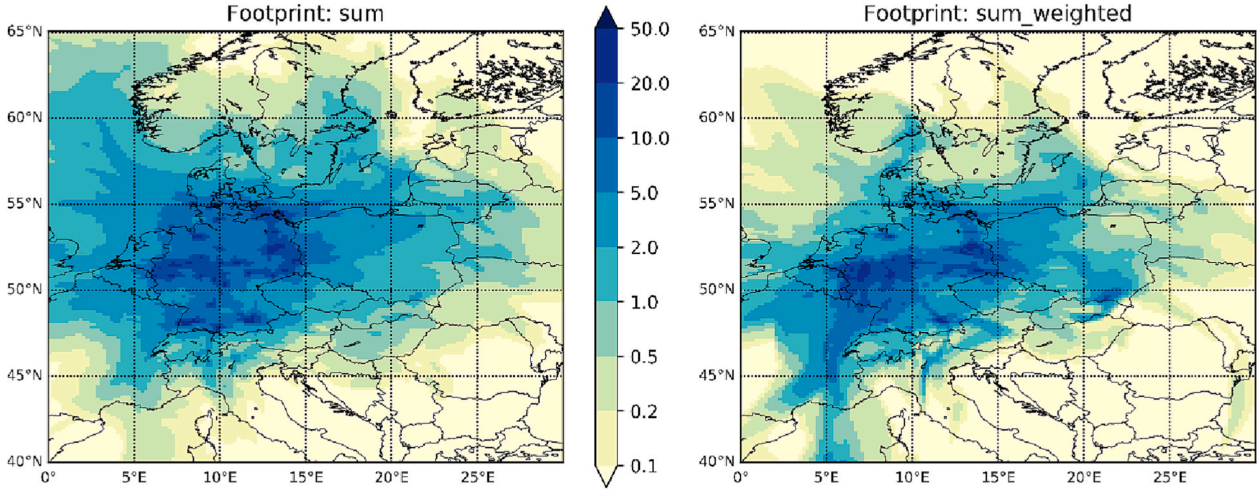


Fig. 6. Sum of the station footprints over the 2014 birch pollen season: Φ_{total} (left-hand panel) and scaled with the observed concentrations: Φ_{scaled} (right-hand panel).

with two main forests in the northeast and northwest of the country being responsible for the bulk of the birch pollen load in 2014.

The results presented above are for 2014; another year would likely show somewhat different patterns, especially when it comes to long-range transport episodes (section 7.2).

5.2.1. Coverage of reduced network based on the Table 2 ranking

To demonstrate the effect of removal of the redundant sites, the annual footprints for the period from 15.03 to 15.05.2014 were compared for the full network (all 40 sites, Fig. 6), as well as for the 20 and 30 most-informative sites presented in Table 2, respectively (Fig. 7).

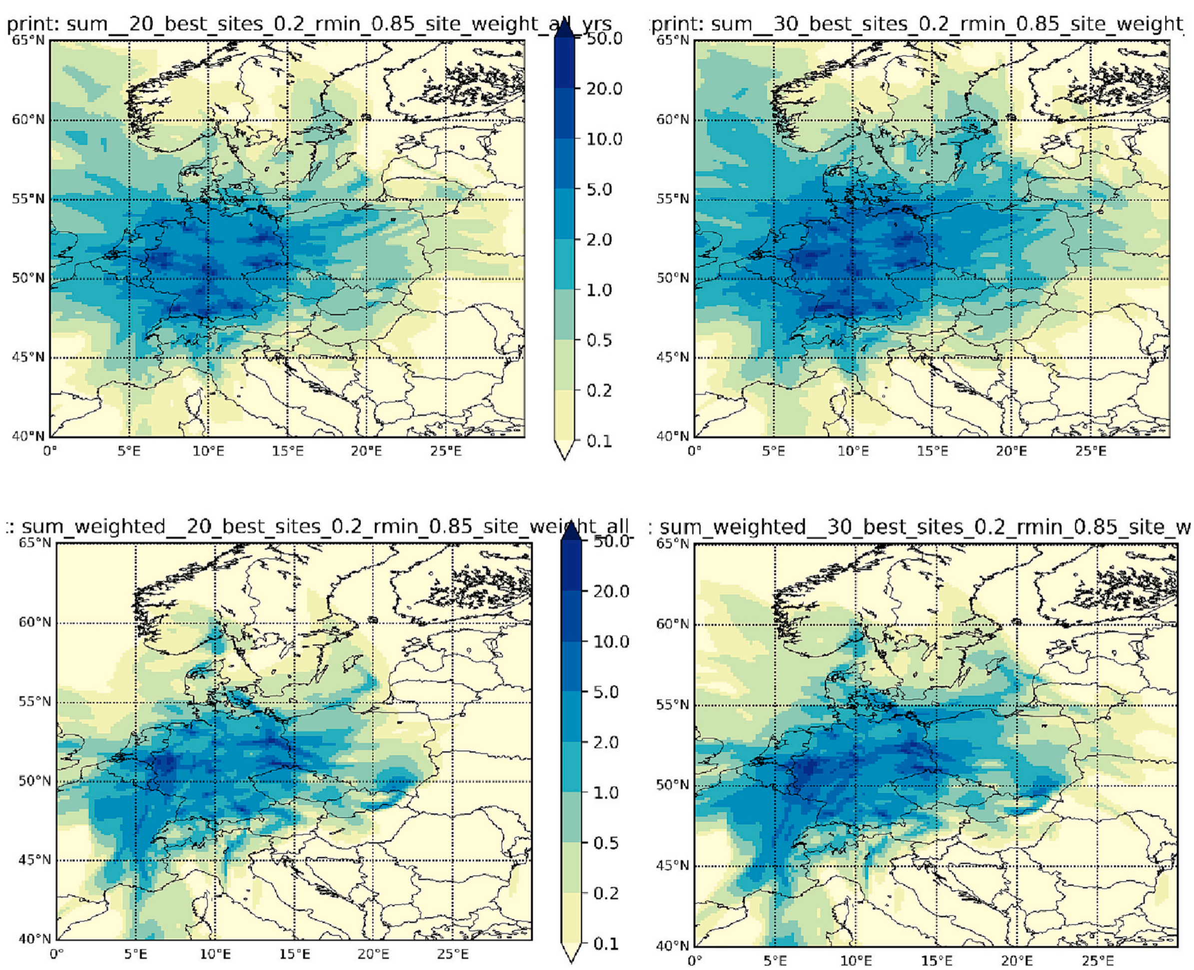


Fig. 7. Sum of the station footprints over the 2014 birch pollen season. Only footprints of the 20 (left) and 30 (right) most informative sites were used. The upper row shows the total footprints, while the lower row shows the weighted footprints.

Comparing Fig. 6 and Fig. 7, one can see that the difference between the 30 best sites and the whole network is relatively small; the entire German territory is covered well, with small observational gaps in both cases. Probably, the only area for improvement is in the north-east and north-west of Bavaria, where new sites were installed as part of the ePIN automatic monitoring project. In contrast, for the 20 best sites, the coverage started to deteriorate, with the north-east of the country being relatively sparsely covered and the coverage being lower in general. This also includes areas adjacent to Germany, particularly in Poland and France.

5.3. Quantifying the network fidelity for direct usage of observations

The above screening procedures result in a non-redundant network and qualitatively showed its coverage. They also provided the site ranking, which can be used as a guideline for adding-removing the stations. This section demonstrates the quantitative assessment of the network fidelity following the formal procedure of Sections 4.2 and 5.3 taking the cost function (5) - (6c) as the target criterion.

To generate the synthetic concentration fields, SILAM was run for the year 2011 for 5 species (alder, birch, grass, mugwort, and ragweed), and its concentration maps (example for birch is in Fig. 8 panel a) were sampled at the station locations of the tested networks. The synthetic observations obtained from the sampling were inter−/extrapolated over the domain (Fig. 8 b) and compared with the SILAM predictions (Fig. 8 c). The RMSE was computed following the Eqs. (5) - (6 c).

An example of hourly fields in Fig. 8 illustrates the challenges of the reconstructing the fields from point observations when the spatial scale of variability of concentrations is shorter than the distance between the stations. Noteworthy, the example of Fig. 8 is for birch, which has comparatively smooth concentration patterns. Nevertheless, the pollen plumes create a pattern with much shorter scale of the spatial inhomogeneity than the distance between the stations. In such situation, a plume can pass unnoticed between the stations, see, e.g., cloud around (52.5 N, 7.5E). Alternatively, if a small plume is accidentally observed by some site(s), it will be treated as a large-scale widespread phenomenon, just as a small cloud at (52.5 N, 13.4E) that contaminated the whole north-eastern part of the domain of Fig. 8.

Considering the whole season (Fig. 9), the evolution of the interpolation error J_d of eq. (5) in comparison with the season strength W_d of eq. (6c) shows that the network of 95 stations provides only a limited improvement compared to 20 well-selected non-redundant sites (panels a and b of Fig. 9, Fig. 10a). Usage of very few stations (e.g., 2) is counterproductive, the error is larger than would be in the case of no observations at all (Fig. 10a). With stations added one-by-one in the order of importance (Table 2), the network fidelity grows, but the improvement stalls after ~10 sites (small fluctuations of the curves is noise due to just one year considered).

The turning points of the curves of Fig. 10 at ~10 stations reflect the separation of scales in the problem: (i) a regional season propagation is easy to catch – already 10 sites over Germany may be sufficient, (ii) small-scale variations driven by source area patchiness, meteorological variability (wind, rain, …) are much more difficult and require dense networks to represent. Relative importance of these two factors depends on source locations and features of plants. Thus, ragweed is easily observed by few sites because its main sources are remote, so that the pollen plumes are wide and smooth.

Overall, the interpolation error is a substantial fraction of the signal itself, sometimes reaching or exceeding it (Fig. 9, Fig. 10). Particularly problematic are times at the beginning and at the end of the season when only part of the region is covered by pollen clouds, so that only few stations observe them (this explains strange dependency for alder – see section 7.5). This is the inevitable feature of the direct usage of observations for restoring the concentration pattern: it quickly (with ~10 sites) provides a rough impression on the situation but for many taxa requires far too many sites (95 was not enough) to obtain the error below 50 % of the signal itself.

6. Discussion

6.1. Selection of the target variable: concentration directly represented from the observations

Selecting the target variable –concentration or emission – is a crucial step defining the purpose of the network, with several major consequences.

A network designed for reproducing concentrations field can be used for direct monitoring given that the location and number of stations are sufficient to reproduce concentrations anywhere in the area of interest using inter/extrapolation techniques purely based on the observations (i.e., without usage of models). The downside is that the number of stations usually needs to be large to ensure reasonable results, which implies a costly dense network or cheap but reliable fully automatic sensors, which are not available today.

6.2. Temporal and spatial patterns of concentrations and the site representativeness

Intuitively, it is clear that relying just on spatial inter- or extrapolation without accounting for dispersion processes would require a dense measurement network, mainly because pollen concentration fields are known to have sharp gradients (Siljamo et al., 2006; Sofiev, 2016;

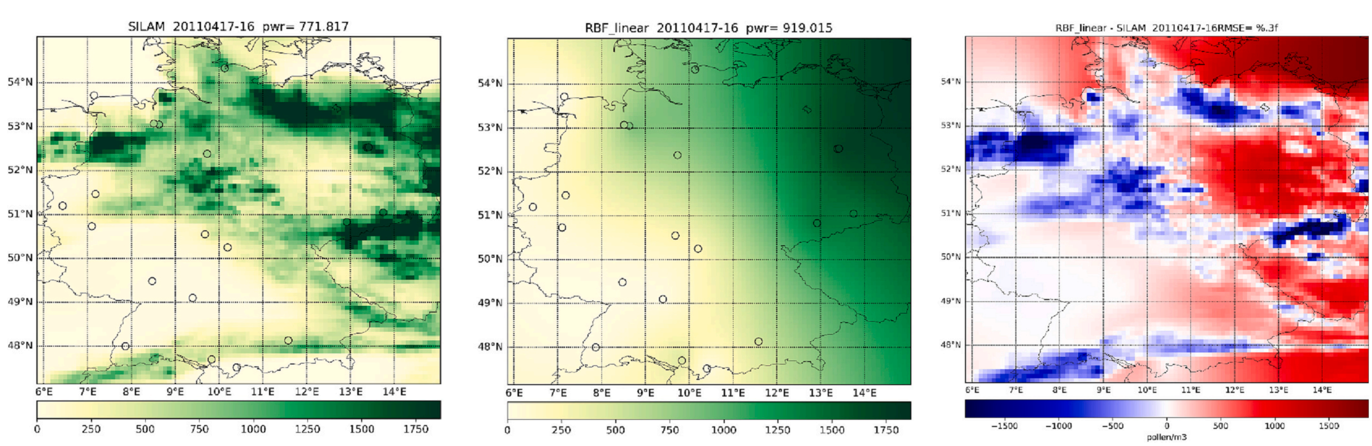


Fig. 8. Testing a network with 20 stations. Station locations and a concentration field obtained from SILAM (a); a field reconstructed from sampling of the SILAM field at the station locations (b); Panel c) shows the difference between the maps a) and b).

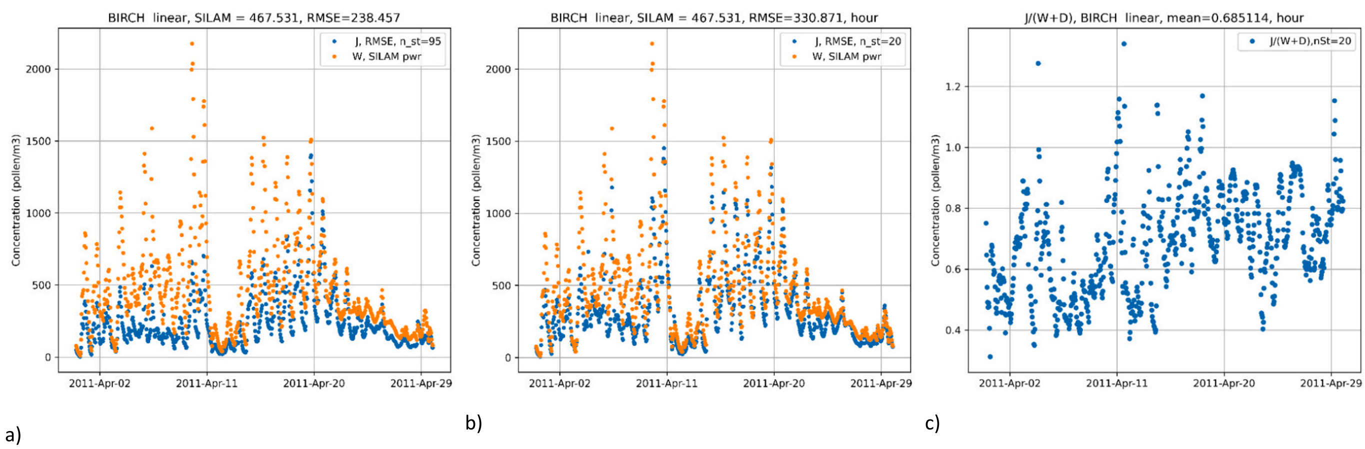


Fig. 9. Time series for the birch season strength W_d *(t)* and the interpolation error J_d *(t)* of eqs. (5), (6c) for the current German network of 95 stations (Panel a)), for 20 important stations Table 2 (Panel b)), and normalized interpolation error J_d/W_d for the 20-station network, Panel c).

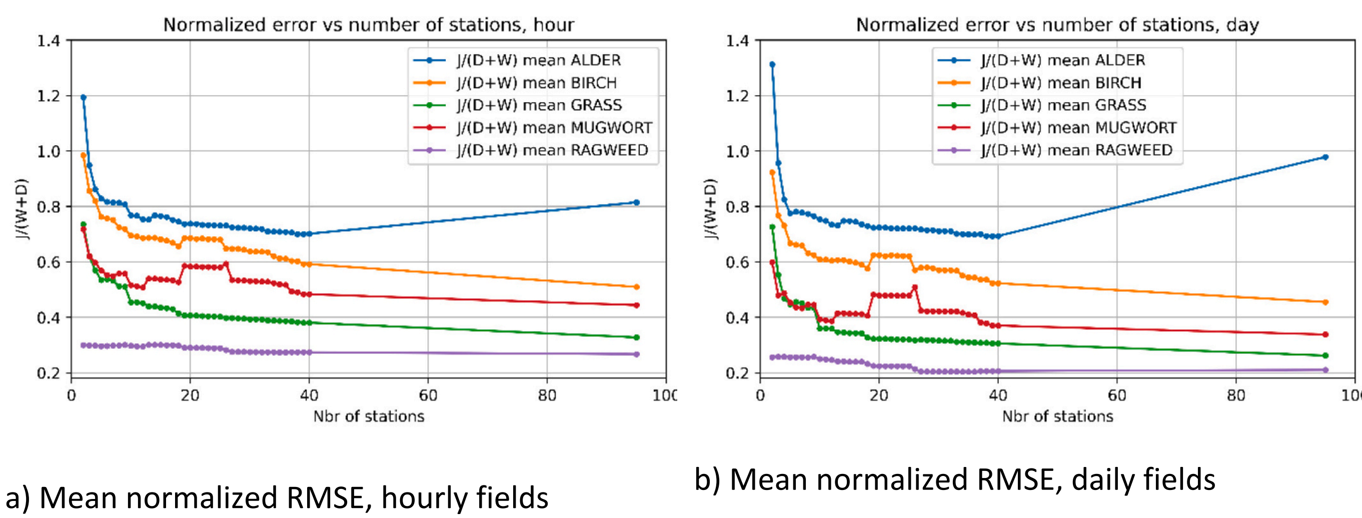


Fig. 10. Dependence of the RMSE of the reconstructed field (Fig. 8) on the number of sampling stations and averaging period of concentration fields (panel a – hourly, panel b - daily). Mean over 2011. The panel b is discussed in section 7.3.

Veriankaitė et al., 2010; Vogel et al., 2008; Zink et al., 2013). A pollen monitoring network satisfying this criterion would probably be much denser than one can afford.

It is also evident that for various pollen types, the homogeneity of the concentration pattern across the network is different. For example, spatial patterns are the most homogeneous for *Betula* and *Poaceae*, so the time series at different sites would correlate highly for these pollen types (Fig. 11). In contrast, *Ambrosia* pollen and *Cladosporium* spores have more varying patterns, so the data from different stations correlate worse. As a result, observations of, e.g., *Betula*, would require fewer stations than for *Ambrosia*. Interestingly, *Betula* is among the far-transported pollen types, which makes the concentration patterns smooth, but *Poaceae* pollen transport is predominantly local – but their sources are so widespread that the resulting concentration fields are also homogeneous.

Furthermore, there are substantial differences between years. Even at the mean level (Fig. 11), 2013 was an extreme year since almost all species showed either the highest or the lowest correlation between the sites. 2010 was also unique, whereas 2011 and 2012 show similar spatial homogeneity.

6.3. Network accuracy vs temporal resolution

Automatic measurements are produced with a high temporal resolution, therefore the above synthetic case used the hourly SILAM output fields. However, a vast majority of allergology-related recommendations are based on daily concentrations, on a seasonal pollen integral (a sum of daily-mean concentrations over a year), which are used for describing intensity of the pollen season. Similarly, most of agriculture and forestry-relevant parameters are also daily, weekly, or seasonal. This is partly historical: with data being daily/seasonal, formulating and evaluating criteria based on, e.g., hourly values, was not possible. The situation may change with automatic data becoming available, which raises the question: how much of the spatial representativeness error does one add when switching, e.g., from daily to hourly averaging?

This question directly refers to the needed network density. The longer the acceptable averaging period, the larger the correlation distance for an individual station and the lower the number of stations required. The latter, however, reaches a minimum and no longer decreases as the averaging period increases beyond the synoptic variability period, approximately 3–4 days.

The synthetic experiment quantifying the network fidelity allows a quick check of the effect. Averaging the SILAM hourly fields to the daily

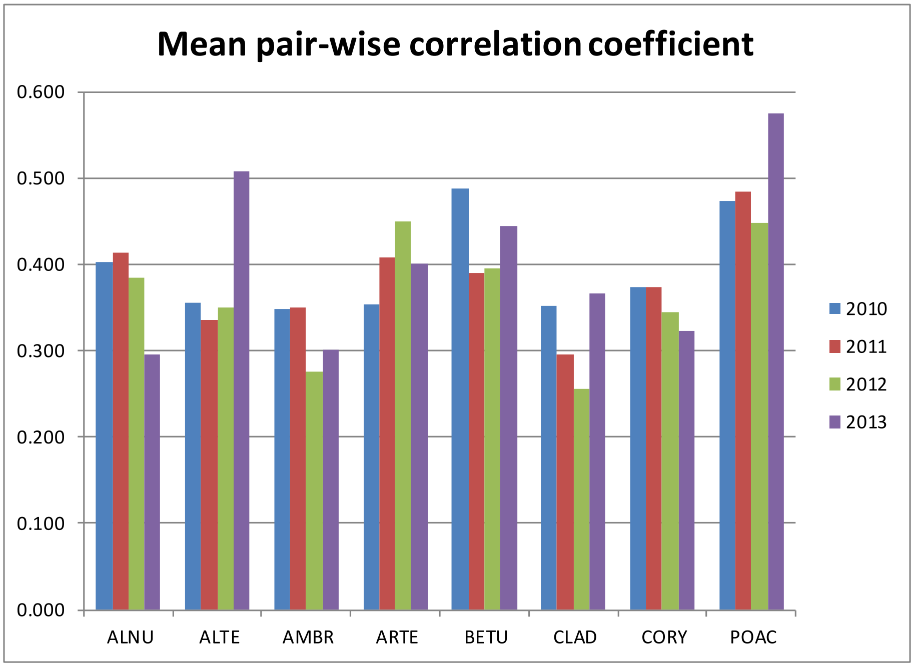


Fig. 11. Mean pairwise correlation coefficients for the four years and eight different taxa considered.

level and repeating the error calculations, one obtains that the impact of the change of the temporal resolution on the spatial representativeness error is within 10–20 % depending on taxon (compare the panels a and b of Fig. 10).

6.4. The value of redundant stations in a network

For designing a minimal network, removing redundant stations is reasonable to reduce the cost of the network. But it should be kept in mind that with this step, advantages of redundant measurements are lost. For example, it will not be possible to control the long-term stability or representativeness of the remaining non-correlated stations. With redundant stations, measured changes in the timing or intensity of the pollen season can be compared between the correlated stations. A station with locally influenced changes that are not representative of a larger area can then be detected.

Additionally, stations with high correlations in the short-term temporal pattern during a single pollen season may show differences in the long-term development of the intensity of the seasons (Gehrig, 2019). With several similar stations, a mean state of the pollen season of a region can be described, and they can bring advantages in recording possible different future developments within this region. An additional benefit is that redundant stations can be used for filling measurement gaps due to the technical failures of an individual monitor. This is especially important when long-term changes are analysed.

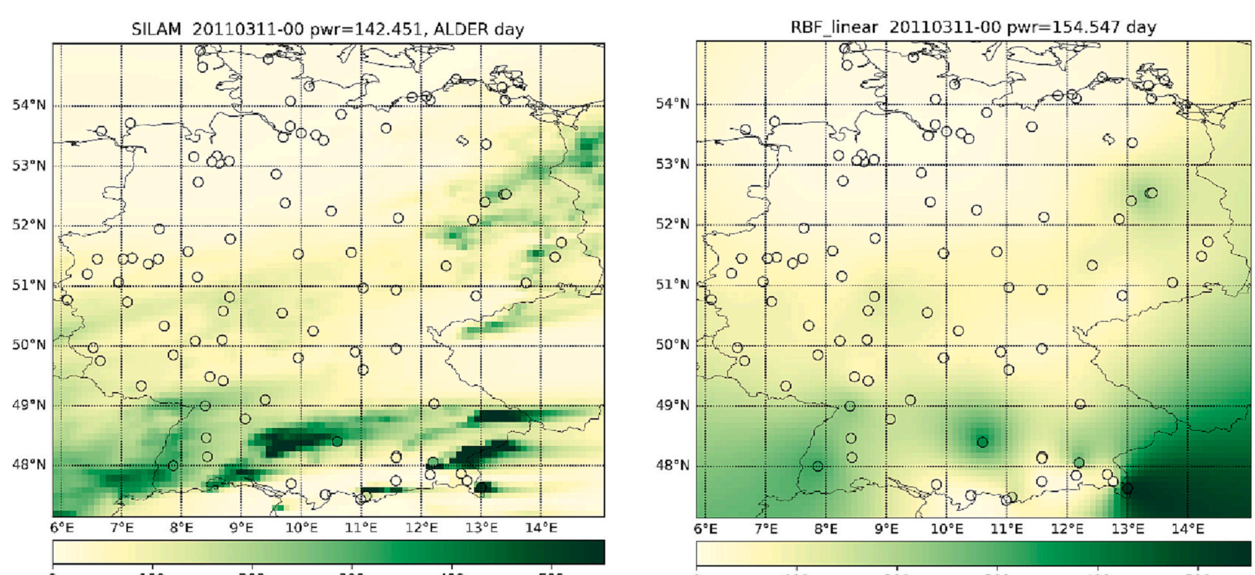


Fig. 12. Panel a): SILAM daily mean *Alnus* pollen concentration on 1 March. Panel b): daily concentration field on 11 March restored from the sampling of the SILAM field with 95 stations.

6.5. Why the fidelity for Alnus is poor for the full network?

As seen from Fig. 9 and Fig. 10, mean RMSE for *Alnus* behaves differently from those for other taxa: the full network of 95 stations turned out to be noticeably worse than, e.g., network of 20 stations. Seemingly counter-intuitive, this behaviour illustrates the problem of insufficient density of the network for direct usage of the observations.

Fig. 12a shows that on 11 March, a small but dense *Alnus* pollen cloud passed over one station in the south-east of Germany. Since there were no sites to the east of this point, this cloud was not constrained and the interpolated field (Fig. 12b) showed the whole lower-right corner of the domain with high concentrations. Simultaneously, other small clouds were missed by all stations. Together, these false and missed plumes resulted in a very poor representation of the pattern. The problem shows up only for the full network because the 20 top-ranked sites do not include the problematic point, i.e., they miss this small cloud completely and avoid over-stating the episode. The issue again stresses the consideration above: local-scale phenomena cannot be captured within the concept of a direct usage of observations for restoring the concentration fields.

7. Summary

Responding to the three obectives of the study, a methodology of a network design for aerobiological automatic monitoring is discussed combining the rigorous principles of Section 4 and practically feasible techniques of Sections 3 and 5 (objective (i)). These techniques are organized in a multi-step practical procedure (objective (ii)), and its application is demonstrated for the example of the network of Germany (objective (iii)).

The suggested procedure includes several steps.

Step 1. Preparations. Configuration of an observational network is critically dependent on several a-priori characteristics, which has to be determined before starting the design: (i) the purpose of the monitoring and its protocol must be established based on the expected data usage, applications, key stakeholder needs, and available resources; (ii) existing observations must be quality-checked and cleaned from poorly performing and non-representative stations, (iii) from the network purpose, the quantitative criterion of its quality should be formulated, as well as its required threshold; (iv) all user- and technology- driven restrictions and additional constraints should be identified and formalized.

The methodology shown in this paper follows the current nearly-consensus concept that the observations are directly used to represent the concentration fields and serve the data users. The corresponding network quality criterion is then the accuracy of reproduction of pollen concentrations anywhere in the target region by means of interpolation between the stations. The target function is the RMSE of the field obtained from the station data interpolation in comparison with the true pattern.

Step 2. Introduction of necessary simplifications. Due to computational limitations, all practical network design projects apply numerous simplifications reducing the problem dimension. A popular approach, also used in this study, is to use an existing “proxy” network, which is optimized for the given criterion. In the provided example for pollen/spores network in Germany, an existing national network was used as a prominent starting point.

Step 3. Analysis of the starting proxy network. The design procedure starts from pre-screening of the initial proxy-network: reduction via redundancy test followed by the site ranking and coverage evaluation. The output of this step is a non-redundant subset of the initial network.

Step 4. Iterative network optimization. The step-by-step removal of least-informative sites using the ranking of the Step 3, and/or introduction of additional stations improving the network skills leads to the final network setup.

Analysis of the synthetic case showed a separation of scales in the pollen concentration patterns: (i) a regional season propagation is smooth and can be resolved by a few well-positioned sites, (ii) smaller-scale variations driven by source area patchiness and meteorological variability (wind, rain, …) require a far too dense network to resolve within the concept of the direct usage of observation data.

CRediT authorship contribution statement

Mikhail Sofiev: Conceptualization, formal analysis, methodology, supervision, funding acquisition, writing original draft, editing.
Jeroen Buters Conceptualization, methodology, data curation, editing.
Fiona Tummon Conceptualization, methodology, funding acquisition, editing, project administration.
Yalda Fatahi, formal analysis, editing.
Olga Sozinova, formal analysis, editing.
Beverley Adams-Groom, editing.
Karl Christian Bergmann Conceptualization, methodology, data curation, editing.
Åslög Dahl, editing.
Regula Gehrig, formal analysis, editing.
Stefan Gilge, formal analysis, editing.
Andrea Kofol Seliger: formal analysis, editing.
Rostislav Kouznetsov: methodology, editing.
Gian Lieberherr, editing.
David O'Connor, editing.
Jose Oteros Conceptualization, formal analysis, methodology, editing.
Julia Palamarchuk, formal analysis, editing.
Helena Ribeiro, editing.
Barbora Werchan, data curation, editing.
Matthias Werchan, data curation, editing.
Bernard Clot Conceptualization, funding acquisition, editing, project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The team of the co-authors expresses the sincere gratitude to the German Pollen Information Service Foundation (PID) for providing pollen data from the PID monitoring network for this manuscript and also thanks the pollen analysts for their work. The current manuscript is one of contributions to the EUMETNET AutoPollen Programme. The computations have been performed within the Academy of Finland project PS4A (grant 318194) and ALL-IMPRESS (grant 329215). Support of Copernicus Atmospheric Monitoring Service CAMS-50 and CAMS2_40 is kindly appreciated. Olga Sozinova acknowledges the support from EC ERDF PostDoc Latvia No 1.1.1.2/VIAA/2/18/283 project.

Appendix A

A.1. Objectives of EUMETNET Autopollen programme with regard to the European pollen monitoring network

Within the Autopollen Programme, the development of methodology and practical recommendations on the future automatic European pollen monitoring network is a responsibility of the Working Group 3. The related objectives, tasks, and the programme deliverables are as follows.

Objective 3

Suggest an optimal network configuration (number of sites, their location, etc.) and alternative, possibly more realistic, configurations and compare those to the optimal (e.g. in terms of costs, spatial coverage, etc.). Develop a description of how the quality of a network can be assessed (criteria, metrics, etc.).

3.1 Set of criteria and metrics to quantitatively design and assess the AutoPollen network across Europe.

3.2 Recommendation for an optimal network based on the selection criteria and current projects across Europe.

3.3 Based on the existing observation network, develop a design for its extension across Europe to meet end-user needs (particularly numerical forecast models).

Table A.1
The relative importance of the station based on population within ~35 km from it and the length of the time series collected by 2015.

	dx = 0.333333						
Site	population count 2005	Years obs. 1970–2014	Days observed 1970–2014	Rank population	Rank years	Sum of ranks	Final rank
DEMUNC	1,360,090	22	6507	6	8	14	1
DEBOCH	1,568,150	21	4614	3	12	15	2
DEBERL	1,872,870	20	5040	2	20	22	3
DEHANN	386,712	26	4320	28	3	31	4
DEERLA	688,240	21	3607	18	14	32	5
DEDRES	446,392	21	6105	25	10	35	6
DEFREI	352,737	23	6888	30	5	35	7
DEGERL	1,074,500	19	2396	11	25	36	8
DEMOEN	685,401	20	5165	19	19	38	9
DEDELM	224,470	33	9743	39	1	40	10
DEBER1	1,872,870	12	3043	1	41	42	11
DEHAGE	1,076,120	15	2954	9	34	43	12
DEBONN	930,525	16	3142	12	31	43	13
DEPOTS	1,324,060	13	3318	8	38	46	14
DELIPP	219,532	27	6617	44	2	46	15
DEHEID	429,803	19	3124	26	23	49	16
DELOEW	247,272	21	3362	35	17	52	17
DECHEM	418,202	18	4477	27	26	53	18
DEWLOG	224,470	21	4714	42	11	53	19
DEKOEN	179,327	25	4079	51	4	55	20
DEROST	226,096	17	4092	37	27	64	21
DEBIED	1,360,090	5	1339	5	60	65	22
DEAACH	636,941	11	1365	21	44	65	23
DEGOET	167,491	21	3659	55	13	68	24
DEREIN	844,631	7	1740	15	54	69	25
DEKARL	527,091	11	1247	24	45	69	26
DEJENA	171,132	21	3428	53	16	69	27
DEFRAN	786,633	8	833	17	53	70	28
DEHAMB	844,631	6	874	14	58	72	29
DELUEB	199,495	19	2976	48	24	72	30
DEWANG	164,513	21	3594	57	15	72	31
DEMANN	849,626	5	1214	13	61	74	32
DEBUXT	792,258	6	739	16	59	75	33
DEBREM	543,429	8	1167	23	52	75	34
DEMSTR	365,943	9	1291	29	47	76	35
DEWEST	224,470	14	2785	41	36	77	36
DEMAGD	198,259	17	3394	49	28	77	37
DEFULD	128,036	22	6698	70	7	77	38
DEBAMB	206,535	14	2054	45	37	82	39
DEFLEN	147,030	20	3116	63	21	84	40
DENEST	94,479.9	23	5161	78	6	84	41
DEMOER	546,055	5	948	22	64	86	42
DEMUNI	1,360,090	2	96	7	81	88	43
DEKIEL	205,347	11	2211	46	42	88	44
DEESSE	1,568,150	1	122	4	89	93	45
DEMUST	151,596	15	3206	60	33	93	46
DEAUKR	154,655	13	2383	59	40	99	47
DEPINN	309,275	4	563	32	68	100	48
DEDONA	223,882	6	876	43	57	100	49
DEGREI	72,594.1	20	6138	82	18	100	50
DEMUEN	1,076,120	1	100	10	91	101	51
DEVECH	113,324	16	3516	72	29	101	52
DENORD	11,904.2	22	3603	92	9	101	53
DEZUSM	146,391	13	2665	64	39	103	54

(continued on next page)

Table A.1 (*continued*)

	dx = 0.333333						
Site	population count 2005	Years obs. 1970–2014	Days observed 1970–2014	Rank population	Rank years	Sum of ranks	Final rank
DESOES	167,692	8	2170	54	50	104	55
DEERFU	148,856	11	1894	62	43	105	56
DEKREU	161,292	9	1209	58	48	106	57
DEOBER	52,122	19	4988	85	22	107	58
DEOLDE	256,466	3	340	33	75	108	59
DEBADE	176,919	6	893	52	56	108	60
DEAAC1	636,941	1	105	20	90	110	61
DELEMW	224,470	3	717	40	73	113	62
DEGIES	249,832	2	195	34	80	114	63
DEROSE	203,359	4	504	47	69	116	64
DEBAYR	140,481	9	1092	68	49	117	65
DETREU	70,433.1	14	3834	83	35	118	66
DEGARZ	34,619.6	16	3294	89	30	119	67
DEWARN	226,096	1	258	38	82	120	68
DEWUER	242,345	1	179	36	86	122	69
DENEUS	94,856.4	9	1827	77	46	123	70
DEBRAU	323,647	1	60	31	93	124	71
DEPRER	10,899.6	15	3590	93	32	125	72
DEMARB	182,298	2	533	50	76	126	73
DEWASS	149,722	4	167	61	72	133	74
DESCHW	84,555.6	7	1365	79	55	134	75
DECOTT	128,083	4	779	69	67	136	76
DEGEES	125,002	5	572	71	65	136	77
DETRIE	141,494	4	409	67	70	137	78
DEDREB	95,062.8	5	1011	76	62	138	79
DEBINZ	34,619.6	8	1290	88	51	139	80
DEGRAF	165,978	1	168	56	87	143	81
DEGAIS	113,233	4	318	73	71	144	82
DEHOMB	142,176	2	245	66	79	145	83
DEFREU	145,338	1	184	65	84	149	84
DEBERC	54,528.8	5	466	84	66	150	85
DETRAU	111,438	2	266	75	78	153	86
DESYLT	14,517.9	5	976	91	63	154	87
DESCHN	80,835.5	2	463	81	77	158	88
DEBORK	51,509.7	3	406	86	74	160	89
DEINZE	111,438	1	130	74	88	162	90
DEGARM	80,835.5	1	190	80	83	163	91
DEAMRU	14,517.9	1	184	90	85	175	92
DEHDMM	43,492.9	1	92	87	92	179	93

Table A.2
List of informative sites for 2010.

Taxon	Informative stations
ALNU	DEAUKR DEBERL DEBOCH DEBONN DEDELM DEDRES DEERLA DEFLEN DEFREI DEGARZ DEHAGE DEHANN DEHEID DEJENA DEKIEL DELIPP DELUEB DEMOEN DEMUNC DENEST DENORD DEOBER DESCHN DEWANG DEWEST
ALTE	DEBER1 DEBIED DEBOCH DEBONN DECHEM DEDELM DEDRES DEERLA DEFULD DEGARM DEKIEL DELOEW DEMANN DEMOEN DEMUNC DEMUST DENORD DEOBER DESCHN DETREU DEVECH DEWANG DEWEST
AMBR	DEAUKR DEBER1 DEBERL DEBIED DEBOCH DECHEM DEDELM DEDREB DEDRES DEERLA DEFREI DEFULD DEGARZ DEGOET DEGREI DEHAGE DEHANN DEHEID DEKIEL DEKOEN DELIPP DELOEW DELUEB DEMANN DEMOEN DEMUNC DENEUS DENORD DEOBER DESCHN DESOES DETREU DEWEST DEWLOG DEZUSM
ARTE	DEAUKR DEBER1 DEBERL DEBIED DEBOCH DEBONN DECHEM DEDELM DEDREB DEDRES DEERLA DEFLEN DEFREI DEFULD DEGARZ DEGOET DEGREI DEHAGE DEHANN DEHEID DEJENA DEKIEL DEKOEN DELIPP DELOEW DELUEB DEMANN DEMOEN DEMUNC DEMUST DENEST DENEUS DENORD DEOBER DEPOTS DEPRER DEREIN DEROST DESCHN DESOES DETREU DEVECH DEWANG DEWEST DEWLOG DEZUSM
BETU	DEBERL DEBOCH DECHEM DEDELM DEDRES DEERLA DEFLEN DEFREI DEGOET DEHAGE DEHANN DEHEID DEJENA DEKIEL DEKOEN DELOEW DEMANN DEMOEN DEMUNC DEMUST DENEUS DEOBER DEPOTS DEPRER DEREIN DEROST DESOES DEVECH DEWANG DEWEST DEZUSM
CLAD	DEBIED DEBONN DECHEM DEDELM DEERLA DEFULD DEGARM DEKIEL DEMANN DEMOEN DEMUNC DEMUST DENORD DEOBER DEREIN DESCHN DEWANG DEWEST
CORY	DEAUKR DEBERL DEBOCH DEBONN DEDELM DEDRES DEERLA DEFLEN DEFREI DEGARM DEGARZ DEGOET DEHAGE DEHANN DEHEID DEKIEL DEKOEN DELUEB DEMANN DEMOEN DEMUNC DEMUST DENEST DENEUS DENORD DEPRER DESCHN DESYLT DETREU DEWEST
POAC	DEAUKR DEBER1 DEBERL DEBIED DEBOCH DEBONN DECHEM DEDELM DEDREB DEDRES DEERLA DEFLEN DEFREI DEGARM DEGARZ DEGOET DEHAGE DEHANN DEHEID DEJENA DEKIEL DEKOEN DELIPP DELOEW DELUEB DEMAGD DEMANN DEMOEN DEMUNC DEMUST DENEUS DENORD DEOBER DEROST DESCHN DESOES DETREU DEVECH DEWANG DEWEST DEWLOG DEZUSM

References

Abida, R., Bocquet, M., Vercauteren, N., Isnard, O., 2008. Design of a monitoring network over France in case of a radiological accidental release. Atmos. Environ. 42, 5205–5219. https://doi.org/10.1016/j.atmosenv.2008.02.065.

Adamov, S., Lemonis, N., Clot, B., Crouzy, B., Gehrig, R., Graber, M.-J., Sallin, C., Tummon, F., 2021. On the measurement uncertainty of Hirst-type volumetric pollen and spore samplers. Aerobiologia. https://doi.org/10.1007/s10453-021-09724-5.

Aguilar, E., Auer, I., Brunet, M., Peterson, T.C., Wieringa, J., 2003. Guidelines on climate metadata and homogenisation (WMO-TD No. WCDMP-No. 53, WMO-TD No. 1186). World Meteorological Organization, Geneva.

Amato, G., Spieksma, F., Liccardi, G., Jager, S., Russo, M., Kontou-Fili, K., Nikkels, H., Wuthrich, B., Bonini, S., 1998. Pollen-related allergy in Europe. Allergy 53, 567–578.

Behrendt, H., Becker, W., 2001. Localization, release and bioavailability of pollen allergens : the influence of environmental factors. Curr. Opin. Immunol. 13, 709–715.

Ben Dhiab, A., Ben Mimoun, M., Oteros, J., Garcia-Mozo, H., Domínguez-Vilches, E., Galán, C., Abichou, M., Msallem, M., 2017. Modeling olive-crop forecasting in Tunisia. Theor. Appl. Climatol. 128, 541–549. https://doi.org/10.1007/s00704-015-1726-1.

Berger, U.E., Jaeger, S., Smith, M., Bergmann, K.-C., 2011. PHD, the electronic patient ‘s Hayfever diary. Pulmologie 4, 24–25.

Berliner, L.M., Lu, Z.-Q., Snyder, C., 1999. Statistical Design for Adaptive Weather Observations. J. Atmos. Sci. 56, 2536–2552. https://doi.org/10.1175/1520-0469(1999)056<2536:SDFAWO>2.0.CO;2.

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G.R., Flemming, J., Inness, A., Pagowski, M., Pérez Camaño, J.L., Saide, P.E., San Jose, R., Sofiev, M., Vira, J., Baklanov, A., Carnevale, C., Grell, G., Seigneur, C., 2015. Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. Atmos. Chem. Phys. 15, 5325–5358. https://doi.org/10.5194/acp-15-5325-2015.

Böhner, J., Bechtel, B., 2018. GIS in Climatology and Meteorology. In: Comprehensive Geographic Information Systems. Elsevier, pp. 196–235. https://doi.org/10.1016/B978-0-12-409548-9.09633-0.

Bosch-Cano, F., Bernard, N., Sudre, B., Gillet, F., Thibaudon, M., Richard, H., Badot, P.-M., Ruffaldi, P., 2011. Human exposure to allergenic pollens: a comparison between urban and rural areas. Environ. Res. 111, 619–625. https://doi.org/10.1016/j.envres.2011.04.001.

Brasseur, G.P., Xie, Y., Petersen, A.K., Bouarar, I., Flemming, J., Gauss, M., Jiang, F., Kouznetsov, R., Kranenburg, R., Mijling, B., Peuch, V.-H., Pommier, M., Segers, A., Sofiev, M., van der Timmermans, R., Walters, S., Xu, J., Zhou, G., 2019. Ensemble forecasts of air quality in eastern China – part 1: model description and implementation of the MarcoPolo–panda prediction system, version 1. Geosci. Model Dev. 12, 33–67. https://doi.org/10.5194/gmd-12-33-2019.

Broomhead, D.S., Lowe, D., 1988. Multivariable functional interpolation and adaptive networks. Complex Systems 2, 321–355.

Buters, J.T.M., Antunes, C., Galveias, A., Bergmann, K.C., Thibaudon, M., Galán, C., Schmidt-Weber, C., Oteros, J., 2018. Pollen and spore monitoring in the world. Clin. Transl. Allergy 8, 9. https://doi.org/10.1186/s13601-018-0197-8.

Buters, J., Gonzalez-Alonso, M., Klimek, L., Simon, J.C., Treudler, R., 2021. Automatisches Pollen-Monitoring. AL 44, 932–942. doi:https://doi.org/10.5414/ALX02286.

Cecchi, L., D'Amato, G., Ayres, J.G., Galan, C., Forastiere, F., Forsberg, B., Gerritsen, J., Nunes, C., Behrendt, H., Akdis, C., Dahl, R., Annesi-Maesano, I., 2010. Projections of the effects of climate change on allergic asthma: the contribution of aerobiology. Allergy 65, 1073–1081. https://doi.org/10.1111/j.1398-9995.2010.02423.x.

Chappuis, C., Tummon, F., Clot, B., Konzelmann, T., Calpini, B., Crouzy, B., 2020. Automatic pollen monitoring: first insights from hourly data. Aerobiologia 36, 159–170. https://doi.org/10.1007/s10453-019-09619-6.

Chmielewski, F., Rötzer, T., 2002. Annual and spatial variability of the beginning of growing season in Europe in relation to air temperature changes. Clim. Res. 19, 257–264.

CIESIN, 2016. Documentation for gridded population of the world, version 4 (GPWv4). Center For International Earth Science Information Network-CIESIN-Columbia University. doi:https://doi.org/10.7927/H4D50JX4.

Clot, B., Gilge, S., Hajkova, L., Magyar, D., Scheifinger, H., Sofiev, M., Bütler, F., Tummon, F., 2020. The EUMETNET AutoPollen programme: establishing a prototype automatic pollen monitoring network in Europe. Aerobiologia. https://doi.org/10.1007/s10453-020-09666-4.

Crouzy, B., Stella, M., Konzelmann, T., Calpini, B., Clot, B., 2016. All-optical automatic pollen identification: towards an operational system. Atmos. Environ. 140, 202–212. https://doi.org/10.1016/j.atmosenv.2016.05.062.

D'Amato, G., Cecchi, L., Bonini, S., Nunes, C., Liccardi, G., Popov, T., Cauwenberge, P.V., 2007. Allergenic pollen and pollen allergy in Europe. Allergy 976–990. https://doi.org/10.1111/j.1398-9995.2007.01393.x.

de Weger, L., Bergmann, K.-C., Rantio-Lehtimäki, A., Dahl, AAslög, Buters, J.T.M., Dechamp, C., Belmonte, J., Thibaudon, M., Cecchi, L., Besancenot, J.-P., Gala, C., Waisel, Y., 2013. Impact of pollen. In: Sofiev, M., Bergmann, K.-C. (Eds.), Allergenic Pollen. A Review of the Production, Release, Distribution and Health Impacts. Springer, Netherlands, Dordrecht. https://doi.org/10.1007/978-94007-4881-1 p. x + 247.

Elbern, H., Strunk, A., Schmidt, H., Talagrand, O., 2007. Emission rate and chemical state estimation by 4-dimensional variational inversion. Atmos. Chem. Phys. 21.

Elkamel, A., Fatehifar, E., Taheri, M., Al-Rashidi, M.S., Lohi, A., 2008. A heuristic optimization approach for air quality monitoring network design with the simultaneous consideration of multiple pollutants. J. Environ. Manag. 88, 507–516. https://doi.org/10.1016/j.jenvman.2007.03.029.

ETC/BD, 2017. Biogeographical regions in Europe.

Galán, C., García-Mozo, H., Vázquez, L., Ruiz, L., De La Guardia, C.D., Trigo, M., 2005. Heat requirement for the onset of the Olea europaea L. pollen season in several sites in Andalusia and the effect of the expected future climate change. Int. J. Biometeorol. 49, 184–188.

Galán, C., Smith, M., Thibaudon, M., Frenguelli, G., Oteros, J., Gehrig, R., Berger, U., Clot, B., Brandao, R., Group, E.Q.W, 2014. Pollen monitoring: minimum requirements and reproducibility of analysis. Aerobiologia 30, 385–395.

Galán, C., Alcázar, P., Oteros, J., García-Mozo, H., Aira, M.J., Belmonte, J., Diaz de la Guardia, C., Fernández-González, D., Gutierrez-Bustillo, M., Moreno-Grau, S., Pérez-Badía, R., Rodríguez-Rajo, J., Ruiz-Valenzuela, L., Tormo, R., Trigo, M.M., Domínguez-Vilches, E., 2016. Airborne pollen trends in the Iberian Peninsula. Sci. Total Environ. 550, 53–59. https://doi.org/10.1016/j.scitotenv.2016.01.069.

Gehrig, R., 2019. Representativeness of pollen traps: a review of the national pollen network of Switzerland. Aerobiologia 35, 577–581. https://doi.org/10.1007/s10453-019-09593-z.

Gray, L.J., Beer, J., Geller, M., Haigh, J.D., Lockwood, M., Matthes, K., Cubasch, U., Fleitmann, D., Harrison, G., Hood, L., Luterbacher, J., Meehl, G.A., Shindell, D., van Geel, B., White, W., 2010. Solar influence on climate. Rev. Geophys. 48, RG4001. https://doi.org/10.1029/2009RG000282.

Hansen, P.C., 1992. Analysis of discrete ill-posed problems by means of the L-curve. SIAM Rev. 34, 561–580. https://doi.org/10.1137/1034115.

Hao, Y., Xie, S., 2018. Optimal redistribution of an urban air quality monitoring network using atmospheric dispersion model and genetic algorithm. Atmos. Environ. 177, 222–233. https://doi.org/10.1016/j.atmosenv.2018.01.011.

Hirst, J.M., 1952. An automatic volumetric spore trap. Ann. Appl. Biol. 39, 257–265. https://doi.org/10.1111/j.1744-7348.1952.tb00904.x.

Hjort, J., Hugg, T.T., Antikainen, H., Rusanen, J., Sofiev, M., Kukkonen, J., Jaakkola, M.S., Jaakkola, J.J., 2015. Fine-scale exposure to allergenic pollen in the urban environment: evaluation of land use regression approach. Environ. Health Perspect. 124, 619–626. https://doi.org/10.1289/ehp.1509761.

Huijnen, V., Eskes, H.J., Poupkou, A., Elbern, H., Boersma, K.F., Foret, G., Sofiev, M., Valdebenito, A., Flemming, J., Stein, O., Gross, A., Robertson, L., D'Isidoro, M., Kioutsioukis, I., Friese, E., Amstrup, B., Bergstrom, R., Strunk, A., Vira, J., Zyryanov, D., Maurizi, A., Melas, D., Peuch, V.-H., Zerefos, C., 2010. Comparison of OMI NO2 tropospheric columns with an ensemble of global and European regional air quality models. Atmos. Chem. Phys. 10, 3273–3296.

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. Sci. Total Environ. 653, 1213–1222. https://doi.org/10.1016/j.scitotenv.2018.11.020.

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.

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. Science 220, 671–680. https://doi.org/10.1126/science.220.4598.671.

Koohkan, M.R., Bocquet, M., Wu, L., Krysta, M., 2012. Potential of the international monitoring system radionuclide network for inverse modelling. Atmos. Environ. 54, 557–567. https://doi.org/10.1016/j.atmosenv.2012.02.044.

Kouznetsov, R., Sofiev, M., 2012. A methodology for evaluation of vertical dispersion and dry deposition of atmospheric aerosols. J. Geophys. Res. 117. https://doi.org/10.1029/2011JD016366.

Kouznetsov, R., Sofiev, M., Vira, J., Stiller, G., 2020. Simulating age of air and the distribution of SF₆ in the stratosphere with the SILAM model. Atmos. Chem. Phys. 20, 5837–5859. https://doi.org/10.5194/acp-20-5837-2020.

Lauvaux, T., Schuh, A.E., Bocquet, M., Wu, L., Richardson, S., Miles, N., Davis, K.J., 2012. Network design for mesoscale inversions of CO_2 sources and sinks. Tellus Ser. B Chem. Phys. Meteorol. 64, 17980. https://doi.org/10.3402/tellusb.v64i0.17980.

Li, J., Zhang, H., Luo, Y., Deng, X., Grieneisen, M.L., Yang, F., Di, B., Zhan, Y., 2019. Stepwise genetic algorithm for adaptive management: application to air quality monitoring network optimization. Atmos. Environ. 215, 116894 https://doi.org/10.1016/j.atmosenv.2019.116894.

Meinander, O., Kontu, A., Kouznetsov, R., Sofiev, M., 2020. Snow samples combined with long-range transport modeling to reveal the origin and temporal variability of black carbon in seasonal snow in Sodankylä (67°N). Front. Earth Sci. 8, 1–11. https://doi.org/10.3389/feart.2020.00153.

Mofarrah, A., Husain, T., 2010. A holistic approach for optimal design of air quality monitoring network expansion in an urban area. Atmos. Environ. 44, 432–440. https://doi.org/10.1016/j.atmosenv.2009.07.045.

Mofarrah, A., Tahir, H.B., 2011. Design of Urban air Quality Monitoring Network: Fuzzy based multi-criteria decision making approach. In: Mazzeo, N. (Ed.), Air Quality Monitoring, Assessment and Management. InTech. https://doi.org/10.5772/16716.

Nagovitsyn, Yu.A., Kuleshova, A.I., 2012. The Waldmeier rule and early diagnostics of the maximum of the current solar cycle. Astron. Rep. 56, 800–804. https://doi.org/10.1134/S1063772912090065.

Newnham, R.M., Sparks, T.H., Skjøth, C.A., Head, K., Adams-Groom, B., Smith, M., 2013. Pollen season and climate: is the timing of birch pollen release in the UK

approaching its limit? Int. J. Biometeorol. 57, 391–400. https://doi.org/10.1007/s00484-012-0563-5.
Oteros, J., Buters, J., Laven, G., Röseler, S., Wachter, R., Schmidt-Weber, C., Hofmann, F., 2017. Errors in determining the flow rate of Hirst-type pollen traps. Aerobiologia 33, 201–210. https://doi.org/10.1007/s10453-016-9467-x.
Oteros, J., Sofiev, M., Smith, M., Clot, B., Damialis, A., Prank, M., Werchan, M., Wachter, R., Weber, A., Kutzora, S., Heinze, S., Herr, C.E.W., Menzel, A., Bergmann, K.-C., Traidl-Hoffmann, C., Schmidt-Weber, C.B., Buters, J.T.M., 2019. Building an automatic pollen monitoring network (ePIN): selection of optimal sites by clustering pollen stations. Sci. Total Environ. 688, 1263–1274. https://doi.org/10.1016/j.scitotenv.2019.06.131.
Oteros, J., Weber, A., Kutzora, S., Rojo, J., Heinze, S., Herr, C., Gebauer, R., Schmidt-Weber, C.B., Buters, J.T.M., 2020. An operational robotic pollen monitoring network based on automatic image recognition. Environ. Res. 191, 110031 https://doi.org/10.1016/j.envres.2020.110031.
Petersen, A.K., Brasseur, G.P., Bouarar, I., Flemming, J., Gauss, M., Jiang, F., Kouznetsov, R., Kranenburg, R., Mijling, B., Peuch, V.-H., Pommier, M., Segers, A., Sofiev, M., van der Timmermans, R., Walters, S., Xie, Y., Xu, J., Zhou, G., 2019. Ensemble forecasts of air quality in eastern China – part 2: evaluation of the MarcoPolo–panda prediction system, version 1. Geosci. Model Dev. 12, 1241–1266. https://doi.org/10.5194/gmd-12-1241-2019.
Pfaar, O., Bastl, K., Berger, U., Buters, J., Calderon, M.A., Clot, B., Darsow, U., Demoly, P., Durham, S.R., Galán, C., Gehrig, R., Gerth van Wijk, R., Jacobsen, L., Klimek, L., Sofiev, M., Thibaudon, M., Bergmann, K.C., 2016. Defining pollen exposure times for clinical trials of allergen immunotherapy for pollen-induced rhinoconjunctivitis - an EAACI position paper. Allergy. https://doi.org/10.1111/all.13092.
Prank, M., Chapman, D.S., Bullock, J.M., Belmonte, J., Berger, U., Dahl, A., Jäger, S., Kovtunenko, I., Magyar, D., Niemelä, S., Rantio-Lehtimäki, A., Rodinkova, V., Sauliene, I., Severova, E., Sikoparija, B., Sofiev, M., 2013. An operational model for forecasting ragweed pollen release and dispersion in Europe. Agric. For. Meteorol. 182–183, 43–53. https://doi.org/10.1016/j.agrformet.2013.08.003.
Prautzsch, H., Boehm, W., Paluszny, M., 2002. Bézier and B-Spline Techniques, Mathematics and Visualization. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-04919-8.
Ramaker, J., Mackby, J., Marshall, P.D., Geil, R., 2003. The Final Test: A History of the Comprehensive Nuclear-Test-Ban Treaty Negotiations. Provisional Technical Secretariat of the Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization.
Reza Koohkan, M., Bocquet, M., 2012. Accounting for representativeness errors in the inversion of atmospheric constituent emissions: application to the retrieval of regional carbon monoxide fluxes. Tellus Ser. B Chem. Phys. Meteorol. 64, 19047. https://doi.org/10.3402/tellusb.v64i0.19047.
Ritenberga, O., Sofiev, M., Siljamo, P., Saarto, A., Dahl, A., Ekebom, A., Sauliene, I., Shalaboda, V., Severova, E., Hoebeke, L., Ramfjord, H., 2017. A statistical model for predicting the inter-annual variability of birch pollen abundance in northern and North-Eastern Europe. Sci. Total Environ. 615 https://doi.org/10.1016/j.scitotenv.2017.09.061 in press.
Rodríguez-Fernández, A., Oteros, J., Vega-Maray, A.M., Valencia-Barrera, R.M., Galán, C., Fernández-González, D., 2022. How to select the optimal monitoring locations for an aerobiological network: a case of study in central northwest of Spain. Sci. Total Environ. 827, 154370 https://doi.org/10.1016/j.scitotenv.2022.154370.
Rodríguez-Rajo, F.J., Jato, V., Fernández-González, M., Aira, M.J., 2010. The use of aerobiological methods for forecasting *Botrytis* spore concentrations in a vineyard. Grana 49, 56–65. https://doi.org/10.1080/00173130903472393.
Rojo, J., Oteros, J., Pérez-Badia, R., Cervigón, P., Ferencova, Z., Gutiérrez-Bustillo, A.M., Bergmann, K.-C., Oliver, G., Thibaudon, M., Albertini, R., Rodríguez-De la Cruz, D., Sánchez-Reyes, E., Sánchez-Sánchez, J., Pessi, A.-M., Reiniharju, J., Saarto, A., Calderón, M.C., Guerrero, C., Berra, D., Bonini, M., Chiodini, E., Fernández-González, D., García, J., Trigo, M.M., Myszkowska, D., Fernández-Rodríguez, S., Tormo-Molina, R., Damialis, A., Kolek, F., Traidl-Hoffmann, C., Severova, E., Caeiro, E., Ribeiro, H., Magyar, D., Makra, L., Udvardy, O., Alcázar, P., Galán, C., Borycka, K., Kasprzyk, I., Newbigin, E., Adams-Groom, B., Apangu, G.P., Frisk, C.A., Skjøth, C.A., Radišić, P., Šikoparija, B., Celenk, S., Schmidt-Weber, C.B., Buters, J., 2019. Near-ground effect of height on pollen exposure. Environ. Res. 174, 160–169. https://doi.org/10.1016/j.envres.2019.04.027.
Rybnicek, O., Jaeger, S., 2001. Ambrosia (ragweed) in Europe. Allergy Clin. Immunol. Int. 13, 60–65.
Saarikoski, S., Sillanpaa, M., Sofiev, M., Timonen, H., Saarnio, K., Teinela, K., Karppinen, A., Kukkonen, J., Hillamo, R., 2007. Chemical composition of aerosols during a major biomass burning episode over northern Europe in spring 2006: experimental and modelling assessments. Atmos. Environ. 41, 3577–3589. https://doi.org/10.1016/j.atmonsenv.2006.12.053.
Šaulienė, I., Šukienė, L., Daunys, G., Valiulis, G., Vaitkevičius, L., Matavulj, P., Brdar, S., Panic, M., Sikoparija, B., Clot, B., Crouzy, B., Sofiev, M., 2019. Automatic pollen recognition with the rapid-E particle counter: the first-level procedure, experience and next steps. Atmos. Meas. Tech. 12, 3435–3452. https://doi.org/10.5194/amt-12-3435-2019.
Siljamo, P., Sofiev, M., Severova, E., Ranta, H., Polevova, S., 2006. On influence of long-range transport of pollen grains onto pollinating seasons. In: Borrego, C., Renner, E. (Eds.), Developments in Environmental Science, 6. Air Polution Modelling and Its Applications XVIII, pp. 708–716. https://doi.org/10.1016/S1474-8177(70)06074-3.
Siljamo, P., Sofiev, M., Filatova, E., Grewling, L., Jäger, S., Khoreva, E., Linkosalo, T., Ortega Jimenez, S., Ranta, H., Rantio-Lehtimäki, A., Svetlov, A., Veriankaite, L., Yakovleva, E., Kukkonen, J., 2012. A numerical model of birch pollen emission and dispersion in the atmosphere. Model evaluation and sensitivity analysis. Int. J. Biometeorol. e-pub. https://doi.org/10.1007/s00484-012-0539-5.
Skjøth, C.A., Ørby, P.V., Becker, T., Geels, C., Schlünssen, V., Sigsgaard, T., Bønløkke, J. H., Sommer, J., Søgaard, P., Hertel, O., 2013. Identifying urban sources as cause of elevated grass pollen concentrations using GIS and remote sensing. Biogeosciences 10, 541–554. https://doi.org/10.5194/bg-10-541-2013.
Smith, M., Jäger, S., Berger, U., Sikoparija, B., Hallsdottir, M., Sauliene, I., Bergmann, K.-C., Pashley, C.H., de Weger, L., Majkowska-Wojciechowska, B., Rybníček, O., Thibaudon, M., Gehrig, R., Bonini, M., Yankova, R., Damialis, A., Vokou, D., Gutiérrez Bustillo, A.M., Hoffmann-Sommergruber, K., van Ree, R., 2014. Geographic and temporal variations in pollen exposure across Europe. Allergy 69, 913–923. https://doi.org/10.1111/all.12419.
Sofiev, M., 2002. Extended resistance analogy for construction of the vertical diffusion scheme for dispersion models. J. Geophys. Res.-Atmos. 107 https://doi.org/10.1029/2001JD001233. ACH 10-1-ACH 10-8.
Sofiev, M., 2016. On impact of transport conditions on variability of the seasonal pollen index. Aerobiologia 33, 167–179. https://doi.org/10.1007/s10453-016-9459-x.
Sofiev, M., 2019. On possibilities of assimilation of near-real-time pollen data by atmospheric composition models. Aerobiologia 1. https://doi.org/10.1007/s10453-019-09583-1.
Sofiev, M., Bergmann, K.-C. (Eds.), 2013. Allergenic Pollen. A Review of Production, Release, Distribution and Health Impact. Springer-Verlag, Berlin, Heidelberg.
Sofiev, M., Siljamo, P., Valkama, I., Ilvonen, M., Kukkonen, J., 2006. A dispersion modelling system SILAM and its evaluation against ETEX data. Atmos. Environ. 40, 674–685. https://doi.org/10.1016/j.atmosenv.2005.09.069.
Sofiev, M., Genikhovich, E., Keronen, P., Vesala, T., 2010. Diagnosing the surface layer parameters for dispersion models within the meteorological-to-dispersion modeling Interface. J. Appl. Meteorol. Climatol. 49, 221–233. https://doi.org/10.1175/2009JAMC2210.1.
Sofiev, M., Siljamo, P., Ranta, H., Linkosalo, T., Jaeger, S., Rasmussen, A., Rantio-Lehtimaki, A., Severova, E., Kukkonen, J., 2012. A numerical model of birch pollen emission and dispersion in the atmosphere. Description of the emission module. Int. J. Biometeorol. 57, 54–58. https://doi.org/10.1007/s00484-012-0532-z.
Sofiev, M., Berger, U., Vira, J., Arteta, J., Belmonte, J., Bergmann, K.-C., Chéroux, F., Elbern, H., Friese, E., Galan, C., Gehrig, R., Khvorostyanov, D., Kranenburg, R., Kumar, U., Marécal, V., Meleux, F., Menut, L., Pessi, A.-M., Robertson, L., Ritenberga, O., Rodinkova, V., Saarto, A., Segers, A., Severova, E., Sauliene, I., Siljamo, P., Steensen, B.M., Teinemaa, E., Thibaudon, M., Peuch, V.-H., 2015a. MACC regional multi-model ensemble simulations of birch pollen dispersion in Europe. Atmos. Chem. Phys. 15, 8115–8130. https://doi.org/10.5194/acp-15-8115-2015.
Sofiev, M., Vira, J., Kouznetsov, R., Prank, M., Soares, J., Genikhovich, E., 2015b. Construction of an Eulerian atmospheric dispersion model based on the advection algorithm of M. Galperin: dynamic cores v.4 and 5 of SILAM v.5.5. Geosci. Model Dev. 8, 3497–3522. https://doi.org/10.5194/gmd-8-3497-2015.
Sofiev, M., Ritenberga, O., Albertini, R., Arteta, J., Belmonte, J., Bonini, M., Celenk, S., Damialis, A., Douros, J., Elbern, H., Friese, E., Galan, C., Gilles, O., Hrga, I., Kouznetsov, R., Krajsek, K., Plu, M., Prank, M., Robertson, L., Steensen, B.M., Thibaudon, M., Segers, A., Stepanovich, B., Valdebenito, A.M., Vira, J., Vokou, D., 2017. Multi - model ensemble simulations of olive pollen distribution in Europe in 2014. Atmos. Chem. Phys. 17, 12341–12360. https://doi.org/10.5194/acp-17-12341-2017.
Sofiev, M., Kouznetsov, R., Hänninen, R., Sofieva, V.F., 2020. Technical note: intermittent reduction of the stratospheric ozone over northern Europe caused by a storm in the Atlantic Ocean. Atmos. Chem. Phys. 20, 1839–1847. https://doi.org/10.5194/acp-20-1839-2020.
Tarasova, O.A., Brenninkmeijer, C.A.M., Assonov, S.S., Elansky, N.F., Roeckmann, T., Sofiev, M., 2007. Atmospheric CO along the trans-Siberian railroad and river Ob: source identification using isotope analysis. J. Atmos. Chem. 57, 135–152. https://doi.org/10.1007/s10874-007-9066-x.
Tešendić, D., Boberić Krstićev, D., Matavulj, P., Brdar, S., Panić, M., Minić, V., Šikoparija, B., 2020. RealForAll: real-time system for automatic detection of airborne pollen. Enterprise Information Systems 1–17. https://doi.org/10.1080/17517575.2020.1793391.
Tummon, F., Adamov, S., Clot, B., Crouzy, B., Gysel-Beer, M., Kawashima, S., Lieberherr, G., Manzano, J., Markey, E., Moallemi, A., O'Connor, D., 2021. A first evaluation of multiple automatic pollen monitors run in parallel. Aerobiologia. https://doi.org/10.1007/s10453-021-09729-0.
Tummon, F., Bruffaerts, N., Celenk, S., Choël, M., Clot, B., Crouzy, B., Galán, C., Gilge, S., Hajkova, L., Mokin, V., O'Connor, D., Rodinkova, V., Sauliene, I., Sikoparija, B., Sofiev, M., Sozinova, O., Tesendic, D., Vasilatou, K., 2022. Towards standardisation of automatic pollen and fungal spore monitoring: best practises and guidelines. Aerobiologia. https://doi.org/10.1007/s10453-022-09755-6.
Veriankaitė, L., Siljamo, P., Sofiev, M., Sauliene, I., Kukkonen, J., 2010. Modelling analysis of source regions of long-range transported birch pollen that influences allergenic seasons in Lithuania. AEROBIOLOGIA 26, 47–62. https://doi.org/10.1007/s10453-009-9142-6.
Vira, J., Sofiev, M., 2012. On variational data assimilation for estimating the model initial conditions and emission fluxes for short-term forecasting of SOx concentrations. Atmos. Environ. 46, 318–328. https://doi.org/10.1016/j.atmosenv.2011.09.066.
Vogel, H., Pauling, A., Vogel, B., 2008. Numerical simulation of birch pollen dispersion with an operational weather forecast system. Int. J. Biometeorol. 52, 805–814. https://doi.org/10.1007/s00484-008-0174-3.

Wang, C., Zhao, L., Sun, W., Xue, J., Xie, Y., 2018. Identifying redundant monitoring stations in an air quality monitoring network. Atmos. Environ. 190, 256–268. https://doi.org/10.1016/j.atmosenv.2018.07.040.
Werchan, B., Werchan, M., Mücke, H.-G., Bergmann, K.-C., 2018. Spatial distribution of pollen-induced symptoms within a large metropolitan area—Berlin, Germany. Aerobiologia 34, 539–556. https://doi.org/10.1007/s10453-018-9529-3.
WMO, 2017. WMO Guidelines on the Calculation of Climate Normals (WMO Guidelines No. 1203). World Meteorological rganization, Geneva.
WMO, 2020. Guidelines on homogenization (No. 1245). WMO.
Xian, P., Reid, J.S., Hyer, E.J., Sampson, C.R., Rubin, J.I., Ades, M., Asencio, N., Basart, S., Benedetti, A., Bhattacharjee, P.S., Brooks, M.E., Colarco, P.R., da Silva, A. M., Eck, T.F., Guth, J., Jorba, O., Kouznetsov, R., Kipling, Z., Sofiev, M., Perez Garcia-Pando, C., Pradhan, Y., Tanaka, T., Wang, J., Westphal, D.L., Yumimoto, K., Zhang, J., 2019. Current state of the global operational aerosol multi-model ensemble: an update from the international cooperative for aerosol prediction (ICAP). Q.J.R. Meteorol. Soc. 145, 176–209. https://doi.org/10.1002/qj.3497.
Ziello, C., Böck, A., Estrella, N., Ankerst, D., Menzel, A., 2012. First flowering of wind-pollinated species with the greatest phenological advances in Europe. Ecography 35, 1017–1023. https://doi.org/10.1111/j.1600-0587.2012.07607.x.
Zink, K., Pauling, A., Rotach, M.W., Vogel, H., Kaufmann, P., Clot, B., 2013. EMPOL 1.0 : a new parameterization of pollen emission in numerical weather prediction models. Geosci. Model Dev. 6, 1961–1975. https://doi.org/10.5194/gmd-6-1961-2013.