UNIVERSITY OF LATVIA
FACULTY OF GEOGRAPHY AND EARTH SCIENCES

Olga Ritenberga

FORECASTING THE GEOSPATIAL AND TEMPORAL PATTERNS OF POLLEN SEASON IN EUROPE USING STATISTICAL AND DETERMINISTIC MODELLING

DOCTORAL THESIS

Submitted for the Degree of Doctor of Geography
Subfield of Physical Geography

RIGA 2017

The doctoral thesis was carried out in Department of Physical Geography, Faculty of Geography and Earth Science, University of Latvia from the year 2010 to 2012 and in University of Latvia and Finnish Meteorological Institute from the year 2014 to 2017.

This study was supported by science-based funding of Latvian Ministry of Education and Science via "Attraction of Human Resources to Development of Scientific Study in the area of Earth and Environmental Sciences" programme; performance-based funding of University of Latvia Nr. AAP2016/B041//ZD2016/AZ03 within the "Climate change and sustainable use of natural resources" programme; National Research Program of Latvia 2014.10-4/VPP-1/27-LATENERGI.

Supervisors:

Mikhail Sofiev, Professor, *Doc. Phys., Dr. Sci. Tech.*, Finnish Meteorological Institute, Helsinki, Finland
Laimdota Kalniņa, Assoc. Professor, *Dr. geogr.*, University of Latvia, Riga, Latvia

Reviewers:

Jānis Kleperis, *Dr. phys.*, Institute of Solid State Physics, Riga, Latvia
Agrita Briede, *Dr. geogr.*, University of Latvia, Riga, Latvia
Lukasz Grewling, *PhD*, Adam Mickiewicz University, Poznan, Poland

Doctoral Council:

Agrita Briede, *Dr. geogr.* – Head of the Doctoral Council
Oļģerts Nikodemus, *Dr. geogr.*
Raimonds Kasparinskis, *Dr. geogr.*
Māris Kļaviņš, *Dr. habil. chem.*
Viesturs Melecis, *Dr. biol.*
Solvita Rūsiņa, *Dr. geogr.* – Secretary of the Doctoral Council
Iveta Šteinberga, *Dr. geogr.*

The thesis will be defended in the public session of the Doctoral Committee of Geography University of Latvia, at the Faculty of Geography and Earth Sciences (1 Jelgavas iela, Riga, Latvia) on December 29, 2017 at 13:00.

The thesis is available at the University of Latvia Library, Raiņa bulvāris 19, Riga, Latvia.

Address for submitting comments:
Dr. Solvita Rūsiņa, Department of Geography, Faculty of Geography and Earth Sciences, University of Latvia, Raiņa bulvāris 19, LV-1586, Riga. E-mail: solvita.rusina@lu.lv

ISBN 978-9934-18-294-5

ANNOTATION

The thesis introduces forecasting of the temporal and geospatial patterns of pollen season in Europe (including influencing factors and daily concentration of pollen) through complex, multi-step transformation of data and parametric statistical analysis for model development and validation. Simulations were performed for several plant taxa at various (local, regional) spatial scales.

The first part of the thesis develops a local flowering model for Riga. It is based on 12 years of observational data and aims at predicting the concentration of pollen in the air using the data from meteorological forecast. The model accuracy exceeds 80%.

The second model is developed for regional-scale predictions of seasonal pollen index (SPI) – for the region covering Finland, Sweden, Lithuania, Latvia, Belarus, partly Russia and Norway. The SPI model is designed to be universally applicable to the entire region and has an accuracy of 65% in south-eastern part of the region, and up to 92% in the northern part of the region.

The third part of the doctoral thesis describes an optimised ensemble built over simulations of six models for olive pollen over Europe in 2014. The optimization procedure included observations of previous days and was shown to noticeably improve the accuracy of the pollen forecasts generated by the individual models and simple ensembles (mean and median) built over their predictions.

The result of doctoral thesis demonstrates the possibility to predict the amount of pollen in the air at different temporal and spatial scales using historical and forecasted meteorological information and past-time pollen counts. Forecasts are important for allergic people, as well as for agricultural purposes (potential crop production), and in phenological research. Practically applicable methodologies were constructed for the regional seasonal pollen index predictions, daily pollen forecasts in Riga, and European-scale ensemble fusion.

LIST OF ABBREVIATIONS

CAMS – Copernicus Atmospheric Monitoring Service
EAS – European Aerobiology Society
ECMWF – European Centre for Medium-Range Weather Forecast
IAA – International Association for Aerobiology
MACC – Monitoring Atmospheric Composition & Climate
SILAM – System for Integrated Modelling of Atmospheric Composition
WHO – World Health Organisation

DD – degree day – the unit of Heat Sum
H – The amount of accumulated heat, Heat Sum
SPI – Seasonal Pollen Index – integral of pollen concentrations over the season, numerically equal to the sum of daily pollen concentration during the pollen season

Season – refers to the period associated with the presence of pollen in the air, originating partly from local sources, and partly as a result of pollen long range transport

Predictor – input parameter for statistical analysis influencing the value of predictant
Predictant – variable to be predicted

CONTENTS

SATURS

INTRODUCTION

Studies of pollen and spores in the air have been carried out in Europe since the second half of XX century. The motivation of the studies was associated with several factors: (i) allergenicity of many pollen and spores (Newson et al., 2014; Ring et al., 2012); (ii) phenological and agricultural needs (Aguilera and Ruiz-Valenzuela, 2014; Orlandi et al., 2005).

During the last 30 years, the prevalence of airborne allergy and asthma in Europe has increased fourfold, reaching 15–40% of population. The World Health Organization underlined the problem of continuously increasing number of people suffering from respiratory disease caused by pollen and spores (Huynen et al., 2003). According to the European Federation of Allergy and Airway Diseases Patients Associations, 80 million (24.4%) of adults living in Europe are allergic; the allergy prevalence in children is 30–40% and increasing (Laatikainen et al., 2011; Rönmark et al., 2009). Socio-economic impact of these allergies was emphasized by European Academy of Allergy and Clinical Immunology (Muraro and et al., 2015).

The cost of asthma in Europe is estimated at € 33.9 billion a year, the productivity loss due to poor asthma control is € 14.4 billion a year (ERS, White Book, 2013, http://www.erswhitebook.org).

Asthma and allergy prevalence are among the major public concerns, with 87% of Europeans seeing it as a serious or very serious problem – on par with cardiovascular and respiratory diseases. Pollen allergy, being a widespread cause of quality of life deterioration, has one of the strongest potentials for citizens actively participating in their health management and reduction of acute cases. Several studies, such as ISAAC and ECRHS, (Pearce et al., 2007; Sunyer et al., 2004), show large variations among the European countries (Laatikainen et al., 2011). Among the possible reasons of the contrast are variations in the pollen features, differences in lifestyle and in susceptibility due to co-exposure to chemical and aerosol pollutants.

The burden of allergic disease is recognised by a Written Declaration of the European Parliament (Declaration, 2014). The Declaration has been followed by the creation of the Allergy Interest Group of European Parliament. Pollen forecasting has been taken as a new breakthrough development in Copernicus Atmospheric Monitoring Service (CAMS).

Topicality of the study

The allergic reactions, even those caused by natural allergens, such pollen and spores, can be noticeably reduced by forehanded treatment, which needs to be initiated before the before the appearance of allergens in the air. That brings about the challenge of allergenic pollen forecasts.

There are around 300 monitoring sites in Europe, providing regular observational and forecasted data on pollen and spore concentration in the air (Figure 1.1). Being of vital importance for mere existence of aerobiological forecasts, the observations are available with a delay of 1–2 weeks or more, which makes their direct usage for the model-based forecasting highly problematic. Automatic real-time pollen monitors, potentially capable of revolutionising aerobiology, are still too expensive for massive deployment. Therefore, the main attention currently is dedicated to forecasting models that do not use observations in daily routine, being only calibrated and evaluated against them in offline mode.

Two types of forecasting models are the most popular: regional-to-continental dispersion models and local-scale statistical models. Dispersion models (Helbig et al., 2004; Prank et al., 2013; Sofiev et al., 2015, 2012, 2006; Zink et al., 2013, 2012) are capable of predicting the pollen distribution over large areas but their accuracy strongly varies in space and depends on available information on plant distribution (Siljamo et al., 2012; Sofiev et al., 2015).

The local-scale statistical models exploit empirically established relations between the predicted quantity (predictant, such as pollen concentration) and independent predictors (meteorological factors and historical pollen concentrations) (Rodriguez-Rajo, 2000, as referred by Castellano-Méndez et al., 2005). The ways of establishing these relations vary widely (see Paper I and references therein).

One of the most-important parameters quantifying the strength of an allergenic pollen season is a Seasonal Pollen Index, SPI, which is defined as a sum of all daily-mean pollen concentrations, i.e. a season-long integral of pollen concentrations. It was related to:

a. severity of human allergy (Bastl et al., 2016; D'Amato et al., 2007; Huynen et al., 2003);
b. used as an indicator of the productivity of trees, such as olives (Galán et al., 2014; Myszkowska, 2013; Orlandi et al., 2005a; Oteros et al., 2013b; Prasad et al., 1999);
c. used as predictive parameter for grape and wine (Cunha and Ribeiro, 2015) or olives (Dhiab et al., 2016) production;
d. used as a bio indicator of plant reaction to the on-going climate change (Hatfield and Prueger, 2015; Hedhly et al., 2009; Storkey et al., 2014; Zhang et al., 2014);
e. used in numerous pollen forecasting models as a scaling factor determining the predicted pollen concentrations (Helbig et al., 2004; Prank et al., 2013; Puc, 2012; Ranta et al., 2008; Ritenberga et al., 2016; Siljamo et al., 2012; Sofiev et al., 2012; Stach et al., 2008; Toro et al., 1998; Veriankaitė et al., 2009; Zhang et al., 2013; Ziello et al., 2012).

The SPI is known to change substantially from year to year depending on combination of meteorological factors and physiology of the plant (Masaka, 2001; Ranta and Satri, 2007), see also a review of Dahl et al. (2013). Its prediction, therefore, is crucial for accurate intra-seasonal forecasts.

Scientific novelty

The thesis develops missing components of an over-arching model for pollen season, at all scales from local to continental (Figure 1).

The thesis is built along three lines bringing the following innovative elements –

1. For local *intra*-seasonal short-term predictions of pollen concentrations:
   a. A novel methodology of constructing linear regression models accounting for non-stationarity, non-normality and non-linearity of the problems was developed for pollen based on approaches suggested for local AQ now-casting;
   b. This methodology has been applied to Latvian pollen observational dataset constructing and evaluating the first short-term pollen model for Riga.
2. For regional *inter*-seasonal pollen load forecasting:
   a. A novel methodology of constructing models predicting the next-year SPI for large regions is developed further expanding the principles of non-linear data transformations suggested for intra-seasonal model;
   b. Efficiency of the methodology is demonstrated by constructing the first SPI-predicting model universally applicable over Fennoscandia and Baltic States.

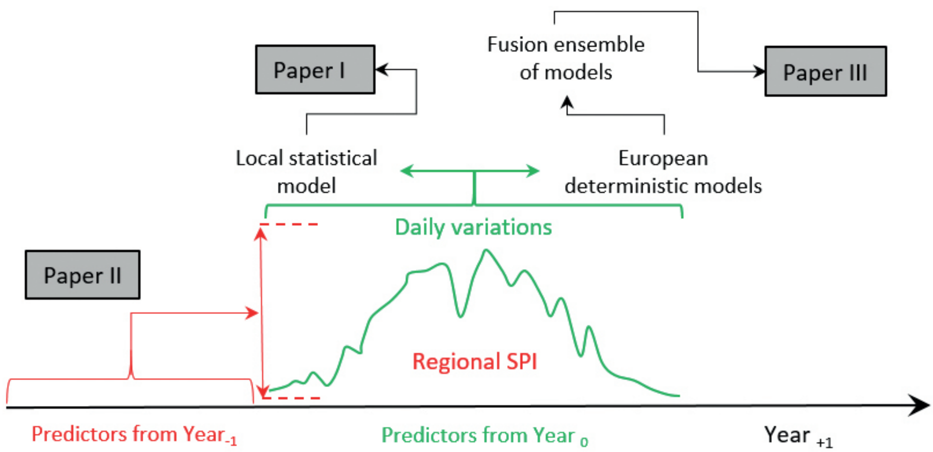


Figure 1. Components, spatial and temporal scales of pollen forecasting models

3. For European-scale ensemble of deterministic pollen forecasts:
   a. For the first time, the data-fusion ensemble treatment, which creates the optimal-ensemble model using the observations over preceding days for optimal combination of the ensemble members, is suggested and evaluated.

Hypotheses of the thesis

Along the above three research directions, the following working hypotheses were formulated:

1. For the local-scale intra-seasonal model, it was suggested that poor performance of most local statistical models constructed for pollen this-far originates from violation of the basic assumptions behind the linear regression methods applied in

practically all such models. A series of non-linear transformations applied to input data should ensure the consistency of the data features and requirements of the statistical procedures used for the model construction.

2. For the inter-season SPI model, we assume that the SPI is a regional parameter determined by the synoptic-scale meteorological processes, i.e. a few hundreds of kilometres. The corresponding temporal scale from several days up to 1–2 weeks characterizes the maximum temporal resolution of input data. Absolute values of the SPI are not important: spatial and temporal variations inside these scales are separable and can be homogenized via normalisation. It should thus be possible to construct a statistical model for the SPI variation over such regions taken as "boxes", i.e. not resolving individual stations.
3. For European-scale fusion ensemble of deterministic forecasting, it was assumed that lasting improvement of the ensemble forecasting skills can be obtained via fusion methods. Unlike the standard data assimilation, which loses its skills within a few hours, the past-time model weighting coefficients should maintain applicability over several days in the future with minor losses of efficiency.

Aim of the thesis

To develop universal and easy-to-use forecasting models with a high accuracy for predicting the concentration of different type of pollen on local-to-European spatial scales and temporal forecasting horizon spanning from days for *intra*-seasonal and months for *inter*-seasonal predictions.

Tasks of the thesis

The above main aim of the study was decomposed into the following tasks:

1. To define meteorological and environmental parameters influencing short and long – term changes of pollen concentration.
2. To construct a local forecasting model that depicts the seasonal parameters of the pollen (i.e., beginning, end, seasonal variations) based on meteorological observations and short-term meteorological forecast data.
3. To construct regional model for next-year SPI based on meteorological and pollen data from current year.
4. To find the best combination of models for olive pollen concentration forecast in South-European region.

Publications

The doctoral thesis consists of three consecutive articles, reflecting the essence of the problem and offering a methodology for solving problems at different levels – local (Paper I) and regional (Paper II and Paper III), by applying statistical (Paper I, Paper II, Paper III) and deterministic (Paper III) modelling methods. Sequential original publications have been published or accepted for publication in high-ranking journals with a JIF from 4.6 to 5.7 (2016) during the last 2 years (2015–2017):

1. **Paper I.** Ritenberga, O., Sofiev, M., Kirillova, V., Kalnina, L., Genikhovich, E. (2016). Statistical modelling of non-stationary processes of atmospheric pollution from natural sources: example of birch pollen. *Agricultural and Forest Meteorology,* 226–227, 96–107.
2. **Paper II.** Ritenberga, O., Sofiev, M., Siljamo, P., Saarto, A., Dahl, A., Ekebom, A., Šauliene, 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. *Science of the Total Environment* (Accepted).
3. **Paper III.** Sofiev, M., Ritenberga, O., Siljamo, P., Albertini, R., Arteta, J., Belmonte, J., Bonini, M., Damialis, T., Elbern, H., Friese, E., Galan, C., Hrga, I., Kouznetsov, R., Plu, M., Prank, M., Robertson, L., Selenc, S., Thibaudon, M., Segers, A., Stepanovich, B., Valdebentino, A.M., Vira, J., Vokou, D. (2017). Multi-model ensemble simulations of olive pollen distribution in Europe in 2014; current status and outlook. *Atmospheric Chemistry and Physics* (Accepted).

Author's contribution

All the tasks regarding the **Paper I**, including (i) pollen monitoring for data collection (years 2006–2016); (ii) sample processing and data analysis; (iii) model construction and evaluation were performed by the author of the current thesis. The author's contribution of **Paper II** consists of: (i) pollen monitoring for Riga site; (ii) data analysis and regional model construction and evaluation. As to the third part of the study (**Paper III**) dedicated to assessing the performance of the existing model ensemble based on the olive pollen season in 2014, the author's contribution is construction and evaluation of statistical fusion model for the ensemble.

Pollen monitoring has been performed under the supervision of Assoc. Prof. Laimdota Kalniņa, modelling – under the supervision of Prof. M. Sofiev (Finnish Meteorological Institute) and the consultations with Prof. Y. Genikhovich (Voeikov Main Geophysical Observatory).

Approbation of the results

Approbation of constructed models has been performed on the European level by Copernicus Atmospheric Monitoring Service, using methodology from constructed regional model for Seasonal Pollen Index (Paper II) forecast, as well as model ensemble (Paper III) as a tool for increase of forecast accuracy. It is planned to use local model for pollen forecast in Riga, as well as to adopt the local model output as an input for regional deterministic models (i.e., SILAM) as a substitute for the non-existing real-time observational data resolving the problem of observational delay.

Related articles

Three related articles were published in journals with JIF of over 2,5 (2016):

1. Sofiev, M., Berger, U., Prank, M., 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., and Peuch, V.-H. 2015. MACC regional multi-model ensemble simulations of birch pollen dispersion in Europe, Atmos. Chem. Phys., 15, 8115–8130. DOI:10.5194/acp-15-8115-2015.
2. Kasprzyk, I., Rodinkova, V., Šauliene, I., **Ritenberga, O.,** Grinn-Gofran, A., et al. 2015. *Air pollution by allergenic spores of the genus Alternaria in the air of central and Eastern Europe*. In: Environmental Science and Pollution Research. DOI 10.1007/s11356-014-4070-6.
3. Sikoparija, B., Skjøth, C.A., Celenk, S., Testoni, C., Abramidze, T., Alm Kübler, K., Belmonte, J., Berger, U., Bonini, M., Charalampopoulos, A., Damialis, A., Clot, B., Dahl, Å., de Weger, L., Gehrig, R., Hendrickx, M., Hoebeke, L., Ianovici, N., Seliger, A.K., Magyar, D., Mányoki, G., Milkovska, S., Myszkowska, D., Páldy, A., Pashley, C.H., Rasmussen, K., **Ritenberga, O.,** Rodinkova, V., Rybníček, O., Shalaboda, V., Šaulienė, I., Ščevková, J., Stjepanović, B., Thibaudon, M., Verstraeten, C., Vokou, D., Yankova, R., Smith, M. 2016. *Spatial and temporal variations in* air*borne Ambrosia pollen across Europe.* In: Aerobiologia. DOI:10.1007/s10453-016-9463-1.

The total number of articles in international and local scientific journals is 11.

Selected international conference proceedings

1. **Ritenberga, O.**, Sofiev, M. Forecasting of *inter-* annual variability of Olive seasonal pollen load. *Mediterranean Palynology Symposium* 4–6 September 2017, Barcelona, Spain.
2. **Ritenberga, O.**, Šauliene, I., Berger, U., Sofiev, M. Towards developing Personal Allergy Symptom Forecasting System in Baltic States. *Congress of the European Academy of Allergy and Clinical Immunology* 17–21 June 2017, Helsinki, Finland.
3. **Ritenberga, O.**, Siljamo, P., Sofiev, M. Modelling of intra-seasonal fluctuation and Inter-annual variability of birch pollen concentration: Example of Latvia and Finland. *European Symposium of Aerobiology* 18–22 July 2016, Lyon, France.
4. **Ritenberga, O.**, Kalnina, L. Temporal changes of ragweed (*A. artemisiifolia, A. trifida, A. psilostachya*) pollen concentration in Latvia. *European Symposium of Aerobiology* 18–22 July 2016, Lyon, France.
5. **Ritenberga, O.**, Sofiev, M., Genikhovich, E. Towards developing of short-term statistical model for birch pollen forecast. *European Aerobiology Network and European Aerobiological Society Symposium*, 10–11 November, 2014. Vienna, Austria.

6. **Ritenberga, O.**, Veselova, A., Kalnina, L., Ustupe, L. The result of 10-year aerobiological monitoring of early spring allergenic pollen in Riga (Latvia). 9th *Pollen Monitoring Programme*. Prague. Czech Republic, 26–30 August, 2013.
7. **Ritenberga, O.**, Kalnina, L., Rodinkova, V., Sofiev, M., Vill, M., Korgmaa, V., Shalaboda, V. Regional differences and comparison of observed and modelled birch pollen data. *14th Nordic Aerobiology Society Symposium*. Riga, Latvia, 19–21 August, 2013.
8. Veselova, A., **Ritenberga, O.**, Ustupe, L., Kalnina, L. Influence of meteorological parameters on Alnus and Corylus pollen concentration in Latvia. *14th Nordic Aerobiology Society Symposium*. Riga, Latvia, 19–21 August, 2013.
9. Kalnina, L., **Ritenberga, O.**, Sauliene, I., Sukiene, L., Severova, E. Character of Betulaceae pollen season in Riga and their comparison with Vilnius and Moscow. *14th Nordic Aerobiology Society Symposium*. Riga, Latvia, 19–21 August, 2013.
10. **Ritenberga, O.**, Kalnina, L., Gudovicha, M. Influence of meteorological parameters on Artemisia pollen concentration in Latvia in the period 2003–2011. *5th European Symposium on Aerobiology*. Krakow, Poland, 3–7 September 2012.
11. Kalnina, L., Šauliene, I., **Ritenberga, O.** Variability of grass pollen concentration during the 9 year period (2003–2011). *5th European Symposium on Aerobiology*. Krakow, Poland, 3–7 September 2012.
12. **Ritenberga, O.**, Kalnina, L., Gudovicha, M. Three ragweed species in Latvia. *Second International Ragweed Conference*, Lyon, France, 28–29 March 2012.
13. **Ritenberga, O.**, Kalnina, L. Development of Aerobiological Monitoring in Latvia. *The conference of European Integration and Baltic Sea Region: Diversity and Perspectives*. Riga, Latvia, 26–27 September 2011.
14. **Ritenberga, O.**, Kalnina, L. Seasonal fluctuations of the airborne pollen concentration in Latvia. *Pollen Monitoring Program* (PMP) 8th International Meeting. University of Tartu, Tartu, Estonia, 20–22 May 2011.
15. **Ritenberga, O.** Aerobiology as geographer's field of research. *Next generation insights into geosciences and ecology*. Tartu, Estonia, 12–15 May 2011.
16. **Ritenberga, O.** Airborne pollen concentration in Riga (Latvia) 2003–2010. *53rd International Scientific Conference of Daugavpils University*. Daugavpils, Latvia, 13–15 April 2011.

The total number of international conference proceedings is 26.

European Institutions interested in the pollen research development in Latvia:

Supporting letters from the presidents of EAS – *European Aerobiological Society*, IAA – *International Association on Aerobiology* was received in 2015 and 2017. The author received two grants from European Aerobiological Society (in 2012 and 2016) for the excellent pollen research. The best presentation award was given by Russian Geographical Society in 2016 during the congress of Young Geographers.

A highly interested and powerful institutional user is Copernicus programme, in particular, Copernicus Atmospheric Monitoring Service (CAMS). The study of the paper III is directly inspired by CAMS, whereas the results of Paper II will be used for predicting the next-year SPI map as one of the inputs to the CAMS European pollen forecasts.

Projects during the doctoral thesis preparation

1. "Statistical modelling of non-stationary processes of atmospheric pollution from natural sources: example of birch pollen" (2015) Science-based funding of Latvian Ministry of Education and Science via "*Attraction of Human Resources to Development of Scientific Study in the area of Earth and Environmental Sciences*" programme. 2016. Completed.
2. National research programme "*Sustainable climate policy and effective energotechnological solutions*". 2015–2017. Researcher.
3. Performance-based funding of University of Latvia "Climate change and sustainable use of natural resources" programme. Researcher 2016–2018. Researcher.
4. EU ECMWF Copernicus Atmospheric Monitoring Service *Personal Allergy Symptom Forecasting system (PASYFO*), leader of Latvian group. Ongoing, 2017–2019.

Current doctoral thesis consists of Summary in English and Latvian languages and three sequential scientific publications – parts of the thesis.

Summary of the thesis consists of annotation, introduction, two main chapters, conclusions and list of references.

1. DATA AND METHODS

The modelling of aerobiological (i.e. *aeropalinological*) processes is based on pollen and/or spore bi-hourly, daily or seasonal data analysis. Bi-hourly data was used in the first part of the study (Paper I), daily data in regional study for Southern Europe (Paper III) and, finally, seasonal data was presented in Paper II. There are different geospatial scales used – from the local (Paper I), to the regional (Paper II, III), – and types of analysed pollen – birch and olives.

Pollen from different plants usually is observed during seven month in Latvia – from the end of February (alder, hazel) to the end of September (mugwort, nettle). Within seven months, there are about 35–40 types of pollen in airflows that are mostly from anemophylous plants.

Pollen grains differ by shape, size, composition, aerodynamic properties and effect on human health. The presence of pollen in the air depends on distribution of plant (influences emission) and meteorological situation, which regulates dispersion and accumulation mechanisms.

In the Northern and North-Eastern Europe, the most common cause of pollen allergy is **birch** and grasses (Bastl et al., 2016; D'Amato et al., 2007; Huynen et al., 2003). In the Central part of Europe, about 25% of people are sensitive to ragweed pollen (Sikoparija et al., 2016), while the plant itself is harmful to agriculture and is able to adapt quickly to different growth conditions. **Olive** pollen is one of the most important causes of respiratory allergies in the Southern Europe and Northern Africa (D'Amato et al., 2007). High rates of sensitization to olive pollen have been documented: 44% in Spain and 20% in Portugal (Pereira et al., 2006), 31.8% in Greece (Gioulekas et al., 2004), 21.6% in Turkey (Kalyoncu et al., 1995), and 15% in France (Spieksma, 1990). At the same time olive is one of the most extensive crops and its oil is one of the major economic resources in Southern Europe. The bulk of olive habitation (95% of the total area worldwide) is concentrated in the Mediterranean basin (Barranco et al., 2008).

1.1. Data for model development

1.1.1. Aerobiological data

The thesis is focused on **birch** (Northern Europe) and **olive** pollen (Southern Europe) data analysis and modelling.

Aerobiological data analysis consists of regular, continuous air monitoring during growing season for several years. The longest data series used in the analysis come from Finland (Figure 1.1) for the period from 1974 to 2015. Aerobiological monitoring in Riga was started in 2003, and the length of data series is sufficient (Figure 1.1) to generate short-term local forecast based on bi-hourly and daily pollen concentration data (Paper I).

Data acquisition was carried out in accordance with the requirements developed by data quality control group (Galán et al., 2014; Oteros et al., 2013a), who formulated the recommendations for devices, substances, sample and data analysis.

Aerobiological monitoring was performed by using Hirst type 7-day Burkard pollen-spore trap (Hirst, 1954) (Paper I), whose specifications make at least a one-week delay in observational data. Seven days are needed for data collection and at least one day for manual microscopic analysis of pollen samples. Automatic pollen monitoring trials have begun at several European monitoring stations (Scheifinger et al., 2013), but for the time being, its accuracy is far behind the manual monitoring accuracy. Pollen recognition and counting was performed in UL ĢZZF Quaternary laboratory using Primo Star Light Microscope under × 400 magnification by choosing vertical counting method – 12 vertical traverses (Carinanos et al., 2000) with the distance of 2 mm, thus, covering daily sample of 14×48 mm.

For the regional studies, aerobiological data were taken from the European Aeroallergen Network database (Figure 1.1), in agreement with the representatives of the countries. The requirements for data in case of birch SPI study (Paper II) was at least 11 continuous years of observations. In case of olive ensemble model construction – at least 30 consecutive observation days, where at least 10 days with pollen concentration exceeded 3 pollen/m^3 (Paper III).

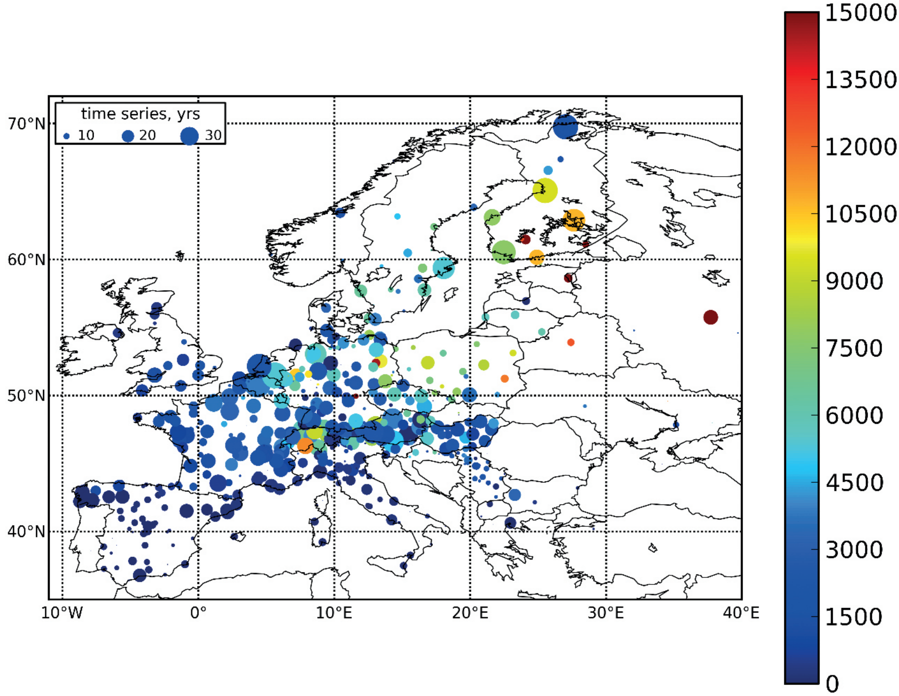


Figure 1.1. EAN monitoring stations and geometric mean of birch SPI [pollen day/m^3] (output of SILAM model)

1.1.2. Meteorological and air quality data

Aerobiological processes have been explained by meteorological (air temperature, short wave solar radiation, wind speed and direction, cloud cover, relative humidity of air, precipitation rate) and CO_2 concentrations in the air.

For local model construction (Paper I) hourly and daily data from Latvian Centre of Geology Environment and Meteorology (LVĢMC) observational station Rīga-Universitāte were used. For the local model verification, data were taken from operational archives of ECMWF.

As the most common factors describing SPI, temperature and precipitation from different time periods are usually mentioned (Dahl and Strandhede, 1996; Latałowa et al., 2002; Yli-Panula et al., 2009). It was decided to use additional related parameters, such as short-wave solar radiation and amount of accumulated heat. 3-hours mean meteorological data were taken from European Re-analysis ERA-Interim data (Dee et al., 2011; Simmons et al., 2010) for the period of 1980–2015.

Carbon dioxide (CO_2) is one of the factors possibly influencing the annual amount of pollen from various taxa (Albertine et al., 2014). Carbon dioxide has a well-expressed positive trend during at least last 40 years, somewhat resembling the trends in the birch pollen abundance. The CO_2 data used for the analysis were downloaded from the NOAA Earth System Research Laboratory (ESRL) public archive (http://www.esrl.noaa.gov/gmd/ccgg/) for the period from 1980 to 2015.

1.2. Basic methods for statistical pollen modelling

Three parts of the pollen forecasting models were defined as separated tasks of pollen forecasting modelling:

1. Determination of the start and end of the pollen season (Paper I, Paper III);
2. Season propagation (Paper I, Paper III), i.e. inter-seasonal daily variation of pollen concentration;
3. Seasonal pollen index calculation (Paper II) – each part is affected by different processes.

Aerobiological processes should be studies in a complex in order not to lose the interactions between different factors. At the same time, at a certain stage of the study they need to be separated (in essence, the model needs to be linearized) for identification of influencing parameters. Following this approach, the influencing factors were identified for (i) productivity (i.e. SPI influencing factors (Paper II); (ii) factors influencing emission and dispersion (Paper I, Paper III); (iii) accumulation-related parameters (Paper I, Paper III). All the aforementioned parameters were identified using multi-linear regression analysis.

The data used for statistical analysis should comply with several requirements: be normally distributed, represent stationary and ergodic process and have linear (or near-linear) dependences between predictor and predictant.

The distribution function of daily mean pollen concentrations is closer to a log-normal than to a normal distribution (Limpert et al., 2008; Toro et al., 1998). The most common method for data transformation applied in the literature is the log-transform of pollen concentrations (Masaka, 2001; Méndez et al., 2005) as a precaution against log-normally distributed data. It was used in case of local birch pollen model and regional SPI model. Non-stationarity was reduced by using normalization and temporal homogenization by replacing the astronomic time-scale with the heat sum scale. Relations between the main predictors and daily mean pollen concentrations are by no means linear. Linearization of dependences was achieved by projection of meteorological variables on pollen scale.

1.2.1. Heat sum calculation

For the calculation the start and the end of pollen season (Paper I, II, III), the idea of G. Linsser (1867) of accumulated heat sum was used. It states that the trees are capable to accumulate the heat, and timing of phenological phases is regulated by the accumulated heat sum. This function in a differential form has been used for construction of birch and olive source terms in the European scale SILAM model (Sofiev et al., 2012):

$$H(d) = \sum_{d=d_s}^{D} [\overline{T(d)} - T_{c-o}]_{+} \quad (1)$$

Here, H is temperature sum (heat sum), d is day, d_s is starting day of the heat accumulation, $T(d)$ is daily temperature, T_{c-o} is cut-off temperature (temperatures below this threshold are not summed up), $[x]_+$ equals 0 for $x<0$ and x for $x\geq 0$ (it excludes the temperatures below the cut-off level).

Equation (1) has two adjustable parameters, which have to be identified for every specific location: cut-off temperature T_{c-o} and start day of accumulation D_s.

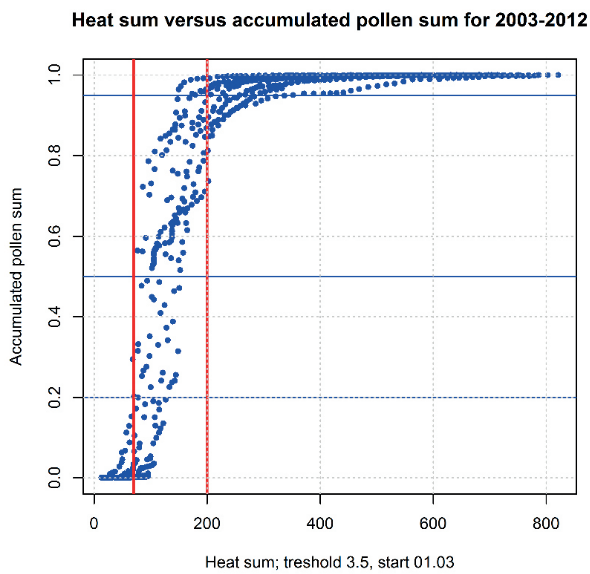


Figure 1.2. Heat sum versus accumulates sum of pollen concentration in Riga, Latvia (Paper I)

The parameters were found simply by testing the different dates – 20.02, 01.03, and 10.03 – and the cut-off temperatures from 0°C to 5°C with a step of 0.5 degrees. The criterion was the smallest standard deviation at three levels of the accumulated pollen sum (a fraction of SPI): 0.2, 0.5 and 0.95 of the SPI for all years (horizontal lines in Figure 1.2).

As a result, the start date of the 1st of March and the cut-off temperature of 3.5°C were identified as the best combination for the heat sum calculation – for Riga. Interestingly, these values are identical to the parameters calculated from phenological data for the Europe-wide birch source term of SILAM.

Having the parameters of the heat sum formula identified, the thresholds for the start and end of the season were estimated from the SPI 5–95% criterion (Nilsson and Persson, 1981): 70 degree day and 200 degree day respectively in Riga (Paper I). Interestingly, these thresholds were not optimal for some of the years, with one of the explanations being the impact of long-distance pollen transport or unusually early or late flowering of birch in the region.

This method with little modifications was used in every part of the thesis.

1.2.2. Temporal homogenization of pollen data

Heat sum was used for the reduction of temporal inhomogeneity as well. One of the problems in aerobiological modelling relates to the temporal variability of perennial data – i.e. the season start, duration and end are all different in different years. Astronomic time is, therefore, an inconvenient variable. It should be replaced by a variable that is clearly associated with the astronomic time but organized so that the start, duration and end of the pollen season would be the same throughout the years. The accumulated amount of heat meets these criteria; it is a monotonous function in a time easily computed from the temperature data. Replacing the temporal scale against the heat scale leads to temporally homogeneous data set.

Several studies have shown that SPI largely depends on the meteorological conditions of the previous year (Introduction of Paper II). Astronomical time sets a significant shift of phenological phases in case of SPI, as well. In order to make comparable year-to-year variations, heat sum was selected as a variable for time axes.

The challenge of the regional consideration is that even expressed via heat-sum, the phenological phases still occur at different moments: the further to the north, the less heat is needed for the season progression (Sofiev et al., 2012). To overcome this difficulty, we express the time scale in % of the total heat accumulated during the whole year, normalized to its long-term mean value, at each station (Paper II). Heat sum differs by up to a factor of two between the aerobiological stations within the region, whereas, e.g., the heat sum threshold for flowering expressed in relative terms is nearly constant.

1.2.3. Reducing of non-stationarity: normalization

Productivity of the plants strongly varies from year to year (Figure 1.3.) depending on several long-term and large-scale parameters, such as the previous-year intensity of flowering, the environmental conditions in winter, pre-seasonal spring conditions, etc. (Dahl and Strandhede, 1996; Linkosalo et al., 2006; Stach et al., 2008).

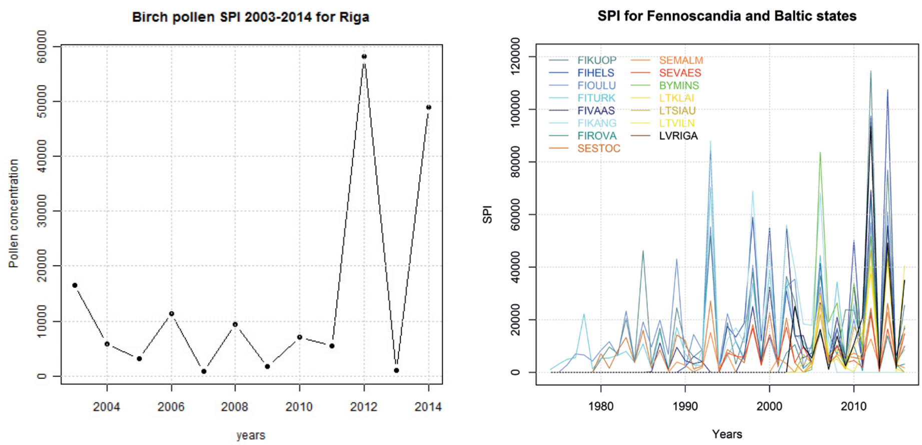


Figure 1.3. Local (left) and regional (right) birch SPI

Characteristic temporal scales of these processes are completely different from the local short-term meteorological and biological processes such as the intensity of flowering the previous year. Such scale separation allows for splitting of the problem: (i) determination of the general flowering intensity characterized by the SPI, (ii) intra-seasonal development of daily mean pollen concentrations.

In practice, these problems can be separated by a simple normalization of the daily concentration $C_i(d)$ with the SPI of the corresponding year, for each considered year i, thus obtaining multi-annual concentration time series $c_i(d)$, which sum up to unity for each season (2).

$$c_i(d) = \frac{C_i(d)}{\sum_{d \in i} C_i(d)} \quad (2)$$

In case of regional model, the SPI_i of each station i for the particular year Y should be normalized with its multi-annual geometric mean $SPI_i^{geomean}$ taken over the whole observed period of the station:

$$SPI_i^{norm}(Y) = \frac{SPI_i(Y)}{SPI_i^{geomean}} \quad (3)$$

thus eliminating the dependence on local birch abundance and making all sites within the region comparable.

2. RESULTS AND DISCUSSION

2.1. Statistical modelling of non-stationary processes – local intra-seasonal forecasts: example of Latvia

Paper I. Ritenberga, O., Sofiev, M., Kirillova, V., Kalnina, L., Genikhovich, E. (2016). Statistical modelling of non-stationary processes of atmospheric pollution from natural sources: example of birch pollen. *Agricultural and Forest Meteorology,* 226–227, 96–107

Local model for short-term daily birch pollen concentration forecast is presented. The model is constructed on the basis of observed meteorological data taken as predictors.

The author's contribution includes data analysis, model construction and verification.

The novelty of the study is twofold:

a. A novel methodology of constructing linear regression models accounting for non-stationarity, non-normality and non-linearity of the problems was developed for pollen based on approaches suggested for local AQ now-casting.
b. This methodology has been applied to Latvian pollen observational dataset, constructing and evaluating the first short-term pollen model for Riga.

The result of data transformation is the significant improvement of efficiency of statistical procedures and increase of the model accuracy. Transformed data sets were used for constructing a model through linear regression. For model application in Riga, nine years of data were used for the model construction and 3 years of data were retained for the model verification. The model evaluation showed an accuracy of over 80% and odds ratio of 30.

The transformations of the input data were performed, as follows:

The mean seasonal propagation of pollen concentration was related to accumulated heat sum, whereas the intra-seasonal concentration variations were associated with the deviations from the mean seasonal curve. Then the dependences between predictors and predictant were linearized by projecting the meteorological data to deviations from the mean seasonal curve of pollen concentration (Figure 2.1). The whole range of the input parameters (relative humidity in Figure 2.1) was split to sub-ranges and for each of those, a mean value of pollen concentrations was computed. The obtained dependence was then used as a lookup table replacing the actual value of relative humidity with the corresponding mean pollen concentration.

Statistically significant ($p > 0.05$) predictors and their weighting coefficients were determined by multilinear regression analysis: Independent (0.03); daily mean temperature of air (0.65), daily mean cloud cover (1.05), sum of daily precipitation (0.53), wind component *u* (0.68).

Model evaluation consisted of two parts:

1. Ability of model to predict seasonal parameters as start, end, propagation.
2. The ability of model to reproduce the above-threshold episodes.

The accuracy of the model was about the same for both learning and control data sets. The overall model accuracy is 5% lower for the control data set, while the OR and other parameters (POD, FAR – described in the annex of Paper I), were even slightly higher. The differences are related to the uncertainty of the model control: the control data set includes 70 days with a concentration above the threshold, about half of the cases were reproduced; 170 days where the concentration was lower than threshold and were reproduced in 97% of cases (Table 2 of Paper I).

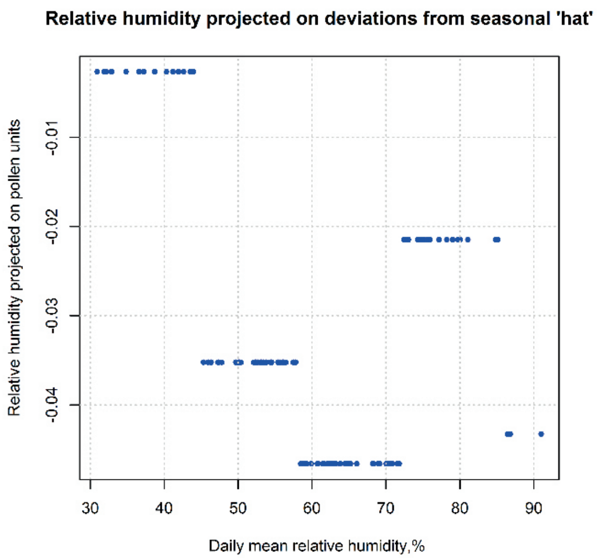


Figure 2.1. Linearization of dependences between relative humidity and deviations from mean seasonal pollen concentration curve

Comparison of the model results with existing models is problematic, as it is the first model for Riga and only a few models in Europe are able to forecast pollen concentrations day-by-day. One can, however, note some important similarities and differences of the new model and other approaches. A quite common practice in short-term forecasting models is to “nudge” them towards observations via auto-regression methods: the forecast is taken as a function of the observed previous-day concentrations, see (Inatsu et al., 2014) for birch and (Stach et al., 2008) for grasses. The current model does not use the nudging because of two reasons: (i) the birch season in Riga is short with very strong day-to-day variability; (ii) pollen data for “today” are never available when the forecast for “tomorrow” is to be generated. However, despite the exclusion of this strong predictor, the temporal correlation of the new model − with all reservations against this parameter − is within the same range from 0.64 to 0.94 as the scores of (Inatsu et al., 2014) − from 0.6 up to 0.9–and (Stach et al., 2008) − from 0.6 to 0.7.

One can also compare the current scores with those of SILAM (Table 3 of Paper I). Expectedly, the new Riga model showed noticeably better performance − as one would expect for the strictly localized statistical development. The difference between the models can be illustrated via their comparison for 2014, one of the control years (Figure 2.2.). Since the models are based on completely different principles, their joint application to the forecasting has a certain potential for further improvement of the forecast accuracy.

Complicated data transformations implemented in the current study is a requirement of statistical methods used for training and evaluation of the model. The normality of the data distribution, stationarity of the time series, ergodicity and near-linear dependencies of the variables are necessary for the model construction – in theory. One can, of course, formally apply the multi-linear regression to non-transformed datasets and still obtain a "technically-working" model with some forecasting skills

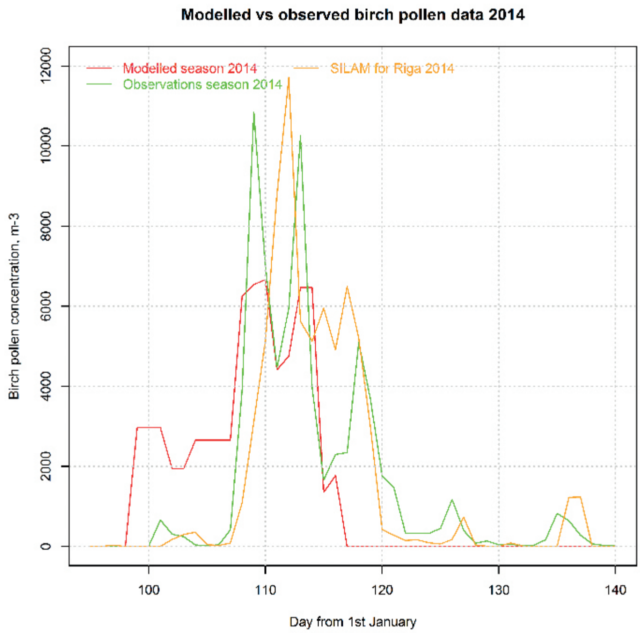


Figure 2.2. Evaluation of accuracy of local and regional (SILAM) models vs. observed birch pollen concentration in 2014

Its performance, however, will be substantially worse, penalized for every transformation step skipped (Table 4 in Paper I). A separate issue is that the inverse transformations, being inevitable for practical use of the forecasts, also introduce errors. Indeed, the regression model is built and optimized in the transformed space of both predictors and predictants. Returning to the physical space for concentrations is a non-linear transformation, which may not preserve the optimality of the solution. Therefore, the scores for the transformed concentrations are higher than for concentrations after the inverse transformations to the physical space. A similar effect was noticed by (Toro et al., 1998).

It was not possible to include the current year SPI calculation in the construction of a local model, so, at the beginning of the season, knowing only a meteorological forecast, the constructed model allows to predict the normalized daily pollen concentrations. In order to de-normalize and express concentrations in absolute values, the development of a separate, regional SPI model was required.

2.2. Statistical modelling of SPI for Northern and North-Eastern Europe

Paper II. Ritenberga, O., Sofiev, M., Siljamo, P., Saarto, A., Dahl, A., Ekebom, A., Šauliene, 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. *Science of the Total Environment* (Accepted)

The paper suggests a methodology for predicting next-year seasonal pollen index (SPI, a sum of daily mean pollen concentrations) over large regions and demonstrates its performance for birch in Northern and North-Eastern Europe. A statistical model is constructed using meteorological, geophysical and biological characteristics of the previous year. We built the model for the northern cluster of stations, which covers Finland, Sweden, Baltic States, part of Belarus, and, probably, Russia and Norway, where the lack of data did not allow for conclusive analysis.

The thesis author's contribution includes data analysis, model construction and verification.

Birch pollen SPI has well expressed bi-annual periodicity (Figure 2.3) (Zink et al., 2013) and a positive trend (Spieksma et al., 2003) noticeable since at least 1974.

It was assumed that the SPI is a regional parameter determined by the synoptic-scale meteorological processes, i.e. a few hundreds of kilometres. It should therefore be possible to identify the regions that react synchronously and demonstrate similar patterns of the SPI year-to-year variations. Moreover, absolute values of the SPI are of essentially no importance: they are decided by vegetation density in proximity to the station, which is a static parameter. Therefore, spatial and temporal variations inside these regions are separable. With the above assumptions, it should be possible to construct a statistical model for the SPI variation over these regions taken as "boxes", i.e. not resolving individual stations but taking each region as a single entity with the normalized SPI averaged over the region.

The primary goal of the study was to build the unified model suitable for applications over large regions, thus demonstrating the feasibility of spatial generalization of the SPI predictions using large-scale meteorological features as the controlling parameters.

The bulk of previous studies concentrated on a single or a few closely-located stations (Introduction of Paper II)

The second principal difference of the suggested approach is that we applied a series of non-linear transformations of the input data and changed the governing variable from time to normalized heat sum, thus eliminating spatial variability of SPI and reducing spatial variability of year-to-year variation after that, construction of the predictive models followed standard procedures for multi-component optimal model fitting. Predictors from the previous year *Y-1* showed a greater influence on the SPI of the year *Y* than those from the year *Y*. Thus, six main influencing parameters were identified:

- precipitation $Pr_{0.13}$ for the period 0.13–0.25 of the annual heat sum accumulation
- temperature $T_{0.13}$ for the period 0.13–0.25, of the annual heat sum accumulation
- temperature $T_{0.25}$ for the period 0.25 to 0.45 of the annual heat sum accumulation
- short wave solar radiation SW_0 for the period from 0 to 0.05 of the annual heat sum accumulation
- SPI_{Y-1}
- $CO_{2\ Y-1}$

The bootstrap included 15 iterations, one for each of 15 station datasets withheld from both regional SPI and regional meteorological averaging. The SPI data and meteorological parameters for the remaining stations were averaged over the region. After that, the multilinear regression was calculated obtaining the model coefficients for the regional time series with one site excluded. The fitting coefficients were averaged across the iterations, finally obtaining the mean regression coefficients and their standard deviation (Table 3 of Paper II).

During the data analysis, it was realized that it is possible to build the model using *only available meteorological data* and the trend-describing variable $CO_{2,\ Y-1}$. This opportunity allows predicting the SPI relative deviation from the long-term mean knowing neither this mean level nor the previous-year SPI. The model becomes detached from the SPI observations and applicable also in places with no aerobiological observations whatsoever.

The regional formula for $\Delta SPI^{reg}(Y)$ of the year *Y* based on meteorological data and data of CO_2 of the previous year *Y-1* is:

$$\Delta SPI^{reg}(Y) = a_0 + a_{CO_2} CO_{2,Y-1} + a_{pr} Pr_{0.13} + a_{sw} sw_0 + a_{T_{0.13}} T_{0.13} + a_{T_{0.25}} T_{0.25} \quad (4)$$

Adding SPI_{Y-1} as a predictor still significantly increases the quality of the model. The final formula includes the same meteorological parameters as the *meteo-only* model, CO_2 and SPI from the previous year

For the regional SPI model evaluation, same parameters as in Paper I were used. (i) Odds Ratio as a measure of differentiating between the high and low seasons, (ii) Model Accuracy and (iii) a fraction of the SPI predictions that fall within the factor 2 from the observations were considered. Behaviour of individual stations inside the region differs somewhat but the regional formula tends to work better for the stations with longer time series, i.e. those, which contributed mostly to the model identification (Turku, Kuopio, etc.). The limited size of the dataset and high scores of the models resulted in just few cases of the predictions being wrong. For several sites, the low-high differentiation appeared always correct during the observed years. The sample OR for such sites is infinitely high (Figure 3a from Paper II).

Since the suggested models are first in their class, direct comparison with published studies is quite difficult. However, certain conclusions can be derived. The most direct comparison is possible with the study of (Ranta and Satri, 2007), who built three local models for Turku, Kuopio and Oulu stations.

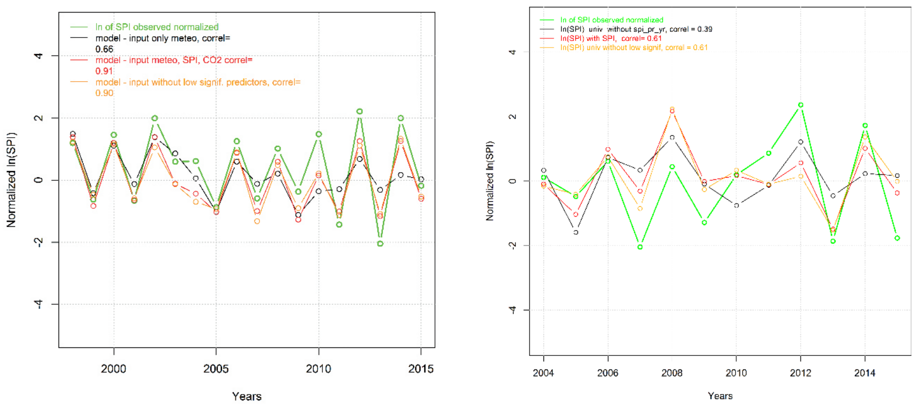


Figure 2.3. Regional SPI model result for Vaasa (left) and Riga (right)

Temporal correlation of these models however trails significantly behind the current unified model: it was 0.59 vs. 0.82 for Turku and 0.61 vs. 0.77 in Oulu. Only in Kuopio, both models scored to 0.65. Similar predicting capacity to that of (Ranta and Satri, 2007) was reported by (Dahl and Strandhede, 1996) but direct comparison is not possible due to sqrt-transformed values reported in that paper. An important difference of that work was that the predictors were taken as heat sums rather than the mean temperature over some period, which potentially can further improve the model's accuracy.

Quite high correlations of the local SPI and several meteorological drivers were reported for Poznan by (Grewling et al., 2012) but no predictive model was built. The authors also pointed out that they found no parameters capable of explaining the break points in the bi-annual cycle. This study has succeeded but Poznan is outside the current region and the developed models are not directly applicable there.

Considering the relative importance of various parameters, one can see that the *meteo-only* model, albeit quite good, trails behind the full-input model, thus suggesting that the biological processes also significantly contribute to the SPI – in agreement with (Dahl and Strandhede, 1996). In particular, the dynamic range of the SPI variability is the largest in the observations, closely followed by the full-input model, whereas the *meteo-only* model is more conservative. One can argue then that the plant response serves as an amplifier for the meteorological signals. This also corroborates with conclusions of the biological model (Dahl et al., 2013) that stressed the importance of the combination of meteorological and biological parameters for adequate next-season prediction.

Finally, the *meteo-only* model reproduces both the bi-annual cycle and the years when it breaks down, also strongly suggesting that such behaviour of birch is at least inspired by the regional-scale weather phenomena. The suggested models were made for a large but still limited region in the north of Europe. Since the climatic conditions and plant response to the stress change gradually along north-south and west-east directions, the models score lower in Lithuania and southern Sweden, which delineate the southern border of the region. Outside the region, the time series are still quite good for the years with pronounced bi-annual cycle but large errors show up when this cycle breaks down. Such "unusual" years also become more frequent, especially in Brussels, where the bi-annual cycle is practically non-existent. It indicates a presence of some un-accounted factors, which control the SPI in temperate climate I, for example, precipitation and winter-time chilling.

2.3. Model ensemble for pollen forecast in Southern Europe

Paper III. Sofiev, M., Ritenberga, O., Siljamo, P., Albertini, R., Arteta, J., Belmonte, J., Bonini, M., Damialis, T., Elbern, H., Friese, E., Galan, C., Hrga, I., Kouznetsov, R., Plu, M., Prank, M., Robertson, L., Selenc, S., Thibaudon, M., Segers, A., Stepanovich, B., Valdebentino, A.M., Vira, J., Vokou, D. (2017). Multi-model ensemble simulations of olive pollen distribution in Europe in 2014; current status and outlook. *Atmospheric Chemistry and Physics* (Accepted)

The thesis author's contribution is construction and evaluation of the statistical optimization procedure and its application to ensemble of CAMS six deterministic models. The performance of the obtained fusion ensemble was compared with simple ensemble treatments and with individual members.

The aim of the study was to present the first Europe-wide ensemble-based evaluation of the olive pollen dispersion during the season of 2014. The first modelling experiment of the European-scale olive pollen dispersion analyses the quality of the predictions and outlines the research needs. A 6-models (EMEP, EURAM-IM, LOTOS-EUROS, MATCH, MOCAGE, SILAM) strong ensemble of Copernicus Atmospheric Monitoring Service (CAMS) was run through the season of 2014 computing the olive pollen distribution. The study followed the approach of the multi-model simulations for birch (Sofiev et al., 2015) with several amendments reflecting the peculiarity of olive pollen distribution in Europe. Further steps towards fusion of model predictions and observations and demonstration of its value in the forecasting regime were made, as well.

One of possible ways to improve the quality of model predictions without direct application of data assimilation is to combine them with observations via ensemble-based data fusion methods (Potempski and Galmarini, 2009). Their efficiency has been demonstrated for air quality problems (Johansson et al., 2015 and references therein) and climatological models (Genikhovich et al., 2010) but the technology has never been applied to pollen.

Three **ENSEMBLE** models were generated by (i) arithmetic average, (ii) median and (iii) optimal combination of the 6 model fields. Averaging and median were taken on hourly basis, whereas optimization was applied at the daily level following the temporal resolution of the observational data. For the current work, we used simple linear combination c_{opt} (5) of the models c_m, $m = 1..M$ minimizing the regularized RMSE J (6) of the optimal field:

$$c_{opt}(i,j,k,t,\tau,A) = a_0(\tau) + \sum_{m=1}^{M} a_m(\tau)\, c_m(i,j,k,t), \quad A = [a_1..a_M], \quad a_m \geq 0\ \forall m \quad (5)$$

$$\begin{aligned} J(t,\tau) = sqrt\left[\frac{1}{O}\sum_{o=1}^{O}\left(c_{opt}(i_o,j_o,k_o,t,\tau,A) - c_o(t)\right)^2\right] + \\ \alpha\sum_{m=1}^{M}\left(a_m(\tau) - \frac{1}{M}\right)^2 + \beta\sum_{m=1}^{M}\left(a_m(\tau-1) - a_m(\tau)\right)^2, \qquad \tau = \{d_{-k}, d_0\} \end{aligned} \quad (6)$$

Here, i, j, k, t are indices along the x, y, z, and time axes, M is the number of models in the ensemble, O is the number of observation stations, $\tau = \{d_{-k}:d_0\}$ is the time period of $k+1$ days covered by the analysis window, starting from d_{-k} until d_0, τ-1 is the previous-day analysis period $\tau\text{-}1 = \{d_{-k-1}:d_{-1}\}$, c_m is concentration of pollen predicted by the model m, c_o is observed pollen concentration, a_m is time-dependent weight coefficient of the model m in the ensemble, a_0 is time-dependent bias correction.

In the Eq. (5), the first term represents the RMSE of the assimilated period τ, the second term limits the departure of the coefficients from the homogeneous weight distribution, the third one limits the speed of evolution of the a_m coefficients in time. The scaling values α and β decide on the strength of regularization imposed by these two terms.

Eq. (6) requires three parameters to prescribe: the regularization scaling parameters α and β, and length of the assimilation window T. For the purposes of the current feasibility study, several values for each of the parameters were tested and the robust performance of the ensemble was confirmed with very modest regularization strength and for all considered lengths of the analysis window – from 1 to 15 days. Finally, $\alpha = 0.1$, $\beta = 0.1$, $T = 5$ *days* were selected.

The ensemble was constructed mimicking the forecasting mode. Firstly, the analysis is made using data from the analysis period τ. The obtained weighting coefficients a_i are used over several days forwards from day d_0: from d_1 until d_{nf}, which constitute the forecasting steps. The performance of the ensemble is evaluated for each length of the forecast, from 1 to n_f days.

The optimized ensemble showed that each of the 6 models had substantial contribution over certain parts of the period prior to and after the main season, concentrations were very low and noisy, so the regularization terms of Eq.(6) took over and pushed the weights to a-priori value of 1/6.

Comparison with other forecasts expectedly shows that the optimized ensemble not only has significantly better skills than any of the individual models, but is up to 25–30% better than mean and median of the ensemble (Figure 2.4.).

A stronger competitor was the “persistence forecast” when the next-day(s) concentrations are predicted to be equal to the last observed daily value. The one-day persistence appeared to be the best-possible “forecast”, which showed at the beginning of May almost twice lower RMSE than the one-day forecast of the optimal ensemble (Figure 2.4).

Strong performance of the one-day persistence forecast is not surprising and, with the current standards of the pollen observations, has no practical value: the data are always late by more than one day (counting can start only the next morning and become available about mid-day).

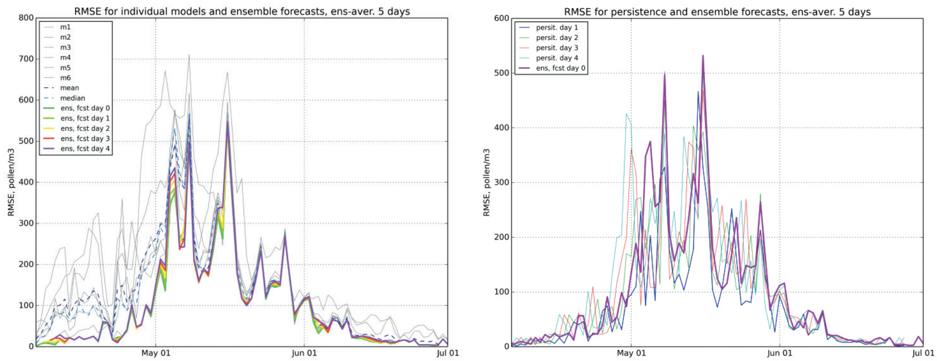


Figure 2.4. RMSE of the of individual models against the ensemble mean (left) against persistence-based forecasts (right)

The second problem of the persistence forecast is that it needs actual data, i.e. the scarcity of pollen network limits its coverage. Thirdly, persistence loses its skills very fast: already day+2 forecast has no superiority to the optimal ensemble, whereas day+3 and +4 persistence-based predictions are useless. Finally, at local scale, state-of-art statistical models can outperform it – see the discussion in (Paper I).

The most-evident issue highlighted by the exercise is the shift of the pollen season in some key regions, which is similar in all the models suggesting some unresolved inconsistencies between the heat-sum calculations of the source term and the features of the temperature predictions by the weather model. The issue suggests some factor(s) currently not included or misinterpreted in the source term. One of the candidate processes is the chilling-sum accumulation suggested by some studies, e.g., (Aguilera et al., 2014). A switch to different types of phenological models with genetic differentiation of the populations following Chuine and Belmonte, (2004) is another promising option. The set of questions refers to the pollen load prediction, i.e. a possibility to forecast the overall season severity before it starts. Several statistical models have been presented in the literature, e.g., (Dhiab et al., 2016) for total annual load and (Chuine and Belmonte, 2004) for relative load. Their evaluation and implementation in the context of dispersion models is important.

The first steps towards ensemble-based fusion of the model forecasts and pollen observations showed a strong positive effect. Further development of these techniques combined with progress towards near-real-time pollen data has a very high potential for improving the forecasts.

CONCLUSIONS

Three studies on statistical pollen modelling were chosen as parts for doctoral thesis – one local, one regional (Northern Europe) and one of European scale.

Each of constructed models are formulated in generic terms, thus, its methodology is applicable for any location and different pollen types. Such extrapolation, however, will require full re-parameterization and applicability verification because statistically established relations may or may not hold outside the area and time period where/when they were identified.

The local intra-seasonal model demonstrated solid performance and stability with the model accuracy MA exceeding 80% and the odds ratio OR = 30.

The inter-annual variability of the regional birch seasonal pollen index is synchronized over large regions of Europe with significant correlations holding at the distances associated with the synoptic spatial scale. This open the possibility to generate comparatively simple linear statistical models valid within such regions.

The best-performing regional SPI predicting model was based on combination of the meteorological, CO_2 data and aerobiological data from the preceding season. All the constructed models successfully reproduced both bi-annual cycle of the SPI and the years when this cycle breaks down. In particular, the model with only meteorological input captured the bi-annual cycles and its breaking years, which highlights the key role of meteorology in formation of this cycle. The dynamic range of the variations is, however, under-stated by this model, pointing out the importance of the plant response to the meteorological stress.

For the large-scale deterministic models, it was shown that their forecasts can be improved via statistical fusion algorithms applied to the multi-model ensemble.

An optimal linear combination of the individual ensemble models showed strong skills, routinely outperforming all individual models and simple ensemble approaches

Two of three constructed models are already in use of Copernicus Atmospheric Monitoring Service.

Future challenges

A specific future challenge is to construct the European model for the SPI predictions. At the current stage, there are two possibilities for constructing it: (i) increase the number of parameters and allow for strong non-linearities in the dependencies (possible, as shown in Paper I), (ii) construct different model(s) for Central and Eastern Europe as it was done, for instance, for the olive season timing forecasting by (Aguilera et al., 2014). Each approach has its advantages and drawbacks. The first approach, being clearly preferable from the user standpoint, faces a wide variety of interplay of the governing parameters, which have to be identified and quantified, expansion will require different treatment of meteorology, since the area will be much larger than the synoptic scale, i.e. the homogenity assumption with regard to the meteorological input will no longer be valid. The multi-model approach, albeit being easier in each region, leads to discontinuities of the parameterizations at the borders of the delineated areas, where also none of the models is good.

One of the strengths of continental-scale dispersion models is their ability to predict long-range transport events. However, local models are usually more accurate at their locations than the European-scale dispersion models. Therefore, combination of large-scale deterministic and regional/local statistical models may prove a good way to increase the overall forecast accuracy. Specific ways of combining these approaches can vary but separation of the scales between them (local-to-regional for statistical models and regional-to-continental for transport models) is probably the main watershed. Technologically, the link can be implemented via data assimilation into the transport models or via fusion technique as done in the Paper III. The strengths and weaknesses of each approach are not fully understood yet.

ACKNOWLEDGEMENTS

I would like to thank my family and friends – especially Mama Yelena for encouragement and support, and just for who I am. I want to thank my son Robert for understanding and patience, my brothers (Anton, Pavel, Robert, Nickolay) for the support and motivation. I am very thankful to my cousin Nickolay Chashnikov for support and technical assistance.

I am expressing my deepest gratitude to the excellent supervisor, Professor Mikhail Sofiev of Finnish Meteorological Institute for inspiration, care, responsiveness and invaluable support. I am very grateful to Professor Eugene L. Genikhovich of Voeikov's Main Geophysical Observatory, for advice, kindness and support.

I would like to thank the academic staff, colleagues and friends from University of Latvia – especially Assoc. Prof. Laimdota Kalniņa for showing me this direction, for useful advice and sincerity for more than 10 years. Special thanks to my colleague Antra Dūle for her understanding and responsiveness.

LATVIJAS UNIVERSITĀTE
ĢEOGRĀFIJAS UN ZEMES ZINĀTŅU FAKULTĀTE

Olga Ritenberga

ĢEOTELPISKO UN TEMPORĀLO PUTEKŠŅU SEZONAS IZMAIŅU PROGNOZĒŠANA EIROPĀ AR STATISTISKO UN DETERMINISTISKO MODELĒŠANU

PROMOCIJAS DARBS

Doktora grāda iegūšanai ģeogrāfijas nozarē
dabas ģeogrāfijas apakšnozarē

Rīga 2017

Promocijas darbs izstrādāts Latvijas Universitātē Ģeogrāfijas un Zemes zinātņu fakultātē Ģeogrāfijas nodaļā un Somijas Meteoroloģijas institūtā laika posmā no 2010. gada līdz 2012. gadam un no 2014. gada līdz 2017. gadam.

Promocijas darbs tika izstrādāts ar VPP "Ilgtspējīga klimata politika un inovatīvi, energoefektīvi tehnoloģiski risinājumi (KPIET)" un projekta „Energoefektīvi un oglekļa mazietilpīgi risinājumi drošai, ilgtspējīgai un klimata mainību mazinošai energoapgādei" (LATENERGI) atbalstu.

Darba forma: publikāciju kopa ģeogrāfijas nozarē, dabas ģeogrāfijas apakšnozarē.

Promocijas darba zinātniskie vadītāji:
Profesors *Doc. phys., Dr. sci. Tech.* **Mihails Sofievs (*Mikhail Sofiev*)**, Somijas Meteoroloģijas institūts
Asoc. profesore *Dr. geogr.* **Laimdota Kalniņa**, Latvijas Universitāte

Darba recenzenti:
Dr. phys. **Jānis Kleperis**, Cietvielu fizikas institūts, Rīga, Latvija
Dr. geogr. **Agrita Briede**, Latvijas Universitāte, Rīga, Latvija
PhD **Lukasz Grewling**, Adama Mickeviča Universitāte, Poznaņa, Polija

LU Ģeogrāfijas nozares promocijas padomes sastāvs:
Agrita Briede, *Dr. geogr.* – promocijas padomes priekšsēdētāja
Oļģerts Nikodemus, *Dr. geogr.*
Raimonds Kasparinskis, *Dr. geogr.*
Māris Kļaviņš, *Dr. habil. chem.*
Viesturs Melecis, *Dr. biol.*
Solvita Rūsiņa, *Dr. geogr.* – promocijas padomes sekretāre
Iveta Šteinberga, *Dr. geogr.*

Promocijas darba aizstāvēšana notiks 2017 gada 29. decembrī plkst. 13.00 Latvijas Universitātes Ģeogrāfijas nozares promocijas padomes atklātā sēdē Latvijas Universitātes Ģeogrāfijas un Zemes zinātņu fakultātē Rīgā, Jelgavas ielā 1.

Ar promocijas darbu un tā kopsavilkumu var iepazīties Latvijas Universitātes Zinātniskajā bibliotēkā Rīgā, Raiņa bulvārī 19.

Atsauksmes sūtīt promocijas padomes sekretārei *Dr. geogr*. Solvitai Rūsiņai:
Ģeogrāfijas un Zemes zinātņu fakultāte, Latvijas Universitāte, Raiņa bulvāris 19, LV-1586, Rīga, Latvija. E-pasts: rusina@lu.lv

ISBN 978-9934-18-294-5

ANOTĀCIJA

Promocijas darbs ir pētījums par temporālo un ģeotelpisko putekšņu sezonas prognozēšanu (ieskaitot gaitu, ietekmējošos faktorus un putekšņu diennakts koncentrāciju), izmantojot sarežģītu datu transformāciju un vienkāršus statistiskus paņēmienus modeļa veidošanā un validācijā. Modelēšana tika veikta vairākiem putekšņu veidiem un dažādos telpiskos mērogos (lokālā un reģionālā). Pētījuma rezultāti atspoguļoti trīs rakstos augsti indeksētos (JIF > 4,6) starptautiskos žurnālos.

Pētījuma sākumā pēc 12 gadu diennakts datiem tika izveidots lokāls modelis Rīgai (Raksts I), lai, izmantojot meteoroloģiskās prognozes datus, noteiktu bērza putekšņu koncentrāciju gaisā. Izveidotā lokālā modeļa precizitāte pārsniedz 80%.

Otrs pētījums (Raksts II) izveidots pēc reģionālajiem datiem – iekļaujot Somijas, Zviedrijas, Lietuvas, Latvijas, Baltkrievijas, daļēji Krievijas un Norvēģijas datus –, lai veiktu nākamā gada sezonālā putekšņu indeksa (SPI) prognozēšanu. SPI modelis ir veidots kā universāls visam reģionam, un tā precizitāte ir robežās no 65% (DA daļā) līdz 92% (reģiona ziemeļu daļā).

Promocijas darba trešā daļa veltīta modeļu ansambļa izveidei ar Eiropas pārklājumu (Raksts III) uz olīvkoka putekšņu bāzes. Pierādīts, ka sešu esošo deterministisko CAMS modeļu ansamblis darbojas ievērojami labāk par katru atsevišķu modeli.

Secināts, ka ir iespējams prognozēt putekšņu daudzumu gaisā dažādos temporālos mērogos – pēc noteiktā gada meteoroloģiskās situācijas aprēķināt nākamā gada sezonālo putekšņu indeksu un pēc lokālās meteoroloģiskās prognozes datiem prognozēt diennakts putekšņu koncentrāciju noteiktā vietā. Prognožu rezultāti izmantojami, lai risinātu putekšņu alerģijas problēmu, lai noteiktu ražas apjomus lauksaimniecībā, un augu fenoloģijas pētījumiem.

SAĪSINĀJUMU UN PASKAIDROJUMU SARAKSTS

CAMS – Kopernika Atmosfēras monitoringa serviss (*Copernicus Atmosphere Monitoring Service*)

EAS – Eiropas Aerobioloģijas biedrība (*European Aerobiology Society*)

ECMWF – Eiropas Vidējā mēroga prognožu centrs (*European Centre for Medium-Range Weather Forecast*)

IAA – Starptautiskā Aerobiologu asociācija (*International Association for Aerobiology*)

MACC – atmosfēras sastāva un klimata monitorings (*Monitoring Atmospheric Composition & Climate*)

SILAM – swistēma atmosfēras sastāva integrētai modelēšanai (*System for Integrated modeLling of Atmospheric coMposition*)

WHO – Pasaules Veselības organizācija (*World Health Organisation*)

DD – grādu diena (*degree day*) – uzkrātā siltuma daudzuma (*Heat Sum*) mērvienība

H – uzkrātais siltuma daudzums (*Heat Sum*)

SPI – sezonālais putekšņu indekss (*Seasonal Pollen Index*) – putekšņu diennakts koncentrāciju summa gadā

Sezona – promocijas darbā – periods, kas saistīts ar putekšņu klātbūtni gaisa plūsmās, kur tie nonāk daļēji lokālās ziedēšanas, daļēji putekšņu tālās pārneses rezultātā

Prediktors – ietekmējošais faktors, kas nosaka prediktanta vērtības

Prediktants – prognozējamais mainīgais, ko raksturo prediktori

IEVADS

Sākot ar 20. gs. otro pusi, Eiropā aktīvi tiek veikti putekšņu un sēņu sporu pētījumi gaisa plūsmās. Pētījumu iemesls ir saistīts ar to, ka: (I) putekšņi un sporas izraisa elpceļu alerģiju (Kasprzyk et al., 2015; Newson et al., 2014; Ring et al., 2012); (II) pētījumu rezultāti ir nozīmīgi lauksaimniecībā un fenoloģijā (Aguilera and Ruiz-Valenzuela, 2014; Orlandi et al., 2005b).

Pasaules Veselības organizācija (WHO) uzsver, ka pēdējo 30 gadu laikā cilvēku skaits, kas pakļauti putekšņu alerģijai un astmai, pieaudzis četras reizes un veido 15–40% no Eiropas iedzīvotāju skaita (Huynen et al., 2003). Pēc Eiropas Alerģiju un elpceļu slimību pacientu asociācijas (AADPA) datiem, 80 milj. (24,4%) pieaugušo Eiropā ir alerģiski, bet ap 30–40% bērnu alerģijai ir tendence palielināties (Laatikainen et al., 2011; Rönmark et al., 2009). Eiropas Alerģijas un klīniskās imunoloģijas akadēmija (EAACI) uzsver alerģijas sociālekonomisko ietekmi (Muraro and et al., 2015). Alerģijas un astmas izmaksas Eiropā veido 33,9 miljardus eiro gadā (ERS, White Book, 2013, http://www.erswhitebook.org).

Ņemot vērā, ka astma un alerģija ievērojami pasliktina dzīves kvalitāti, problēma ir jārisina starptautiskā līmenī. Alerģijas problēma tika atzīta Eiropas Parlamentā, par ko liecina rakstiskā deklarācija, Eiropas Parlamenta Alerģijas interešu grupas izveide (Molnár et al., 2015) un putekšņu prognozēšanas iekļaušana Kopernika Atmosfēras monitoringa servisā (CAMS).

Pētījuma aktualitāte

Cilvēka organisma alerģiskās reakcijas var ievērojami samazināt, veicot laicīgu profilaktisku ārstēšanu, kas ir nepieciešama vismaz divas nedēļas pirms alergēnu parādīšanās gaisā. Tas ir izaicinājums putekšņu koncentrācijas prognostiskajai modelēšanai.

Eiropā ir ap 300 monitoringa vietu, kas regulāri sniedz informāciju par putekšņu koncentrāciju gaisā (1.1. attēls). Aerobioloģisko novērojumu dati ir pieejami ar 1–2 nedēļu nobīdi, kas padara problemātisku datu tiešu izmantošanu prognozēšanas vajadzībām. Automātiskais reālā laika putekšņu monitorings eventuāli varētu ieviest globālas izmaiņas aerobioloģijas pētījumos, bet pagaidām tā precizitāte ievērojami atpaliek no manuālās un ierīču izmaksas ir pārāk lielas, lai nodrošinātu monitoringa tīklu visā Eiropā. Līdz ar to pašlaik uzmanība tiek pievērsta tiem prognozēšanas modeļiem, kuros novērojumu dati tiek izmantoti tikai kalibrācijai un modeļa precizitātes novērtēšanai bezsaistes režīmā.

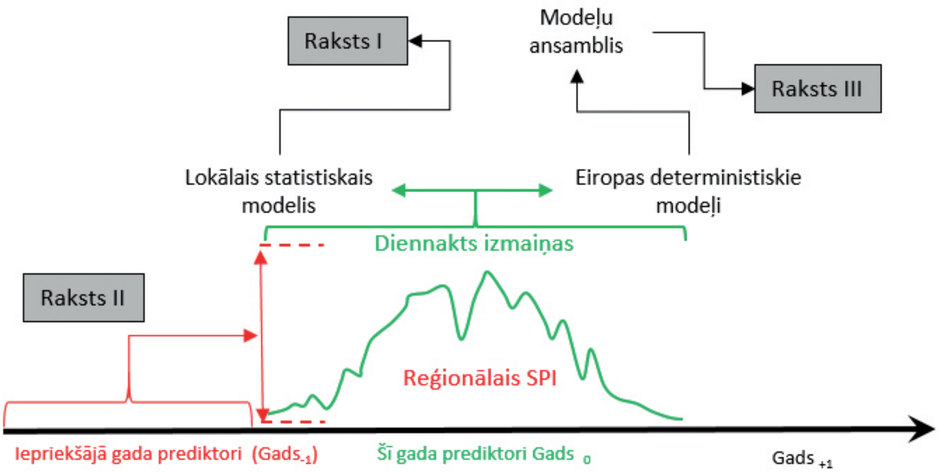
Putekšņu pētījumos pārsvarā tiek izmantoti divu veidu prognozēšanas modeļi: reģionālie – kontinentālie dispersijas modeļi un vietējie (lokālie) statistikas modeļi. Dispersijas modeļi (Helbig et al., 2004; Prank et al., 2013; Sofiev et al., 2015, 2012, 2006; Zink et al., 2013, 2012) ir spējīgi prognozēt putekšņu izkliedi lielās teritorijās, bet to precizitāte ir mainīga un ļoti atkarīga no informācijas par auga (t. i., emisijas avota) izplatību (Siljamo et al., 2012; Sofiev et al., 2015). Vietējā mēroga statistiskie modeļi izmanto empīriski noteiktās sakarības starp prediktantu (piemēram, putekšņu koncentrāciju) un neatkarīgajiem predikotoriem (meteoroloģiskajiem, vides faktoriem un iepriekšējo gadu putekšņu informāciju) (Rodriguez-Rajo, 2000, pēc (Castellano-Méndez et al., 2005)). Minēto sakarību noteikšanas metodes ir dažādas, ko labi atspoguļo publikācijas (sk. Rakstu I un tā bibliogrāfiju).

Viens no parametriem, kas kvantitatīvi nosaka alerģisko putekšņu sezonas stiprumu, ir sezonālais putekšņu indekss (SPI), kas tiek definēts kā diennakts vidējās putekšņu koncentrācijas sezonālā summa, t. i., putekšņu koncentrācijas integrāls sezonas garumā. SPI ir saistīts ar alerģisko reakciju stiprumu (Bastl et al., 2016; D'Amato et al., 2007; Huynen et al., 2003) un tiek izmantots kā indikators augļu koku (piemēram, olīvkoku) produktivitātes noteikšanai (Dhiab et al., 2016; Orlandi et al., 2005a; Oteros et al., 2013; Prasad et al., 1999), kā vīnogu (līdz ar to vīna) produkcijas prognozēšanas parametrs (Cunha and Ribeiro, 2015) un kā bioloģiskais indikators auga reakcijai uz klimata pārmaiņām (Hatfield and Prueger, 2015; Hedhly et al., 2009; Storkey et al., 2014). Putekšņu prognozēšanas modeļos SPI tiek izmantots kā mērogošanas koeficients, jo tas raksturo kopējo prognozēto putekšņu daudzumu (Prank et al., 2013; Puc, 2012; Ranta et al., 2008; Ritenberga et al., 2016; Siljamo et al., 2012; Sofiev et al., 2012; Stach et al., 2008; Toro et al., 1998; Veriankaitė et al., 2009; Ziello et al., 2012).

SPI vērtības ievērojami mainās gadu no gada atkarībā no meteoroloģisko faktoru kombinācijas un auga fizioloģijas (Dahl et al., 2013; Masaka, 2001; Ranta and Satri, 2007), tādēļ tā prognozēšanai ir izšķiroša loma intrasezonālo putekšņu prognozēšanas modeļu veidošanā.

Zinātniskā novitāte

Promocijas darbā tika definētas un izveidotas putekšņu koncentrācijas prognostisko modeļu trūkstošās komponentes no vietējā līdz reģionālajam līmenim (1. attēls).



1. attēls. Prognostisko putekšņu modeļu telpiskā un temporālā mēroga komponentes

Promocijas darbs sastāv no trīs daļām un apvieno šādus novatoriskus elementus:

1. Vietējais intrasezonālais putekšņu koncentrācijas īslaicīgo prognožu modelis:
   a) tika izstrādāta jauna metode lineārās regresijas modeļa izveidei, kas ņem vērā putekšņu koncentrācijas datu nestacionaritātes un nelinearitātes problēmas un pamatojas uz gaisa kvalitātes prognozēšanā izmantoto pieeju;
   b) izstrādātā metodika tika izmantota pirmā putekšņu koncentrācijas prognostiskā modeļa izveidei Latvijā.
2. Reģionālais intersezonālais putekšņu koncentrācijas modelis SPI prognozēšanai:
   a) tika izstrādāta jauna metode nākamā gada SPI prognozēšanai reģionālā līmenī, izmantojot intrasezonālā modeļa datu transformācijas paņēmienus;
   b) metodikas efektivitāti apliecināja pirmais universālais SPI prognozēšanas modelis Ziemeļeiropai un Ziemeļaustrumeiropai.
3. Eiropas mēroga deterministisko putekšņu koncentrācijas prognozēšanas modeļu ansamblis:
   a) pirmo reizi tika izveidots un novērtēts esošo deterministisko modeļu optimālais ansamblis, izmantojot iepriekšējo dienu novērojumus optimālās ansambļa modeļu kombinācijas aprēķināšanai.

Hipotēze

Līdzās iepriekš minētajiem trijiem pētījumu virzieniem formulētas promocijas darba hipotēzes.

1. Lokālā mēroga intrasezonālajam modelim:

   tika pieņemts, ka iemesls, kāpēc lineārās regresijas izmantošana putekšņu modelēšanā nav efektīva, ir ievaddatu neatbilstība statistisko metožu izmantošanas principiem un ka lietotā datu transformācija nodrošinās datu atbilstību statistisko procedūru prasībām, uzlabojot modeļa precizitāti.
2. Reģionālā mēroga SPI modelim:

   tika pieņemts, ka SPI ir reģionālais parametrs, ko kontrolē sinoptiskā mēroga meteoroloģiskie procesi. Tieši tāpēc ir iespējams identificēt reģionus, kas reaģē sinhroni un parāda līdzīgas SPI intersezonālās izmaiņas. Modelim nav svarīgas SPI absolūtās vērtības, tās ir iespējams izlīdzināt (homogenizēt) ar normalizācijas palīdzību, tātad SPI universālā modeļa izveide ir iespējama visam reģionam, nevis atsevišķām monitoringa stacijām.
3. Eiropas mēroga deterministisko modeļu ansamblim:

   tika izteikta hipotēze, ka putekšņu prognozēšanas gadījumā modeļa ansambļa precizitāti var paaugstināt, izmantojot optimālo ansambli ar regularizācijas koeficientiem no iepriekšējo dienu simulācijām. Līdz ar to, atšķirībā no standarta datu asimilācijas, modeļa ansamblis nezaudēs precizitāti dažu stundu laikā.

Darba mērķis

Izstrādāt un izvērtēt universālus, viegli lietojamus, augstas precizitātes prognozēšanas modeļus dažāda veida putekšņu koncentrācijas prognozēšanai gaisā no vietējā līdz reģionālajam ģeotelpiskajam mērogam un ar temporālo izšķirtspēju no diennakts (intrasezonālais modelis) līdz sezonālajam (intersezonālais modelis) līmenim.

Uzdevumi mērķa sasniegšanai

1. Novērtēt bērza koncentrācijas īslaicīgās un ilglaicīgās izmaiņas, saistīt tās ar meteoroloģiskajiem parametriem, klimatu, CO_2 koncentrāciju un LAI.
2. Izstrādāt vienkāršu, lokālu modeli, kas atveido putekšņu sezonas parametrus (t. i., sākumu, beigas, intrasezonālās variācijas), balstoties uz meteoroloģisku novērojumu un īslaicīgu prognožu datiem.
3. Izstrādāt universālu reģionālu modeli, kas pēc noteiktā gada bērza SPI un meteoroloģiskajiem datiem varētu aprēķināt nākamā gada SPI.
4. Atrast optimālo ansambli olīvkoku putekšņu intrasezonālajai prognozēšanai Dienvideiropas reģionā.

Publikācijas

Promocijas darbs ir sadalīts trīs secīgos rakstos, kuri atspoguļo problēmas būtību un piedāvā metodiku problēmu risinājumam lokālā līmenī (Raksts I) un reģionālā līmenī (Raksts II un Raksts III), izmantojot statistiskās (Raksts I, Raksts II, Raksts III) un deterministiskās (Raksts III) modelēšanas metodes. Oriģinālpublikācijas ir publicētas vai pieņemtas publicēšanai augsta ranga žurnālos ar JIF 4,6–5,7 (2016) pēdējo divu gadu laikā (2015–2017).

Raksts I. Ritenberga, O., Sofiev, M., Kirillova, V., Kalnina, L., Genikhovich, E. (2016). Statistical modelling of non-stationary processes of atmospheric pollution from natural sources: example of birch pollen. *Agricultural and Forest Meteorology,* 226–227, 96–107.

Lokālā līmeņa pētījums, veltīts diennakts putekšņu koncentrāciju prognozēšanai putekšņu sezonas laikā, nosakot putekšņu daudzumu gaisā ietekmējošos meteoroloģiskos faktorus un parametrizējot modeli, lai to varētu izmantot bērza putekšņu prognozēšanai Rīgā. Modelis ietver vairākas stadijas, ieskaitot daudzpakāpju sarežģīto aerobioloģisko un meteoroloģisko datu transformāciju, iegūstot homogēnu datu kopu, novērtējot datu kopas normālsadalījumu, linearizējot sakarības starp prognozēto vērtību un ietekmējošiem parametriem. Iegūtais daudzsoļu transformācijas rezultāts ir statistisko parametru ievērojama uzlabošana, statistisko procedūru efektivitātes uzlabošana un modeļa precizitātes paaugstināšana.

Rakststs II. Ritenberga, O., Sofiev, M., Siljamo, P., Saarto, A., Dahl, A., Ekebom, A., Šauliene, 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. *Science of the Total Environment* (Pieņemts publicēšanai).

Reģionālā līmeņa pētījums, veltīts bērza putekšņu produktivitātes periodiskuma skaidrošanai ar meteoroloģisko procesu interpretāciju. Pētījums ietvēra 300 Eiropas gaisa monitoringa staciju datu analīzi laikā no 1974. līdz 2015. gadam. Izmantojot klāsteranalīzi, tika nodalīts homogēns reģions, ko reprezentēja 15 monitoringa vietas sešās valstīs. Tika veikta datu transformācija, un prediktoru noteikšanai tika izvēlēta lineārās regresijas analīze. Rezultātā iegūta universāla formula ar konstantiem koeficientiem SPI noteikšanai definētajā reģionā.

Raksts III. Sofiev, M., Ritenberga, O., Siljamo, P., Albertini, R., Arteta, J., Belmonte, J., Bonini, M., Damialis, T., Elbern, H., Friese, E., Galan, C., Hrga, I., Kouznetsov, R., Plu, M., Prank, M., Robertson, L., Selenc, S., Thibaudon, M., Segers, A., Stepanovich, B., Valdebentino, A. M., Vira, J., Vokou, D. (2017). Multi-model ensemble simulations of olive pollen distribution in Europe in 2014; current status and outlook. *Atmospheric Chemistry and Physics* (Pieņemts publicēšanai).

Reģionālā līmeņa pētījums veltīts 2014. gada olīvkoku putekšņu sezonas analīzei Eiropā. Piedāvāts risinājums optimālā modeļa ansambļa izveidei, lai palielinātu putekšņu prognozes precizitāti. Sešu deterministisko modeļu simulāciju rezultāti tika izmantoti par ievaddatiem optimālā ansambļa izveidei.

Autores ieguldījums

Pirmajā pētījumā iekļautais (**Raksts I**), ieskaitot (I) putekšņu monitoringu no 2006. līdz 2016. gadam, (II) paraugu apstrādi un datu analīzi, (III) prognostisko modeļu izstrādi un novērtēšanu, ir autores darbs. Autores ieguldījums promocijas darba reģionālā SPI modeļa izstrādē (**Raksts II**) ietver: (I) putekšņu monitoringu Rīgas stacijā, (II) datu analīzi un reģionālā modeļa izstrādi no pamatiem līdz tā verifikācijai. Promocijas darba trešajā daļā (**Raksts III**) autores ieguldījums ir optimālā modeļu ansambļa izveide, statistiskās daļas izstrāde un ansambļa darbības novērtēšana.

Rezultātu aprobācija

Izveidoto modeļu aprobācija tiek veikta Eiropas līmenī, jo reģionālā SPI modeļa metodoloģija (Raksts II) un modeļu ansambļa (Raksts III) metodoloģija tiek lietota Kopernika Atmosfēras monitoringa dienesta (CAMS) putekšņu prognozēs, lai paaugstinātu to precizitāti. Ir plānots izmantot lokālo modeli dažāda veida putekšņu prognozēšanai Rīgā. Lokālā modeļa prognozes tiks izmantotas arī par deterministisko (t. i., SILAM) modeļu ievaddatiem neeksistējošo reālā laika novērojumu datu vietā, lai atrisinātu novērojumu datu kavēšanās problēmu.

Saistītās publikācijas

Trīs saistīti pētījumi ir publicēti žurnālos ar JIF virs 2,0 (2016), palielinot analizējamo daļiņu dažādību, mainot pētījuma objektu no putekšņiem uz sporām. Daļa saistīto pētījumu rezultāti publicēti arī Latvijas izdevumos. Kopējais publikāciju skaits ir 14.

1. Sofiev, M., Berger, U., Prank, M., 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., and Peuch, V.-H. 2015. MACC regional multi-model ensemble simulations of birch pollen dispersion in Europe, Atmos. Chem. Phys., 15, 8115-8130. DOI:10.5194/acp-15-8115-2015.
2. Kasprzyk, I., Rodinkova, V., Šauliene, I., **Ritenberga, O.**, Grinn-Gofran, A., et al. 2015. *Air pollution by allergenic spores of the genus Alternaria in the air of central and eastern Europe*. In: Environmental Science and Pollution Research. DOI 10.1007/s11356-014-4070-6.
3. Sikoparija, B., Skjøth, C. A., Celenk, S., Testoni, C., Abramidze, T., Alm Kübler, K., Belmonte, J., Berger, U., Bonini, M., Charalampopoulos, A., Damialis, A., Clot, B., Dahl, Å., de Weger, L., Gehrig, R., Hendrickx, M., Hoebeke, L., Ianovici, N., Seliger, A. K., Magyar, D., Mányoki, G., Milkovska, S., Myszkowska, D., Páldy, A., Pashley, C. H., Rasmussen, K., **Ritenberga, O.**, Rodinkova, V., Rybníček, O., Shalaboda, V., Šaulienė, I., Ščevková, J., Stjepanović, B., Thibaudon, M., Verstraeten, C., Vokou, D., Yankova, R., Smith, M. 2016. *Spatial and temporal variations in* air*borne Ambrosia pollen across Europe.* In: Aerobiologia. DOI:10.1007/s10453-016-9463-1.

Izvēlētās starptautiskās konferences

1. **Ritenberga, O.**, Sofiev, M. Forecasting of *inter*- annual variability of Olive seasonal pollen load. *Mediterranean Palynology Symposium* 4–6 September 2017, Barcelona, Spain.
2. **Ritenberga, O.**, Šauliene, I., Berger, U., Sofiev, M. Towards developing Personal Allergy Symptom Forecasting System in Baltic States. *Congress of the European Academy of Allergy and Clinical Immunology* 17–21 June 2017, Helsinki, Finland.
3. **Ritenberga, O.**, Siljamo, P., Sofiev, M. Modelling of intra-seasonal fluctuation and Inter-annual variability of birch pollen concentration: Example of Latvia and Finland. *European Symposium of Aerobiology* 18–22 July 2016, Lyon, France.
4. **Ritenberga, O.**, Kalnina, L. Temporal changes of ragweed (*A. artemisiifolia, A. trifida, A. psilostachya*) pollen concentration in Latvia. *European Symposium of Aerobiology* 18–22 July 2016, Lyon, France.
5. **Ritenberga, O.**, Sofiev, M., Genikhovich, E. Towards developing of short-term statistical model for birch pollen forecast. *European Aerobiology Network and European Aerobiological Society Symposium*, 10–11 November, 2014. Vienna, Austria.

6. **Ritenberga, O.**, Veselova, A., Kalnina, L., Ustupe, L. The result of 10-year aerobiological monitoring of early spring allergenic pollen in Riga (Latvia). 9th *Pollen Monitoring Programme*. Prague. Czech Republic, 26–30 August, 2013.
7. **Ritenberga, O.**, Kalnina, L., Rodinkova, V., Sofiev, M., Vill, M., Korgmaa, V., Shalaboda, V. Regional differences and comparison of observed and modelled birch pollen data. *14th Nordic Aerobiology Society Symposium*. Riga, Latvia, 19–21 August, 2013.
8. Veselova, A., **Ritenberga, O.**, Ustupe, L., Kalnina, L. Influence of meteorological parameters on Alnus and Corylus pollen concentration in Latvia. *14th Nordic Aerobiology Society Symposium*. Riga, Latvia, 19–21 August, 2013.
9. Kalnina, L., **Ritenberga, O.**, Sauliene, I., Sukiene, L., Severova, E. Character of Betulaceae pollen season in Riga and their comparison with Vilnius and Moscow. *14th Nordic Aerobiology Society Symposium*. Riga, Latvia, 19–21 August, 2013.
10. **Ritenberga, O.**, Kalnina, L., Gudovicha, M. Influence of meteorological parameters on Artemisia pollen concentration in Latvia in the period 2003–2011. *5th European Symposium on Aerobiology*. Krakow, Poland, 3–7 September 2012.
11. Kalnina, L., Šauliene, I., **Ritenberga, O.** Variability of grass pollen concentration during the 9 year period (2003–2011). *5th European Symposium on Aerobiology*. Krakow, Poland, 3–7 September 2012.
12. **Ritenberga, O.**, Kalnina, L., Gudovicha, M. Three ragweed species in Latvia. *Second International Ragweed Conference*, Lyon, France, 28–29 March 2012.
13. **Ritenberga, O.**, Kalnina, L. Development of Aerobiological Monitoring in Latvia. *The conference of European Integration and Baltic Sea Region: Diversity and Perspectives*. Riga, Latvia, 26–27 September 2011.
14. **Ritenberga, O.**, Kalnina, L. Seasonal fluctuations of the airborne pollen concentration in Latvia. *Pollen Monitoring Program* (PMP) 8th International Meeting. University of Tartu, Tartu, Estonia, 20–22 May 2011.
15. **Ritenberga, O.** Aerobiology as geographer's field of research. *Next generation insights into geosciences and ecology*. Tartu, Estonia, 12–15 May 2011.
16. **Ritenberga, O.** Airborne pollen concentration in Riga (Latvia) 2003–2010. *53rd International Scientific Conference of Daugavpils University*. Daugavpils, Latvia, 13–15 April 2011.

Par pētījumu rezultātiem ir ziņots 27 starptautiskajās konferencēs.

Eiropas nozaru institūciju interese par rezultātiem

EAS (*European Aeobiological Society*) un IAA (*International Association on Aerobiology*) prezidentu vēstules par nepieciešamību attīstīt nozari Latvijā un Eiropā tika saņemtas 2015. un 2017. gada rudenī. Promocijas darba izstrādes laikā divas reizes (2012. un 2016. gadā) tika saņemti Eiropas Aerobioloģijas asociācijas granti par izcila pētījuma izstrādi putekšņu monitoringa, analīzes un modelēšanas jomā regulāri organizējamos simpozijos un Krievijas Ģeogrāfijas biedrības augstākais novērtējums jauno zinātnieku sasniegumu konkursā (2016. gadā).

Īpaši ieinteresēta par pētījuma rezultātiem ir ietekmīga institūcija – Kopernika Atmosfēras monitoringa serviss (CAMS). Modeļu ansambļa izveides nepieciešamība (Raksts III) ir ieteikta tieši CAMS, savukārt reģionālā SPI (Raksts II) rezultāti tiek izmantoti par ievaddatiem CAMS putekšņu prognozēs Eiropai.

Projekti promocijas pētījuma laikā

1. "Bioloģiskā gaisa piesārņojuma statistiskā modelēšana: bērza putekšņu piemērs" (2015) Izglītības un zinātnes ministrijas bāzes finansējums "Cilvēkresursu piesaiste zinātnei Zemes un Vides zinātnes jomā", pētniece. Noslēdzies.
2. Valsts pētījumu programmas projekts "Ilgtspējīga klimata politika un inovatīvi, energoefektīvi tehnoloģiski risinājumi (KPIET)", pētniece.
3. VPP projekts „Energoefektīvi un oglekļa mazietilpīgi risinājumi drošai, ilgtspējīgai un klimata mainību mazinošai energoapgādei (LATENERGI)", pētniece. Noslēdzies (2014–2017).
4. EU ECMWF Copernicus Atmospheric Monitoring Center *Personal Allergy Symptom FOrecasting system (PASYFO)*, Latvijas grupas vadītāja (2017–2019).

Promocijas darbs sastāv no kopsavilkuma angļu un latviešu valodā un trīs secīgiem rakstiem – promocijas darba daļām.

Kopsavilkums ietver anotāciju, ievadu, metožu un rezultātu daļu, secinājumus un izmantotās literatūras sarakstu.

1. DATI UN PĒTĪJUMA METODOLOĢIJA

Aerobioloģisko (t. s. *aeropalinoloģisko*) procesu modelēšana balstās uz putekšņu un/vai sporu divu stundu, diennakts, sezonu padziļināto datu analīzi. Promocijas darbā izmantoti dati, sākot no divām stundām (Raksts I), ieskaitot diennakti (Raksts III) un, visbeidzot, pārejot uz sezonu (Raksts II) līmeni. Pētījuma daļās atšķiras ģeotelpiskais mērogs – no lokālā (Raksts I), līdz reģionālajam (Raksts II, III).

Latvijā dažādu augu putekšņi gaisa plūsmās atrodas vismaz septiņus mēnešus – no februāra beigām (alkšņa un lazdas putekšņi), līdz septembra beigām – oktobra sākumam, (vībotnes un nātru putekšņi). Septiņu mēnešu laikā gaisā ir sastopams ap 35–40 dažādu putekšņu veidu, kas pārsvarā nāk no anemofiliem augiem. Putekšņi atšķiras pēc formas, izmēra, sastāva, aerodinamiskajām spējām un ietekmes uz alerģisku cilvēku organismu. Putekšņu atrašanās gaisa plūsmās ir atkarīga gan no auga izplatības, kas ietekmē emisiju, gan no meteoroloģiskās situācijas, kas nosaka atrašanās ilgumu gaisa plūsmās.

Ziemeļeiropā un Ziemeļaustrumeiropā visbiežāk alerģiju izraisa **bērza** (Bastl et al., 2016; Huynen et al., 2003) un graudzāļu putekšņi. Tas saistīts gan ar plašu bērza izplatības areālu un ar īpatsvaru mežos, gan ar putekšņu sastāvu – tajos ietilpst agresīvās alerģiju izraisošās olbaltumvielas.

Centrāleiropā bīstamākie putekšņi, kas izraisa alerģijas, ir vērmeļlapu ambrozijas putekšņi (Sikoparija et al., 2016), savukārt paša auga spēja ātri pielāgoties dažādiem augšanas apstākļiem kaitē lauksaimniecībai.

Dienvideiropas reģionā bīstamākais augs, kas izraisa putekšņu alerģijas, ir olīvkoks (D'Amato et al., 2007), augsti sensibilitātes rādītāji reģistrēti Spānijā (44%) un Portugālē (20%) (Pereira et al., 2006), Grieķijā (31,8%) (Dimitrios Gioulekas et al., 2004), Itālijā (24%) (Negrini et al., 1992) u. c. Vidusjūras reģiona valstīs. Tajā pašā laikā olīveļļa ir viens no galvenajiem ekonomiskajiem resursiem Dienvideiropā, lielākā daļa olīvkoku platību (ap 95%) koncentrējas Vidusjūras reģionā (Barranco et al., 2008).

1.1. Dati modeļu izstrādei

1.1.1. Aerobioloģiskie dati

Promocijas darbā uzmanība veltīta bērza putekšņu (Ziemeļeiropā) un olīvkoku putekšņu (Dienvideiropā) datu apstrādei un analīzei.

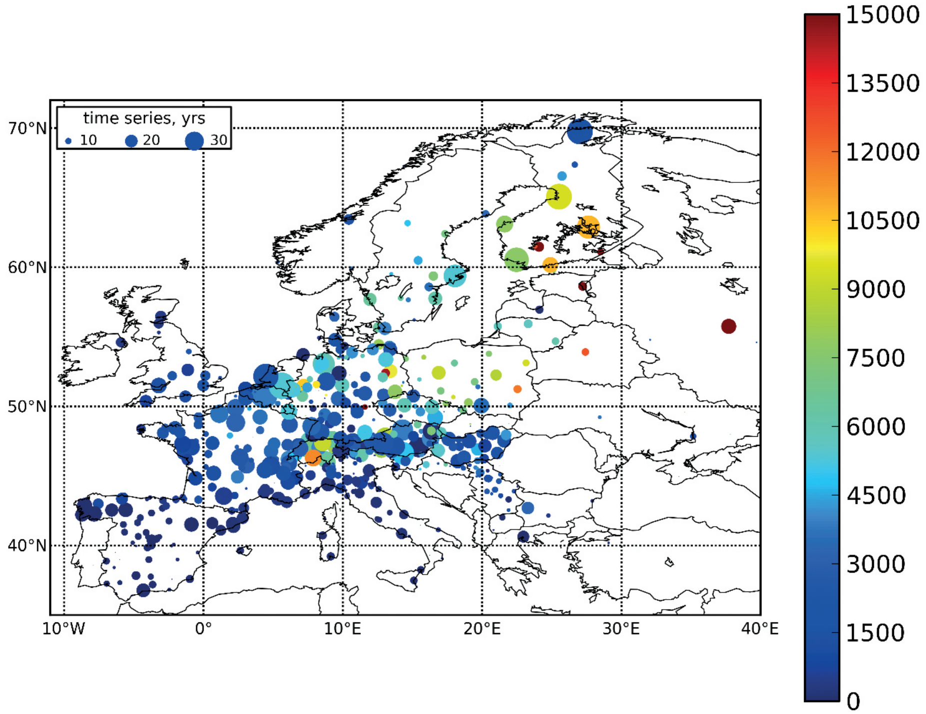
Aerobioloģisko datu analīze sākas ar datu iegūšanu, t. i., regulāru, nepārtrauktu gaisa monitoringu veģetācijas sezonas laikā vairāku gadu garumā. Garākās pētījumā izmantotās datu rindas nāk no Somijas (1.1. attēls) par laika posmu no 1974. līdz 2015. gadam. Rīgā aerobioloģiskais monitorings tika sākts 2003. gadā, un datu apjoms ir pietiekams, lai veidotu īslaicīgas lokālas prognozes, balstoties uz divu stundu un diennakts putekšņu koncentrācijas vērtībām (Raksts I).

Datu iegūšana noritēja atbilstoši datu kvalitātes kontroles izstrādātajām vadlīnijām (Galán et al., 2014; Oteros et al., 2013a), kur ir atrunāts viss, sākot no ierīču un vielu izmantošanas monitoringā, ieskaitot mikroskopiskās analīzes nianses un prasības datu analīzei.

Aerobioloģiskajā monitoringā tika izmantots Hirsta tipa (Hirst, 1954) septiņu dienu *Burkard* putekšņu–sporu uztvērējs, kura specifikācijas paredz vismaz vienu nedēļu nobīdi novērojumu datu iegūšanā. Tas saistīts ar nepieciešamību ievākt (septiņas diennaktis) un manuāli apstrādāt (vienu dienu) putekšņu paraugus. Automatizētā putekšņu monitoringa izmēģinājumi ir sākušies vairākās Eiropas monitoringa stacijās (Scheifinger et al., 2013), bet pagaidām tā precizitāte ievērojami atpaliek no manuālā monitoringa precizitātes.

Putekšņu atpazīšana un uzskaite tika veikta LU ĢZZF kvartārvides laboratorijā, izmantojot *Primo Star* gaismas mikroskopu ×400 palielinājumā un izvēloties vertikālo skaitīšanas metodi (Carinanos et al., 2000), t. i., 12 vertikālās līnijas 2 mm attālumā cita no citas, lai pilnībā pārklātu diennakts paraugu 14×48 mm.

Reģionālo pētījumu veikšanai aerobioloģiskie dati tika ņemti no Eiropas Aeroalergēnu tīkla EAN pieejamām monitoringa stacijām (1.1. attēls), vienojoties ar attiecīgo valstu pārstāvjiem par kopīgu pētījumu veikšanu. Bērza SPI gadījumā (Raksts II) nosacījums bija vismaz 11 novērojumu gadi; olīvu putekšņu modelēšanas datu ansambļa nosacījums bija vismaz 30 secīgas novērojumu dienas, kur vismaz 10 dienu laikā putekšņu koncentrācija pārsniedz 3 putekšņus/m^3 (Raksts III).



1.1. attēls. EAN monitoringa stacijas un bērza vidējais ģeometriskais SPI [putekšņi dienā/m^3] (vizualizēts ar SILAM)

1.1.2. Meteoroloģiskie un vides kvalitātes dati

Promocijas darbā aerobioloģiskie procesi tiek skaidroti ar meteoroloģiskajiem apstākļiem (gaisa temperatūra, īsviļņu saules radiācija, gaisa relatīvais mitrums, nokrišņu summas) un CO_2 koncentrāciju gaisā.

Lokālā modeļa uzbūvei (Raksts I) tika izmantoti Latvijas Vides, ģeoloģijas un meteoroloģijas centra stundu un diennakts meteoroloģiskie dati no meteoroloģiskās stacijas „Rīga-Universitāte”. Lokālā modeļa verifikācijai izmantoti meteoroloģiskās prognozes dati no operatīvā ECMWF arhīva.

Visbiežāk minētie SPI raksturojošie meteoroloģiskie faktori ir temperatūra un nokrišņu daudzums (Dahl and Strandhede, 1996; Latałowa et al., 2002; Yli-Panula et al., 2009) dažādos laika posmos. Bez minētajiem tika izvēlēti vēl divi faktori – īsviļņu saules radiācija un akumulētā siltuma daudzums. Trīs stundu vidējie meteoroloģiskie dati tika iegūti no *European Re-analysis ERA-Interim* datiem (Dee et al., 2011; Simmons et al., 2010). Meteoroloģiskie dati ir pieejami par periodu no 1980. līdz 2015. gadam, kas ierobežoja datu izlasi līdz 36 gadiem.

Oglekļa dioksīds ir viens no parametriem, kas, iespējams, ietekmē putekšņu daudzumu gaisā (Albertine et al., 2014). Analīzei izmantotie CO_2 dati tika lejupielādēti no NOAA Zemes sistēmu pētniecības laboratorijas (ESRL) publiskā arhīva (*Ed Dlugokencky un Pieter Tans, NOAA/ESRL* (www.esrl.noaa.gov/gmd/ /ccgg/trends/)) par laika posmu no 1980. līdz 2015. gadam.

1.2. Pamata metodes nestacionaritātes reducēšanai un datu homogenizācijai

Precīzākā prognozēšanas modeļa izveidei atkarībā no temporālā mēroga ir jādefinē un jānodala laika posmi, kuriem atbilst dažādas augu attīstības fāzes, lai varētu veikt rūpīgāku ietekmējušo faktoru analīzi.

Trīs pamata daļas putekšņu prognozēšanas modeļa uzbūvē tiek definētas kā atsevišķi uzdevumi, izstrādājot prognostiskos putekšņu modeļus:

1) putekšņu sezonas sākuma un beigu posma noteikšana (Raksts I, Raksts III);
2) sezonas gaita (Raksts I, Raksts III) jeb intrasezonālās putekšņu koncentrācijas variācijas;
3) sezonālais putekšņu indekss (Raksts II).

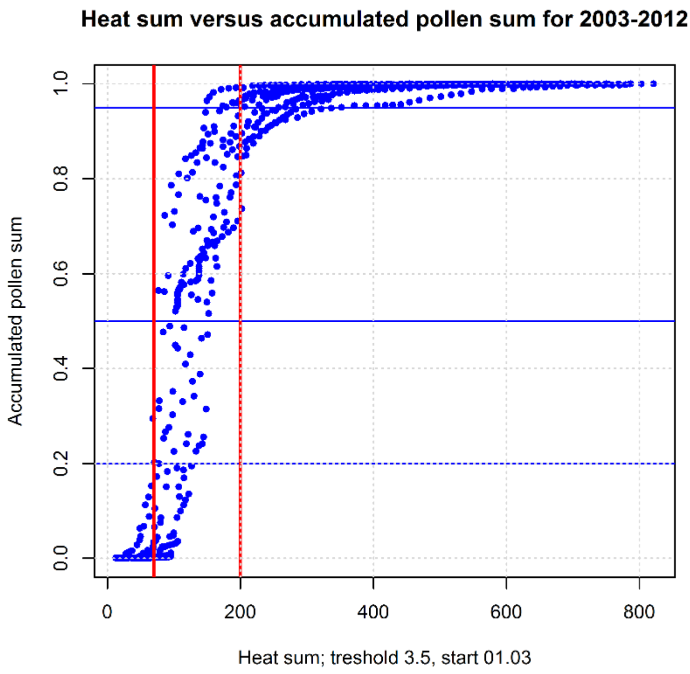
Katru daļu ietekmē ļoti atšķirīgi procesi.

Aerobioloģiskie procesi jāpēta kompleksi. Tajā pašā laikā noteiktā pētījuma stadijā tie jānošķir, jo katru etapu ietekmē atšķirīgi faktori vai to kombinācija. Atsevišķos posmos tika nodalīti faktori, kas nosaka produktivitāti, t. i., SPI (Raksts II); faktori, kas nosaka emisiju un pārnesi; izgulsnēšanos kontrolējošie procesi (Raksts I, Raksts II). Minētie parametri tika izmantoti, par prediktoriem nosakot putekšņu koncentrāciju ietekmējošos faktorus ar daudzpakāpju lineārās regresijas palīdzību.

Datiem, kas tiek izmantoti lineārās regresijas analīzē, jāatbilst vairākiem kritērijiem: jāatbilst normālsadalījumam, jāreprezentē stacionāri un ergodiski procesi un prediktoriem un prediktantam jāveido lineāras sakarības.

1.2.1. Uzkrātā siltuma aprēķināšana

Putekšņu sezonas sākuma un beigu posma noteikšanai (Raksts I) tika izmantota G. Linsser (1867) ideja, t. i., kokiem piemīt spēja uzkrāt siltumu, un to attīstības fāzes ir tieši atkarīgas no uzkrātā siltuma daudzuma (1.2. attēls).



1.2. attēls. Uzkrātā siltuma daudzums pret akumulēto putekšņu daudzumu desmit bērza putekšņu sezonām Rīgā

Promocijas darbā (Raksts I, II, III) tika izmantota modificēta un attiecīgajam reģionam pielāgota uzkrātā siltuma aprēķināšanas formula (Sofiev et al., 2012):

$$H(d) = \sum_{d=d_s}^{D} [\overline{T(d)} - T_{c-o}]_+ \ (1),$$

kur H ir temperatūru summa jeb uzkrātā siltuma daudzums, d ir diena, d_s – sākuma diena siltuma akumulācijai, $T(d)$ – diennakts vidējā temperatūra, T_{c-o} – sliekšņa temperatūra, zem kuras nenotiek siltuma uzkrāšanās, $[x]_+$ vienāds ar 0, ja $x < 0$ un x, ja $x \geq 0$ (neiekļauj temperatūras zem sliekšņa vērtības).

Šis vienādojums prasa divas konstantas vērtības – sākuma dienu siltuma akumulācijai (d_s) un sliekšņa vērtību dienas vidējai temperatūrai (T_{c-o}), virs kuras notiek siltuma uzkrāšanās. Minētās vērtības tika piemeklētas, analizējot 10 bērza putekšņu sezonas Rīgā pret uzkrātā siltuma daudzumu, secīgi pārbaudot datumus no 20.02. līdz 10.03. un sliekšņa temperatūras no 0 °C līdz 5 °C ar 0,5 °C soli. Par optimālo kombināciju tika izvēlēta tā, kuras standartnovirze trīs kumulatīvo putekšņu summu līmeņos ir minimāla (1.2. attēls).

Tā par sākuma dienu siltuma akumulācijai tika noteikts 1. marts, un par sliekšņa temperatūru 3,5 °C. Minētās vērtības sakrīt ar tām, ko izmanto reģionālajā modelī SILAM.

Ņemot vērā, ka par noteiktu putekšņu veidu sezonu tiek uzskatīts periods, kad 90% putekšņu atrodas gaisā, tika noteiktas uzkrātā siltuma robežas, kas ietver 90% putekšņu daudzuma no ikgadējā SPI. Tiek uzskatīts, ka 5% putekšņu sezonas sākumā un beigās ir saistīti ar putekšņu tālo pārnesi, bet 90% – lielākoties ar lokālo ziedēšanu (Jato et al., 2006).

Šī metode ar nelielām modifikācijām tika izmantota par pamatu trīs promocijas darba pētījumos.

1.2.2. Temporālās neviendabības mazināšana putekšņu datos

Uzkrātā siltuma daudzums tika izmantots ne tikai sezonas sākuma un beigu posma noteikšanai, bet arī temporālās neviendabības mazināšanai. Viena no problēmām aerobioloģiskajos pētījumos saistīta ar daudzgadīgo datu temporālo mainīgumu – t. i., sezona sākas, ilgst un beidzas dažādos laikos.

Kalendārais laiks ir neērts mainīgais statistiskajā analīzē. Tas tika aizvietots ar mainīgo, kas ir nepārprotami saistīts ar kalendāro laiku tā, lai putekšņu sezonas sākums, ilgums un beigas būtu vienādi visos pētāmajos gados. Uzkrātā siltuma daudzums atbilst šiem kritērijiem, jo tā ir monotona funkcija laikā, kas vienkārši aprēķināma no diennakts vidējās temperatūras datiem.

Nomainot temporālo skalu pret siltuma skalu, tika iegūta temporāli homogēna datu kopa. Piemēram, lokālā modeļa (Raksts I) gadījumā, uzkrātā siltuma daudzumam sasniedzot 70 vai 200 grādu dienas, tiek konstatēts attiecīgi sezonas sākums vai beigas.

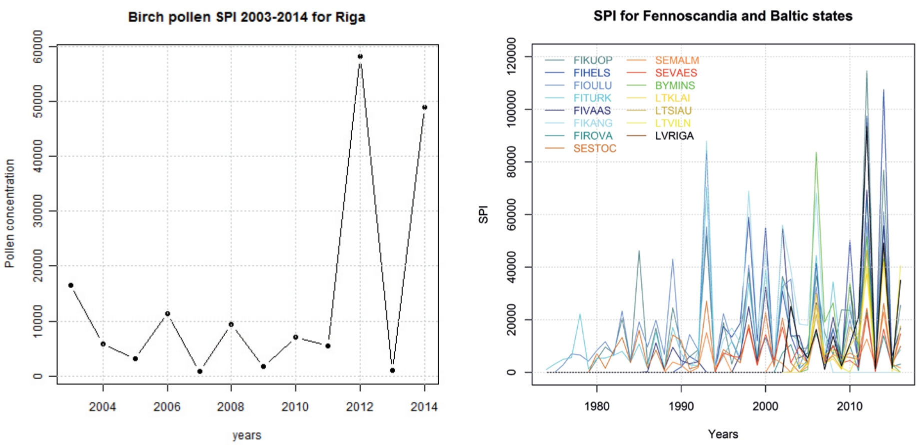
Vairāki pētījumi liecina, ka SPI lielākoties ir atkarīgs no iepriekšējā gada meteoroloģiskajiem apstākļiem noteiktā laika posmā (Raksts II, ievads). Astronomiskais laiks parāda ievērojamas nobīdes attiecībā uz fenoloģisko fāžu iestāšanās laikiem arī SPI analīzes gadījumā. Lai noteiktie periodi sezonas un gada griezumā būtu salīdzināmi pēc fenoloģiskajām fāzēm, atkal tika izvēlēts uzkrātā siltuma daudzums kā fenoloģiski nozīmīgs laika ass mainīgais.

Reģionālā līmenī pat pēc siltuma skalas ieviešanas astronomiskā laika vietā tika konstatētas sezonu iestāšanās laika atšķirības, tas saistīts ar augu spēju pielāgoties – ziemeļos fenoloģiskās fāzes iestājas, ja ir mazāks uzkrātā siltuma daudzums nekā definētā reģiona dienvidu daļā.

Minētās problēmas risināšanai temporālā skala tika izteikta procentos no gadā uzkrātā siltuma daudzuma (Raksts II), normalizējot ar katras stacijas ilglaicīgo vidējo uzkrātā siltuma daudzumu. Ikgadējā uzkrātā siltuma daudzums variē ar divkāršu koeficientu definētā reģiona robežās, turpretī relatīvās vienībās vērtības ir ļoti līdzīgas.

1.2.3. Nestacionaritātes mazināšana putekšņu datos: normalizācija

Augu saražoto putekšņu daudzums un līdz ar to SPI ir ļoti mainīgs gadu no gada (1.3. attēls). Tas ir atkarīgs no vairākiem ilglaicīgiem un liela mēroga procesiem, kā, piemēram, iepriekšējo gadu ziedēšanas intensitātes, meteoroloģiskajiem un vides apstākļiem ziemas mēnešos, pirmssezonālajiem apstākļiem pavasarī utt. (Dahl and Strandhede, 1996; Linkosalo et al., 2006; Stach et al., 2008).



1.3. attēls. Lokālās (kreisajā pusē) un reģionālās (labajā pusē) bērza SPI izmaiņas

Ņemot vērā to, ka (I) intrasezonālās putekšņu variācijas ir saistītas ar īslaicīgām meteoroloģiskajām izmaiņām un ka (II) nav iespējams izveidot pilnvērtīgu SPI modeli, balstoties tikai uz lokāliem datiem, bija nepieciešams mazināt sezonālās SPI atšķirības, padarot dažādas sezonas savstarpēji salīdzināmas. Intrasezonālo un intersezonālo procesu nodalīšanai tika piemērota normalizācija, t. i., diennakts putekšņu koncentrācija $C_i(d)$ tika dalīta ar attiecīgā gada SPI katram aplūkotajam gadam i, tādējādi iegūstot savstarpēji salīdzināmas diennakts relatīvās vērtības (2).

$$c_i(d) = \frac{C_i(d)}{\sum_{d \in i} C_i(d)} \quad (2)$$

Lai iegūtu sezonālās relatīvās vērtības ($SPI_i^{geomean}$) un SPI prognozēšanu varētu modelēt reģionālā līmenī, katra novērojumu punkta i katram gadam Y SPI tika normalizēts ar vidējo ģeometrisko SPI visam analizējamam periodam $SPI_i^{geomean}$,

$$SPI_i^{norm}(Y) = \frac{SPI_i(Y)}{SPI_i^{geomean}} \quad (3)$$

Tādējādi tika ierobežota vietējās ziedēšanas ietekme, padarot analizējamās monitoringa stacijas reģionā savstarpēji salīdzināmas.

2. REZULTĀTI UN DISKUSIJA

2.1. Nestacionāro procesu statistiskā modelēšana – lokālās intrasezonālās prognozes: Latvijas piemērs

Raksts I. Ritenberga, O., Sofiev, M., Kirillova, V., Kalnina, L., Genikhovich, E. (2016). Statistical modelling of non-stationary processes of atmospheric pollution from natural sources: example of birch pollen. *Agricultural and Forest Meteorology, 226–227*, 96–107.

Lokālā līmeņa pētījums, veltīts diennakts putekšņu koncentrāciju prognozēšanai putekšņu sezonas laikā, nosakot putekšņu daudzumu gaisā ietekmējošos meteoroloģiskos faktorus un parametrizējot modeli, lai to varētu izmantot bērza putekšņu prognozēšanai Rīgā.

Autores ieguldījums pētījumā ir: (I) putekšņu monitoringa veikšana no 2006. līdz 2016. gadam; (II) paraugu apstrāde un datu analīze; (III) prognostisko modeļu izstrāde un novērtēšana; (IV) rezultātu sagatavošana publicēšanai.

Pētījumā izmantota novatoriska pieeja lineārās regresijas modeļa izveidei, kas ņem vērā un risina putekšņu koncentrācijas datu ne-normālā sadalījuma, nestacionaritātes un nelinearitātes problēmas ar datu transformācijas palīdzību. Metode pirmo reizi tika piemērota putekšņu koncentrācijas prognozēšanai, un tika izveidots pirmais putekšņu koncentrācijas prognostiskais modelis Latvijai (Rīgai).

Datu daudzsoļu transformācijas rezultāts ir statistisku procedūru efektivitātes uzlabošana un modeļa precizitātes paaugstināšana.

Vidējo sezonas gaitu izskaidro uzkrātā siltuma daudzums (Raksts I), savukārt intrasezonālo putekšņu koncentrācijas izmaiņas iespējams prognozēt, analizējot novirzes no vidējiem sezonas rādītājiem.

Tika veikta sakarību linearizācija (Raksts I) starp prediktoriem (meteoroloģiskie parametri) un prediktantiem (putekšņu koncentrācijas novirzes). Linearizācija tika veikta, projicējot meteoroloģiskus datus uz putekšņu koncentrācijas datiem (2.1. attēls). Deviņu gadu dati par katru prediktoru tika sadalīti intervālos, un katram intervālam tika aprēķināta vidējā putekšņu koncentrācijas noviržu vērtība (relatīvā gaisa mitruma piemērs 2.1. attēlā). Turpmāk analīzē izmantotas linearizētās prediktoru vērtības. Minētā transformācija minimizē nelineārās sakarības un ļauj izmantot lineāro regresiju prediktoru kombinācijas noteikšanai.

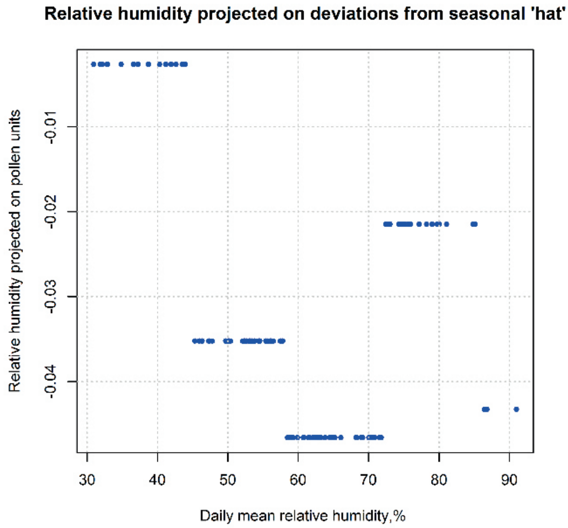
Statistiski nozīmīgi (p > 0,05) prediktori un to koeficienti noteikti ar daudzsoļu lineāro regresiju: brīvais loceklis (0,03), diennakts vidējā temperatūra (0,65), diennakts vidējais mākoņainums (1,05), diennakts nokrišņu summa (0,53), vēja u-komponente (0,68).

Modeļa novērtēšana sastāv no divām daļām:

1) modeļu spēja uzrādīt sezonas parametrus – sākumu, maksimālo koncentrāciju, beigas, intrasezonālās variācijas, dienas ar izteikti zemu vai augstu putekšņu koncentrāciju utt;

2) Modeļa spēja atveidot praktiski nozīmīgās robežvērtības sasniegšanu.

Kopējā modeļu precizitāte MA ir zemāka par 5% kontroles datu kopai, toties OR un citi parametri POD, FAR (Raksts I, pielikums) ir pat nedaudz augstāki. Atšķirības ir saistītas ar modeļa kontroles procedūras nenoteiktību: kontroles datu kopa ietvēra 70 dienas ar putekšņu koncentrāciju virs robežlieluma, aptuveni puse gadījumu tika atveidota; 170 dienas bija zemāka putekšņu koncentrācija par robežlielumu, un tie tika atveidoti 97% gadījumu (Raksts I, 2. tabula).



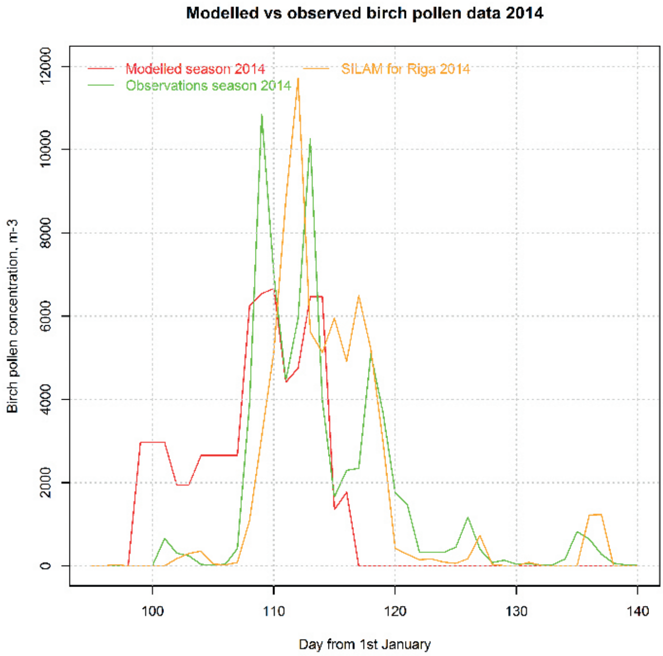
2.1. attēls. Relatīvā mitruma un noviržu no SPI linearizācijas piemērs Rīgai (Raksts I)

Modeļa rezultātu salīdzinājums ar esošajiem modeļiem ir problemātisks, jo tas ir pirmais modelis Rīgai un tikai daži modeļi Eiropā ir spējīgi prognozēt ikdienas putekšņu koncentrāciju. Tomēr jāatzīmē modeļa līdzības un atšķirības: īslaicīgo putekšņu prognožu veidošanā ir izplatīta prakse izmantot autoregresijas metodes – t. i., prognoze ir iepriekšējo dienu koncentrācijas funkcija, piemēram, bērzam (Inatsu et al., 2014) vai graudzālēm (Stach et al., 2008). Rīgas gadījumā autoregresiju nav iespējams lietot, jo dienu no dienas ir novērotas krasas putekšņu koncentrācijas izmaiņas, "šodienas" putekšņu dati "rītdienas" prognozei nav iespējami, jo monitoringa specifika paredz 7 dienu nobīdi. Kaut gan šāds spēcīgs prediktors netiek iekļauts modelī, modeļa precizitāte pēc korelācijas koeficienta ir 0,64–0,94, savukārt autoregresijas modeļiem – 0,6–0,9 bērzam Japānā (Inatsu et al., 2014) un 0,6–0,7 graudzālēm Polijā (Stach et al., 2008).

Modeļa rezultāti tika salīdzināti ar reģionālā modeļa SILAM prognozēm, un tie uzrādīja ievērojami labākus rezultātus (2.2. attēls). Jāpievērš uzmanība tam, ka lokālais modelis precīzi atveidoja dienas ar maksimālo koncentrāciju 2014. gada sezonā, savukārt SILAM labāk atveidoja līknes daļu, kas ir saistīta ar putekšņu tālo pārnesi. Tā kā modeļiem ir atšķirīgi darbības principi, jāņem vērā, ka to kopīga izmantošana paver iespējas uzlabot reģionālo prognožu precizitāti.

Sarežģītā datu transformācija, kas piedāvāta šajā pētījumā, ir nepieciešama modeļa izstrādē un novērtēšanā, lai padarītu iespējamu statistisko metožu izmantošanu. Datu normalizācija, nestacionaritātes reducēšana, linearizācija ir vajadzīgas – teorētiski. Protams, ir iespējams mēģināt izmantot daudzsoļu lineāro regresiju bez datu transformācijas (Raksts I, 4. tabula). Netransformēts datu komplekts uzrādīja ievērojami zemāku modeļa precizitāti.

Iepriekšminētā datu transformācija un daudzsoļu lineārā regresija veido modeli, kas prognozē normalizēto putekšņu koncentrāciju novirzes no daudzgadīgās vidējās sezonālās līknes un uzkrātā siltuma skalas.



2.2. attēls. Modeļu darbības salīdzinājums ar novērojumu datiem 2014. gada bērza putekšņu sezonā (Raksts I)

Inversā transformācija ir nepieciešama lietišķai prognostiskā modeļa izmantošanai. Regresijas modelis tika veidots transformētai datu kopai – t. i., gan transformētiem prediktoriem, gan prediktantiem. Atgriešanās pie netransformētiem datiem sekmē modeļa precizitātes samazināšanos. Līdzīgs efektu aprakstīja Toro et al. (1998).

Lokālā modeļa izstrādē nav iespējams iekļaut tekošā gada SPI aprēķinu, tātad sezonas sākumā, zinot vienīgi meteoroloģisko prognozi, ir iespējams prognozēt normalizētās diennakts putekšņu koncentrācijas. Lai koncentrācijas denormalizētu un pārrēķinātu absolūtās vērtībās, ir nepieciešama atsevišķa, reģionāla SPI modeļa izstrāde.

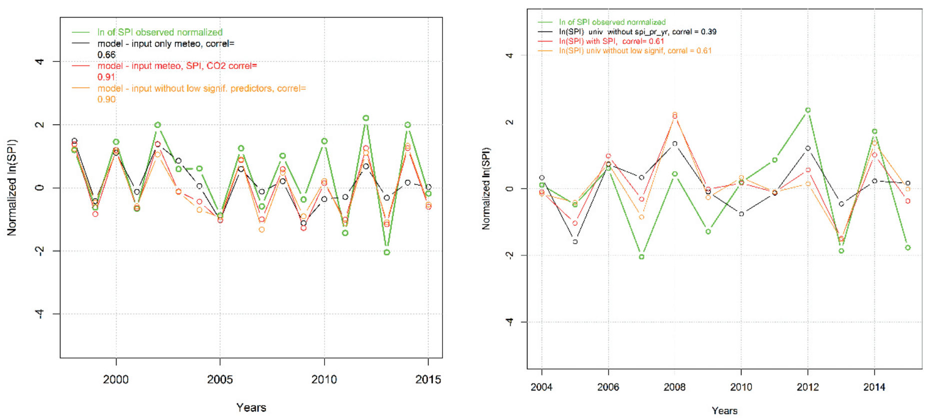
2.2. SPI statistiskā modelēšana Ziemeļeiropā un Ziemeļaustrumeiropā

Raksts II. Ritenberga, O., Sofiev, M., Siljamo, P., Saarto, A., Dahl, A., Ekebom, A., Šauliene, 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. *Science of the Total Environment* (Pieņemts publicēšanai).

Pētījuma rezultātā tika izstrādāta metode putekšņu sezonas indeksa aprēķināšanai lieliem reģioniem, balstoties uz Ziemeļeiropas un Ziemeļaustrumeiropas reģiona piemēru un ietverot Somijas, Zviedrijas, Baltijas valstu, Baltkrievijas teritoriju, savukārt Krievijas un Norvēģijas iekļaušanu analīzē ierobežoja datu trūkums. Statistiskais modelis tika izveidots, izmantojot meteoroloģiskos, ģeofiziskos un bioloģiskos raksturlielumus gadam Y-1, lai prognozētu gada Y sezonālo putekšņu indeksu.

Autores ieguldījums pētījumā iekļauj (I) datu iegūšanu – putekšņu monitoringu Latvijā, sākot ar 2006. gadu; (II) datu analīzi; (III) modeļa izveidi un novērtēšanu.

Ir zināms, ka bērza SPI ir novērojams divu gadu periodiskums (2.3. attēls) (Zink et al., 2013) un augšupejošais trends (Spieksma et al., 2003), sākot vismaz no 1974. gada.



2.3. attēls. Reģionālā SPI modeļa salīdzinājums ar novērojumu datiem stacijās Vāsā, Somijā (kreisajā pusē), un Rīgā (labajā pusē)

Tika izteikta hipotēze, ka SPI ir reģionālais parametrs, ko kontrolē sinoptiska mēroga, t. i., dažu simtu km, meteoroloģiskie procesi. Tieši tāpēc būtu jābūt iespējai identificēt reģionus, kas reaģē sinhroni un parāda līdzīgas SPI starpsezonālās izmaiņas. Modelim nav svarīgas SPI absolūtās vērtības, jo to nosaka tas, cik tuvu uztvērējam augs atrodas. Tātad, nosakot reģionu, tas būtu jāuzskata par "kasti", nedalot atsevišķās monitoringa stacijās, tādēļ reģiona homogenizācijā tika lietota SPI normalizācija (Raksts II).

Pētījuma primārais mērķis bija uzbūvēt universālu modeli, kas būtu piemērojams lieliem reģioniem, demonstrējot SPI telpiskās vispārināšanās iespējamību. Līdz šim lielākā daļa iepriekšējo pētījumu koncentrējās uz vienu vai dažām tuvumā izvietotām stacijām (sk. Raksta II ievadu).

Otrā galvenā piedāvātās pieejas atšķirība ir saistīta ar datu transformāciju – izlīdzinot uzkrātā siltuma atšķirības reģionālā griezumā un ierobežojot SPI telpiskās variācijas. Tālāk prediktoru noteikšanai tika izmantota daudzsoļu lineārā regresija. Par prediktoriem izvēlēti parametri no sešiem intervāliem (Raksts II). Ar regresijas palīdzību tika noteikti tekošā gada parametri, kas ietekmē nākamā gada SPI reģionā:

- nokrišņu summa $Pr_{0,13}$ par periodu 0,13–0,25 no gadā akumulētā siltuma daudzuma;
- vidējā temperatūra $T_{0,13}$ par periodu 0,13–0,25 no gadā akumulētā siltuma daudzuma;
- vidējā temperatūra $T_{0,25}$ par periodu 0,25–0,45 no gadā akumulētā siltuma daudzuma;
- īsviļņu saules radiācija SW_0 par periodu 0–0,05 no gadā akumulētā siltuma daudzuma;
- SPI_{Y-1};
- $CO_{2\ Y-1}$.

Lai noteiktu reģionālās SPI formulas prediktoru koeficientus, tika piemērota *bootrstrap* metode. Analizējamā reģionā atrodas 15 monitoringa stacijas, tādēļ ar *bootstrap* metodi tika veikti 15 atkārtojumi 15 datu kopām, lai atrastu prediktoru koeficientus. Katram prediktoram tika atrastas 15 vērtības, no tām iegūstot vidējos regresijas koeficientus un standartnovirzi (Raksts II, 3. tabula).

Modeļa (4) izveides laikā tika konstatēts, ka, izmantojot tikai un vienīgi meteoroloģiskos datus un CO_2 (bez iepriekšējā gada SPI), ir iespējams atveidot bērza SPI divgadīgo ciklu. Papildus tika izveidots meteoroloģiskais modelis. Šī iespēja ļauj prognozēt nākamā gada SPI, zinot tikai tekošā gada meteoroloģiskos datus, un šis modelis varētu būt piemērojams vietās, kur netiek veikti aerobioloģiskie novērojumi.

Reģionālā sezonālā putekšņu indeksa $\Delta SPI^{reg}(Y)$ formula gadam *Y*, balstoties uz iepriekšējā *Y-1* gada meteoroloģiskajiem un CO_2 datiem:

$$\Delta SPI^{reg}(Y) = a_0 + a_{CO_2}CO_{2,Y-1} + a_{pr}Pr_{0.13} + a_{sw}sw_0 + a_{T_{0.13}}T_{0.13} + a_{T_{0.25}}T_{0.25} \quad (4)$$

SPI_{Y-1} pievienošana prediktoriem ievērojami palielina modeļa kvalitāti. Galīgā formula ietver visus minētos prediktorus ar konstantiem koeficientiem visam reģionam.

Reģionālā modeļa darbības novērtēšanā tika izmantotas līdzīgas metodes kā lokālā modeļa novērtēšanā (Raksts I), t. i., OR, MA, F2, atšķirību novērtēšana starp augsto un zemo sezonu skaitu – virs vai zem ilggadējās vidējās sezonas (metožu aprēķināšanas aprakstu sk. Raksts I, pielikums). Atsevišķās stacijās rezultāti var ievērojami atšķirties, modelim ir tendence labāk strādāt stacijās ar garākām datu rindām (piem., Turku, Kuopio). Modelis pareizi aprēķina *zemās* sezonas, un tikai dažos gadījumos prognozes bija nepareizas *augstajiem* gadiem. OR vērtības ir nepārtraukti augstas (Raksts I, 3a attēls).

Salīdzināšana ar citiem modeļiem ir problemātiska, jo līdzīgi modeļi līdz šim vēl nav veidoti. Vistuvākais piemērs ir aprakstīts Somijas pētījumos (Ranta and Satri, 2007), kur tika veidoti trīs lokālie modeļi Turku, Kuopio un Oulu monitoringa stacijām.

Salīdzinot korelācijas koeficientus, reģionālais SPI uzrādīja lielāku modeļa precizitāti 0,82 pret 0,59 Turku gadījumā, 0,77 pret 0,61 Oulu gadījumā. Līdzīga Turku un Oulu prognozēšanas precizitāte tika sasniegta Zviedrijas pētījumos (Dahl and Strandhede, 1996). Salīdzinoši augsti korelācijas rādītāji, raksturojot SPI, sasniegti Polijā, Poznaņā (Grewling et al., 2012), bet prognostiskais modelis netika veidots. Tika lietota metode, lai aprēķinātu divgadīgā SPI cikla lūzumu, bet Poznaņa ir ārpus analizējamā reģiona, tāpēc reģionālais modelis nav tieši izmantojams salīdzināšanai.

Jāuzsver, ka izmantotās metodes ievērojamās atšķirības no citiem modeļiem ir saistītas ar salīdzinoši labu modeļa spēju prognozēt nākamā gada SPI tikai pēc tekošā gada meteoroloģiskajiem parametriem un CO_2, t. i., neņemot vērā iepriekšējo gadu SPI. Meteoroloģiskā modeļa kvalitāte ir zemāka nekā modelim ar pilnu prediktoru komplektu, un tas liecina par būtisku bioloģiskā signāla nozīmi prognozēs. Tajā pašā laikā meteoroloģisko pediktoru modelis spēj atveidot divgadīgu SPI ciklu un uzrādīt cikla lūzuma gadus – t. i., divgadīgais cikls tiek kontrolēts ar reģionālā mēroga laikapstākļiem. Definētā reģiona dienvidu daļā (Zviedrijas dienvidos, Lietuvā) universālā reģionālā modeļa precizitāte sāk sarukt. Tas saistīts augu reakciju uz klimatiskajiem apstākļiem un to pakāpenisku nomaiņu ziemeļu–dienvidu un rietumu–austrumu virzienā.

Ārpus definētā reģiona stacijās Beļģijā (Briselē), Krievijā (Maskavā) un Norvēģijā (Tronheimā) modelis ir precīzs un darbojas labi gadiem ar izteiktu divgadīgu ciklu, bet, ciklam mainoties, modeļa precizitāte samazinās.

Bērza SPI prognozēšanai dienvidu reģionos nepieciešams veidot atsevišķu modeli – tas saistīts ar citu meteoroloģisko parametru nozīmi SPI veidošanā – piemēram, lielāku nozīmi iegūst nokrišņu daudzums un uzkrātā augstuma daudzums ziemas mēnešos, veidojot prediktoru un prediktantu nelineārās sakarības.

2.3. Modeļu ansambļa izveide putekšņu prognozēm Dienvideiropā

Raksts III. Sofiev, M., Ritenberga, O., Siljamo, P., Albertini, R., Arteta, J., Belmonte, J., Bonini, M., Damialis, T., Elbern, H., Friese, E., Galan, C., Hrga, I., Kouznetsov, R., Plu, M., Prank, M., Robertson, L., Selenc, S., Thibaudon, M., Segers, A., Stepanovich, B., Valdebentino, A. M., Vira, J., Vokou, D. (2017). Multi-model ensemble simulations of olive pollen distribution in Europe in 2014; current status and outlook. *Atmospheric Chemistry and Physics* (Pieņemts publicēšanai).

Autores ieguldījums pētījumā ir optimālā modeļu ansambļa izveide, statistiskās optimizācijas procedūra un tās piemērošana sešu CAMS deterministisko modeļu ansamblim. Tika veikta individuālo modeļu, to vidējo vērtību un mediānas analīze, lai atrastu optimālu prediktoru kombināciju multimodeļu ansamblim.

Pētījuma mērķis bija izstrādāt un prezentēt olīvkoku putekšņu (spēcīgāko alerģiju izraisītāju Vidusjūras reģionā) izkliedes pirmo modelēšanas eksperimentu Eiropas mērogā.

Kopernika Atmosfēras monitoringa servisa sešu modeļu (EMEP, EURAM-IM, LOTOS-EUROS, MATCH, MOCAGE, SILAM) ansamblis tika veidots 2014. gada sezonai, rēķinot olīvkoku putekšņu izplatību. Pētījumā daļēji izmantota multimodeļu ansambļa pieeja bērza putekšņu prognozēšanas simulēšanai (Sofiev et al., 2015) ar izmaiņām, kas attiecas uz olīvkoku putekšņu izplatību un modeļa rezultātu demonstrāciju prognožu režīmā.

Viens no variantiem, kā uzlabot modeļu kvalitāti, ir izmantot esošo modeļu kombināciju – t. i., modeļu ansambli (Potempski and Galmarini, 2009). Ansambļa ideja dažādos veidos izmantota gaisa kvalitātes problēmu risināšanai (Johansson et al., 2015) un klimatiskajiem modeļiem (Genikhovich et al., 2010), bet līdz šim tā nav lietota putekšņu prognožu veidošanai.

Trīs ansambļi tika veidoti pēc (I) vidējā aritmētiskā, (II) mediānas un (III) sešu modeļu optimālās kombinācijas. Vidējās vērtības un mediāna tika rēķināta pēc stundu datiem, bet optimizācija piemērota diennakts līmenī. Optimālā ansambļa izveidei izmantota prediktoru (modeļu c_m, $m = 1..M$) lineārā kombinācija c_{opt} (5) tā, lai minimizētu optimālā lauka RMSE J (6):

$$c_{opt}(i,j,k,t,\tau,A) = a_0(\tau) + \sum_{m=1}^{M} a_m(\tau) c_m(i,j,k,t), \quad A = [a_1..a_M], \quad a_m \geq 0 \ \forall m \quad (5)$$

$$J(t,\tau) = sqrt\left[\frac{1}{O}\sum_{o=1}^{O}\left(c_{opt}(i_o,j_o,k_o,t,\tau,A) - c_o(t)\right)^2\right] + \alpha\sum_{m=1}^{M}\left(a_m(\tau) - \frac{1}{M}\right)^2 + \beta\sum_{m=1}^{M}\left(a_m(\tau-1) - a_m(\tau)\right)^2, \quad \tau = \{d_{-k}, d_0\} \quad (6),$$

kur i, j, k, t – indeksi gar x, y, z un laika asīm, M – modeļu skaits ansamblī, O – monitoringa staciju skaits, $\tau = \{d_{-k}:d_0\}$ – laika posms $k+1$ dienām, kas iekļauts analīzē, sākums no d_{-k} līdz d_0 ,τ-1 – iepriekšējās dienas analīzes periods τ-$1=\{d_{-k-1}:d_{-1}\}$, c_m – modeļa m prognozētā putekšņu koncentrācija, c_o – novērotā putekšņu koncentrācija, a_m – no laika atkarīgs modeļa m koeficients ansamblī, a_0 – no laika atkarīgs sistemātiskās novirzes (*bias*) labojums.

Vienādojumā (6) pirmais mainīgais rāda vidējo kvadrātisko kļūdu RMSE par asimilācijas periodu (logu) τ, otrais notur koeficientu novirzīšanos no vienmērīgā sadalījuma, trešais ierobežo koeficienta a_m izmaiņas ātrumu. Svara vērtības α un nosaka regularizācijas stiprumu.

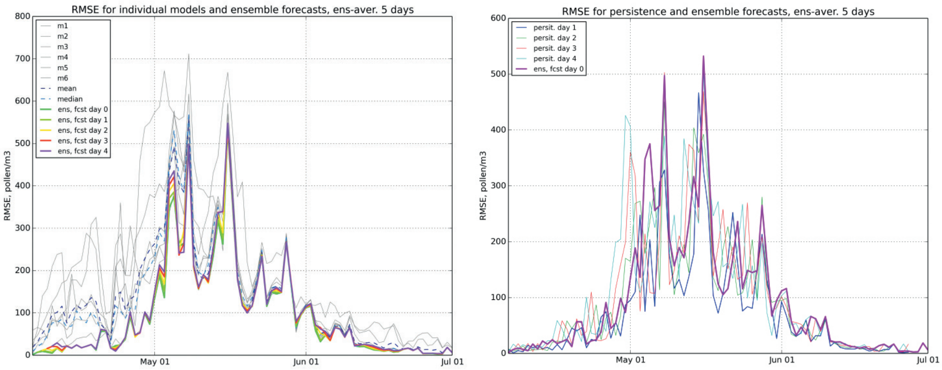
Pētījumā pirmo reizi tiek piedāvāta modeļa ansambļa optimizācijas metode, izmantojot lineāro regresiju. Vienādojums (6) prasa trīs parametru noteikšanu – regularizācijas parametrus α un β un asimilācijas logu T. Tika pārbaudītas vairākas vērtības katram parametram un analīzes periodiem no 1 līdz 15 dienām. Par optimāliem tika izvēlēti α = 0,1 $\beta = 0{,}1$, $T = 5$ *dienas.*

Ansamblis tika veidots, imitējot prognozēšanas režīmu. Pirmkārt, analīze tika veikta, izmantojot datus no analizējamā perioda τ. Iegūtie koeficienti a_i izmantoti vairākas dienas pēc kārtas: no d_1 līdz d_{nf}, kas veido prognozēšanas posmus. Ansambļa veikums tika novērtēts katram prognozes garumam no 1 līdz n_f dienām.

Optimālais ansamblis rāda, ka ikkatrs no sešiem modeļiem ir atbildīgs par noteikta perioda precīzu atveidošanu, savukārt pirms un pēc pamata sezonas, kad putekšņu koncentrācijas ir ļoti zemas, regularizācijas parametri (6) kontrolē sezonas gaitu, pazeminot vērtības līdz apriori 1/6.

Ansambļa prognozēšanas spējas pārbaudes laika tika secināts, ka tam ir ievērojami labāki rezultāti par jebkuru individuālo modeli un par 25–30% labāki nekā ansambļa mediānai vai vidējam rādītājam (4.1. attēls).

Spēcīgākais konkurents prognožu kvalitātes ziņā izrādās vienas dienas *persistence* prognoze – nākamās dienas koncentrācija būs vienāda ar pēdējās novērotās dienas vērtību (2.4. attēls).



2.4. attēls. Individuālo modeļu RMSE un optimālā ansambļa prognoze pret ansambļa vidējo aritmētisko (kreisā puse) un pret *persistence* prognozi (labā puse)

Labs rezultāts vienas dienas *persistence* prognozei ir saprotams, bet, ņemot vērā Eiropas monitoringa sistēmu (dati ienāk ar 7–10 dienu nokavēšanos), minētais modelis nav praktiski izmantojams. Nākamā šāda veida prognozes problēma saistīta ar to, ka modelis spēj darboties tikai monitoringu staciju vietā, jo ir nepieciešami precīzi novērojumu dati, un, visbeidzot, prognozei ātri zūd kvalitāte, skatoties laika griezumā – *persistence* labi strādā ar vienas dienas starpību, bet, tiklīdz to mēģina izmantot +2 vai +3 dienas prognozēm, tā precizitāte ir ļoti vāja, turklāt lokālā līmenī ievērojami labāk strādā vienkāršie statistiskie modeļi (Raksts I).

Nozīmīga individuālo deterministisko modeļu darbībā ir sezonālā nobīde, kas parāda uzkrātā siltuma daudzuma metodes nepilnības. Dienvidu reģionos uzkrātā siltuma formulā jābūt implementētam uzkrātā augstuma parametram, kas aprakstīts, piemēram, (Aguilera et al., 2014), un nav aktuāls ziemeļu reģionos. Vēl viens variants ir pāriešana uz fenoloģiskajiem modeļiem (Chuine and Belmonte, 2004), ņemot vērā putekšņu ģenētisko diferenciāciju.

Iespēja izstrādāto metodi iekļaut dispersijas modeļos ir nākamais solis modeļu precizitātes un prognožu uzlabošanā.

SECINĀJUMI

Promocijas darbā tika veikti trīs secīgi oriģinālpētījumi putekšņu koncentrācijas prognozēšanai Eiropā:

- lokālais modelis diennakts putekšņu koncentrācijas prognozēšanai,
- reģionālais statistiskais modelis SPI prognozēšanai,
- reģionālais deterministisko modeļu ansamblis olīvkoku putekšņu koncentrācijas prognozēšanai Vidusjūras reģionā.

Katrs modelis ir formulēts vispārīgos terminos, tādēļ modeļa metode ir piemērojama jebkurā aerobioloģiskā monitoringa veikšanas vietā un dažādiem putekšņu veidiem. Modeļu izmantošanai citos reģionos ir nepieciešama to reparametrizācija.

Lokāla intrasezonālā modeļa precizitāte MA uzrādīja līdz šim labāko rezultātu, pārsniedzot 80%, koeficients OR = 30.

Reģionālā SPI modeļa gadījumā tika demonstrēts, ka bērza SPI intersezonālās izmaiņas ir sinhronizētās lielu reģionu mērogā – tas paver iespējas veidot samērā vienkāršus prognozēšanas modeļus definētos reģionos.

Labākais intersezonālā modeļa rezultāts panākts, kombinējot iepriekšējā gada meteoroloģiskos parametrus, CO_2 vērtības un iepriekšējā gada SPI. Modelis veiksmīgi spēj noteikt divgadīgā SPI ciklu un atveidot cikla *lūzuma* gadus. Jāuzsver, ka modelis, kurā par ievaddatiem tika izmantoti meteoroloģiskie parametri un CO_2, tāpat spēj atveidot divgadīgu ciklu un cikla lūzuma gadus – tas liecina par meteoroloģijas parametru lielo nozīmi minētā cikla veidošanā.

Reģionālā mēroga deterministisko modeļu prognozēšanas precizitāti var uzlabot, modeļu ansamblim lietojot statistikas datu saplūšanas algoritmus.

Modeļa ansambļa optimālās lineārās kombinācijas bija ievērojami precīzākas par standarta ansambļa pieejām un individuālā modeļa prognozēšanas spējām.

Divi no trīs pētījumiem aktīvi tiek izmantoti Kopernika Atmosfēras monitoringa servisā putekšņu prognožu precizitātes uzlabošanai Eiropā.

Turpmākie pētījuma izaicinājumi

Vienota modeļa izstrāde SPI prognozēšanai Eiropā ir nākamais solis putekšņu prognožu attīstībā. Pastāv divas iespējas: (I) izveidot vienotu modeli, palielinot prediktoru skaitu, un *atļaut* nelineārās sakarības (iespējams, izmantojot metodes no Raksts I); (II) izveidot atsevišķus modeļus Centrāleiropai un Austrumeiropai, kā, piemēram, tika darīts olīvkoku intrasezonālā modeļa gadījumā (Aguilera et al., 2014). Katram variantam ir savi plusi un mīnusi – pirmā pieeja ir izdevīgāka potenciālajiem modeļa lietotājiem, tā ļauj variēt ar prediktoriem, savukārt multimodeļu pieeja, kaut arī ir vienkāršāka katrā atsevišķā reģionā, toties ir neprecīza reģionu robežās, tādēļ to drīzāk varētu izmantot kā starpposmu vienota modeļa izstrādē.

Nākamais solis ir putekšņu prognožu uzlabošana Eiropas līmenī. Dispersijas modeļu priekšrocības ir saistītas ar iespēju prognozēt putekšņu tālo pārnesi, savukārt lokālie modeļi ir ievērojami precīzāki noteiktās vietas putekšņu koncentrācijas prognozēšanā.

Minēto modeļu apvienošana varētu ievērojami uzlabot prognožu kvalitāti Eiropā. Pieejas modeļu apvienošanai var būt dažādas, bet lielākais izaicinājums ir ģeotelpiskā mēroga atšķirības. Domājams, ka modeļus var savienot, izmantojot lokālo modeļu datu asimilāciju dispersijas modeļos vai saplūšanas (*fusion*) metodes, kas tika lietotas Rakstā III. Katras pieejas stiprās un vājās puses tiks izvērtētas turpmāk.

PATEICĪBAS

Vēlos pateikties ģimenei – vecākiem un brāļiem, īpaši mammai Jeļenai – par to, ka iemācīja saprast nemitīgas attīstības nepieciešamību, paldies par sirsnīgu atbalstu vienmēr un visur. Paldies dēlam Robertam par sapratni un pacietību, paldies brāļiem Antonam, Pāvelam, Robertam, Nikolajam par atbalstu un motivāciju pabeigt uzsākto. Paldies brālēnam Nikolajam Čašņikovam par tehnisko palīdzību promocijas darba sagatavošanā.

Milzīgs paldies Somijas meteoroloģijas institūta SILAM komandai un īpaši profesoram Mihailam Sofievam par iedvesmu, rūpēm, atsaucību, atbalstu un uzraudzību. Liels paldies Voeikova Galvenās ģeofizikas observatorijas profesoram Jevgēnijam L. Genihovičam par to, ka licis noticēt saviem spēkiem.

Paldies Latvijas Universitātes Ģeogrāfijas un Zemes zinātņu fakultātes mācībspēkiem, kolēģiem, kursabiedriem, īpaši bakalaura, maģistra un promocijas darba vadītājai asoc. prof. Laimdotai Kalniņai par grūdienu pareizā virzienā, par noderīgiem padomiem un sirsnību.

Īpašs paldies kolēģei Antrai Dūlei par sapratni un atsaucību.

REFERENCES / LITERATŪRAS SARAKSTS

Aguilera, F., Fornaciari, M., Ruiz-Valenzuela, L., Galan, C., Msallem, M., Dhiab, A., Ben, la Guardia, C.D. de, del Mar Trigo, M., Bonofiglio, T., Orlandi, F., 2014. Phenological models to predict the main flowering phases of olive (Olea europaea L.) along a latitudinal and longitudinal gradient across the Mediterranean region. Int. J. Biometeorol. 59, 629–641. doi:10.1007/s00484-014-0876-7

Aguilera, F., Ruiz-Valenzuela, L., 2014. Forecasting olive crop yields based on long-term aerobiological data series and bioclimatic conditions for the southern Iberian Peninsula. Spanish J. Agric. Res. 12, 215–224. doi:10.5424/sjar/2014121-4532

Albertine, J.M., Manning, W.J., Da Costa, M., Stinson, K.A., Muilenberg, M.L., Rogers, C.A., 2014. Projected carbon dioxide to increase grass pollen and allergen exposure despite higher ozone levels. PLoS One 9, 1–6. doi:10.1371/journal.pone.0111712

Barranco, D., Fernández-Escobar, R., Rallo, L., 2008. El Cultivo del olivo (8ª Ed). Madrid.

Bastl, K., Kmenta, M., Pessi, A.M., Prank, M., Saarto, A., Sofiev, M., Bergmann, K.C., Buters, J.T.M., Thibaudon, M., Jäger, S., Berger, U., 2016. First comparison of symptom data with allergen content (Bet v 1 and Phl p 5 measurements) and pollen data from four European regions during 2009–2011. Sci. Total Environ. 548–549, 229–235. doi:10.1016/j.scitotenv.2016.01.014

Carinanos, P., Emberlin, J., Galan, C., Dominguez-Vilches, E., 2000. Comparsion of two pollen counting methods of slides from a hirst type volumetric trap. Aerobiologia (Bologna). 16, 339–346.

Castellano-Méndez, M., Aira, M.J., Iglesias, I., Jato, V., González-Manteiga, W., 2005. Artificial neural networks as a useful tool to predict the risk level of Betula pollen in the air. Int. J. Biometeorol. 49, 310–6. doi:10.1007/s00484-004-0247-x

Chuine, I., Belmonte, J., 2004. Improving prophylaxis for pollen allergies: Predicting the time course of the pollen load of the atmosphere of major allergenic plants in France and Spain 43, 65–80. doi:10.1080/00173130410019163

Cunha, M., Ribeiro, H., 2015. Pollen-based predictive modelling of wine production: Application to an arid region. Eur. J. Agron. doi:10.1016/j.eja.2015.10.008

D'Amato, G., Cecchi, L., Bonini, S., Nunes, C., Annesi-Maesano, I., Behrendt, H., Liccardi, G., Popov, T., Van Cauwenberge, P., 2007. Allergenic pollen and pollen allergy in Europe. Allergy Eur. J. Allergy Clin. Immunol. 62, 976–990. doi:10.1111/j.1398-9995.2007.01393.x

Dahl, Å., Galan, C., Hajkova, L., Pauling, A., Sikoparija, B., Smith, M., Vokou, D., 2013. The onset, course and intensity of the pollen season, in: Sofiev, M., Bergmann, K.-C. (Eds.), Allergenic Pollen. A Review of the Production, Release, Distribution and Health Impacts. Springer, Dordrecht, Heidelberg, New York, London, pp. 29–70. doi:10.1007/978-94-007-4881-1

Dahl, A., Strandhede, S.-O., 1996. Predicting the intensity of the birch pollen season. Aerobiologia (Bologna). 12, 97–106.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. doi:10.1002/qj.828

Dhiab, A.B., Dominguez-Vilches, E., Dhiab, A.B., Mimoun, M.B., Oteros, J., Garcia-mozo, H., 2016. Modeling olive-crop forecasting in Tunisia. doi:10.1007/s00704-015-1726-1

Galán, C., Smith, M., Thibaudon, M., Frenguelli, G., Oteros, J., Gehrig, R., Berger, U., Clot, B., Brandao, R., 2014. Pollen monitoring: minimum requirements and reproducibility of analysis. Aerobiologia (Bologna). 30, 385–395. doi:10.1007/s10453-014-9335-5

Genikhovich, E., Pavlova, T.V., Kattsov, V.M., 2010. On complexing the ensemble of climate models (In Russian: O komplexirovanii ansamblya klimaticheskih modelej). Proc. Voeikov Main Geoiphysical Obs. 7, 28–46.

Gioulekas, D., Balafoutis, C., Damialis, A., Papakosta, D., Gioulekas, G., Patakas, D., 2004. Fifteen years' record of airborne allergenic pollen and meteorological parameters in Thessaloniki, Greece. Int. J. Biometeorol. 48, 128–36. doi:10.1007/s00484-003-0190-2

Gioulekas, D., Papakosta, D., Damialis, A., Spieksma, F.T.M., Giouleka, P., Patakas, D., 2004. Allergenic pollen records (15 years) and sensitization in patients with respiratory allergy in Thessaloniki, Greece. Allergy 174–184.

Grewling, Ł., Jackowiak, B., Nowak, M., Uruska, A., Smith, M., 2012. Variations and trends of birch pollen seasons during 15 years (1996–2010) in relation to weather conditions in Poznań (Western Poland). Grana 51:4, 280–292. doi:10.1080/00173134.2012.700727

Hatfield, J.L., Prueger, J.H., 2015. Temperature extremes: Effect on plant growth and development. Weather Clim. Extrem. 10, 4–10. doi:10.1016/j.wace.2015.08.001

Hedhly, A., Hormaza, J.I., Herrero, M., 2009. Global warming and sexual plant reproduction. Trends Plant Sci. 14, 30–36. doi:10.1016/j.tplants.2008.11.001

Helbig, N., Vogel, B., Vogel, H., Fiedler, F., 2004. Numerical modelling of pollen dispersion on the regional scale. Aerobiologia (Bologna). 3, 3–19.

Hirst, J.M., 1954. An automatic volumetric spore trap. Ann. Appl. Biol. 39, 257–265.

Huynen, M., Menne, B., Behrendt, H., Bertollini, R., Bonini, S., Brandao, R., Clot, B., D'Ambrosio, C., De Nuntiis, P., Ebi, K., Emberlin, J., Orbanne, E., Galán, C., Jäger, S., Kovats, S., Mandrioli, P., Martens, P., Menzel, A., Nyenzi, B., Rantio-Lehtimäki, A., Ring, J., Rybnicek, O., Traidl-Hoffmann, C., Van Vliet, A., Voigt, T., Weiland, S., Wickman, M., 2003. Phenology and Human Health: Allergic Disorders, World Health Organisation.

Inatsu, M., Kobayashi, S., Takeuchi, S., Ohmori, A., 2014. Statistical Analysis on Daily Variations of Birch Pollen Amount with Climatic Variables in Sapporo. Sola 10, 172–175. doi:10.2151/sola.2014-036

Jato, V., Rodríguez-Rajo, F.J., Alcázar, P., De Nuntiis, P., Galán, C., Mandrioli, P., 2006. May the definition of pollen season influence aerobiological results? Aerobiologia (Bologna). 22, 13–25. doi:10.1007/s10453-005-9011-x

Johansson, L., Epitropou, V., Karatzas, K., Karppinen, A., Wanner, L., Vrochidis, S., Bassoukos, A., Kukkonen, J., Kompatsiaris, I., 2015. Fusion of meteorological and air quality data extracted from the web for personalized environmental information services. Environ. Model. Softw. 64, 143–155. doi:10.1016/j.envsoft.2014.11.021

Kalyoncu, A., Qoplii, L., Selguk, Z., Emri, A., Kolagan, B., Kocabas, A., 1995. Survey of the allergic status of patients with bronchial asthma in Turkey: a multicenter study. Allergy 50, 451–456.

Kasprzyk, I., Rodinkova, V., Šaulienė, I., Ritenberga, O., Grinn-Gofron, A., Nowak, M., Sulborska, A., Kaczmarek, J., Weryszko-Chmielewska, E., Bilous, E., Jedryczka, M., 2015. Air pollution by allergenic spores of the genus Alternaria in the air of central and eastern Europe. Environ. Sci. Pollut. Res. 22, 9260–9274. doi:10.1007/s11356-014-4070-6

Laatikainen, T., Von Hertzen, L., Koskinen, J.P., Mäkelä, M.J., Jousilahti, P., Kosunen, T.U., Vlasoff, T., Ahlström, M., Vartiainen, E., Haahtela, T., 2011. Allergy gap between Finnish and Russian Karelia on increase. Allergy Eur. J. Allergy Clin. Immunol. 66, 886–892. doi:10.1111/j.1398-9995.2010.02533.x

Latałowa, M., Mietus, M., Uruska, A., 2002. Seasonal variations in the atmospheric Betula pollen count in Gdansk (southern Baltic coast) in relation to meteorological parameters. Aerobiologia (Bologna). 18, 33–43.

Limpert, E., Burke, J., Galan, C., Trigo, M.D.M., West, J.S., Stahel, W.A., 2008. Data, not only in aerobiology: how normal is the normal distribution? Aerobiologia (Bologna). 24, 121–124. doi:10.1007/s10453-008-9092-4

Linkosalo, T., Häkkinen, R., Hänninen, H., 2006. Models of the spring phenology of boreal and temperate trees: Is there something missing? Tree Physiol. 26, 1165–1172. doi:10.1093/treephys/26.9.1165

Linsser, G., 1867. Die periodischen Erscheinungen des Pflanzenlebens in ihrem Verhältniss zu den Wärmeerscheinungen. Mem. l'Acad. Imp. d. sc. St. Petersbg. XI, 1–44.

Masaka, K., 2001. Modelling the Masting Behaviour of Betula platyphylla var. japonica using the Resource Budget Model. Ann. Bot. 88, 1049–1055. doi:10.1006/anbo.2001.1547

Méndez, J., Comtois, P., Iglesias, I., 2005. Betula pollen: One of the most important aeroallergens in Ourense, Spain. Aerobiological studies from 1993 to 2000. Aerobiologia (Bologna). 21, 115–124. doi:10.1007/s10453-005-4181-0

Molnár, C., Borzan, B., Childers, N., Miriam, D., Renate, S., Kateřina, K., Estefanía, T.M., Lynn, B., Michèle, R., 2015. Written declaration: Rule 136 of the Rules of Procedure on allergic diseases in the European Union. Ugeskr. Laeger 0068/2015.

Muraro, A., et al., 2015. The European Academy of Allergy and Clinical Immunology (EAACI) Advocacy Manifesto Tackling the Allergy Crisis in Europe – Concerted Policy Action Needed.

Myszkowska, D., 2013. Prediction of the birch pollen season characteristics in Cracow, Poland using an 18-year data series. Aerobiologia (Bologna). 29, 31–44. doi:10.1007/s10453-012-9260-4

Negrini, A.C., Ariano, R., Delbono, G., Ebbli, A., Quaglia, A., Arobba, D., Allergologia, A., Paolo, O.S., Ligure, P., Sv, I.-P.L., 1992. Incidence of sensitisation to the pollens of Urticaceae (Parietaria), Poaceae and Oleaceae (Olea europaea) and pollen rain in Liguria (Italy). Aerobiologia (Bologna). 8, 355–358.

Newson, R.B., van Ree, R., Forsberg, B., Janson, C., Lötvall, J., Dahlén, S.-E., Toskala, E.M., Baelum, J., Brożek, G.M., Kasper, L., Kowalski, M.L., Howarth, P.H., Fokkens, W.J., Bachert, C., Keil, T., Krämer, U., Bislimovska, J., Gjomarkaj, M., Loureiro, C., Burney, P.G.J., Jarvis, D., 2014. Geographical variation in the prevalence of sensitization to common aeroallergens in adults: the GA(2) LEN survey. Allergy. doi:10.1111/all.12397

Nilsson, S., Persson, S., 1981. Tree pollen spectra in the stockholm region (Sweden), 1973–1980. Grana 20, 179–182. doi:10.1080/00173138109427661

Orlandi, F., Romano, B., Fornaciari, M., 2005a. Relationship between pollen emission and fruit production in olive (Olea europaea L.). Grana. 44, 98–103. doi:10.1080/00173130510010440

Orlandi, F., Ruga, L., Romano, B., Fornaciari, M., 2005b. An integrated use of aerobiological and phenological data to analyse flowering in olive groves. Grana. 44, 51–56. doi:10.1080/00173130510010431

Oteros, J., Galan, C., Alcazar, P., Dominguez-Vilches, E., 2013a. Quality control in bio-monitoring networks, Spanish Aerobiology Network. Sci. Total Environ. 443, 559–565. doi:10.1016/j.scitotenv.2012.11.040

Oteros, J., Garcia-Mozo, H., Hervas, C., Galan, C., 2013b. Biometeorological and autoregressive indices for predicting olive pollen intensity Biometeorological and autoregressive indices for predicting olive pollen intensity. Int. J. Biometeorol. 307–316. doi:10.1007/s00484-012-0555-5

Oteros, J., Garcia-Mozo, H., Vazquez, L., Mestre, A., Dominguez-Vilches, E., Galan, C., 2013. Modelling olive phenological response to weather and topography. Agric. Ecosyst. Environ. 179, 62–68.

Pearce, N., Ait-Khaled, N., Beasley, R., Mallol, J., Keil, U., Mitchel, E. et al., 2007. Worldwide trends in the prevalence of asthma symptoms: phase III of the International Study of Asthma and Allergies in Childhood (ISAAC). Thorax 62, 758–766.

Pereira, C., Valero, A., Loureiro, C., Dávila, I., 2006. Iberian study of Aeroallergens sensitization in allergig rhinitis. Eur. Ann. Allergy Clin. Immunol. 38, 186–194.

Potempski, S., Galmarini, S., 2009. Est modus in rebus: analytical properties of multi-model ensembles. Atmos. Chem. Phys. 9, 9471–9489.

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. doi:10.1016/j.agrformet.2013.08.003

Prasad, P.V.V., Craufurd, P.Q., Summerfield, R.J., 1999. Fruit number in relation to pollen production and viability in groundnut exposed to short episodes of heat stress. Ann. Bot. 84, 381–386. doi:10.1006/anbo.1999.0926

Puc, M., 2012. Artificial neural network model of the relationship between Betula pollen and meteorological factors in Szczecin (Poland). Int. J. Biometeorol. 56, 395–401. doi:10.1007/s00484-011-0446-1

Ranta, H., Hokkanen, T., Linkosalo, T., Laukkanen, L., Bondestam, K., Oksanen, A., 2008. Male flowering of birch: Spatial synchronization, year-to-year variation and relation of catkin numbers and airborne pollen counts. For. Ecol. Manage. 255, 643–650. doi:10.1016/j.foreco.2007.09.040

Ranta, H., Satri, P., 2007. Synchronized inter-annual fluctuation of flowering intensity affects the exposure to allergenic tree pollen in North Europe. Grana 46, 274–284. doi:10.1080/00173130701653079

Ring, J., Akdis, C., Behrendt, H., Lauener, R.P., Schäppi, G., Akdis, M., Ammann, W., de Beaumont, O., Bieber, T., Bienenstock, J., Blaser, K., Bochner, B., Bousquet, J., Crameri, R., Custovic, A., Czerkinsky, C., Darsow, U., Denburg, J., Drazen, J., de Villiers, E.M., Fire, A., Galli, S., Haahtela, T., zur Hausen, H., Hildemann, S., Holgate, S., Holt, P., Jakob, T., Jung, A., Kemeny, M., Koren, H., Leung, D., Lockey, R., Marone, G., Mempel, M., Menné, B., Menz, G., Mueller, U., von Mutius, E., Ollert, M., O'Mahony, L., Pawankar, R., Renz, H., Platts-Mills, T., Roduit, C., Schmidt-Weber, C., Traidl-Hoffmann, C., Wahn, U., Rietschel, E., 2012. Davos declaration: allergy as a global problem. Allergy 67, 141–3. doi:10.1111/j.1398-9995.2011.02770.x

Ritenberga, O., Sofiev, M., Kirillova, V., Kalnina, L., Genikhovich, E., 2016. Statistical modelling of non-stationary processes of atmospheric pollution from natural sources: example of birch pollen. Agric. For. Meteorol. 226–227, 96–107. doi:10.1016/j.agrformet.2016.05.016

Rönmark, E., Bjerg, A., Perzanowski, M., Platts-Mills, T., Lundbäck, B., 2009a. Major increase in allergic sensitization in school children from 1996 to 2006 in Northern Sweden. J. Allergy Clin. Immunol. 124, 1–19. doi:10.1016/j.jaci.2009.05.011

Scheifinger, H., Belmonte, J., Buters, J., Celenk, S., Damialis, A., Dechamp, C., Garcia-Mozo, H., Gehrig, R., Grewling, L., Halley, J.M., Hogda, K.-A., Jager, S., Karatzas, K., Karlsen, S.-R., Koch, E., Pauling, A., Peel, R., Sikoparija, B., Smith, M., Galan-Soldevilla, C., Thibaudon, M., Vokou, D., de Weger, L., 2013. Monitoring, Modelling and Forecasting of the Pollen Season, in: Sofiev, M., Bergmann, K.-C. (Eds.), Allergenic Pollen. p. 247.

Sikoparija, B., Skjøth, C.A., Celenk, S., Testoni, C., Abramidze, T., Alm Kübler, K., Belmonte, J., Berger, U., Bonini, M., Charalampopoulos, A., Damialis, A., Clot, B., Dahl, Å., de Weger, L.A., Gehrig, R., Hendrickx, M., Hoebeke, L., Ianovici, N., Kofol Seliger, A., Magyar, D., Mányoki, G., Milkovska, S., Myszkowska, D., Páldy, A., Pashley, C.H., Rasmussen, K., Ritenberga, O., Rodinkova, V., Rybníček, O., Shalaboda, V., Šaulienė, I., Ščevková, J., Stjepanović, B., Thibaudon, M., Verstraeten, C., Vokou, D., Yankova, R., Smith, M., 2016. Spatial and temporal variations in airborne Ambrosia pollen in Europe. Aerobiologia (Bologna). doi:10.1007/s10453-016-9463-1

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. doi:10.1007/s00484-012-0539-5

Simmons, A.J., Willett, K.M., Jones, P.D., Thorne, P.W., Dee, D.P., 2010. Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets. J. Geophys. Res. Atmos. 115, 1–21. doi:10.1029/2009JD012442

Sofiev, M., Berger, U., Prank, M., 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., 2015. MACC regional multi-model ensemble simulations of birch pollen dispersion in Europe. Atmos. Chem. Phys. 15, 8115–8130. doi:10.5194/acp-15-8115-2015

Sofiev, M., Siljamo, P., Ranta, H., Linkosalo, T., Jaeger, S., Rasmussen, A., Rantio-Lehtimaki, A., Severova, E., Kukkonen, J., 2012b. A numerical model of birch pollen emission and dispersion in the atmosphere. Description of the emission module. Int. J. Biometeorol. doi:10.1007/s00484-012-0532-z

Sofiev, M., Siljamo, P., Ranta, H., Rantio-Lehtimaki, A., 2006. Towards numerical forecasting of long-range air transport of birch pollen: theoretical considerations and a feasibility study. Int. J. Biometeorol. 50, 392–402. doi:10.1007/s00484-006-0027-x

Spieksma, F.T., 1990. Pollinosis in Europe: New observations and developments. Rev. Palaeobot. Palynol. 64, 35–40.

Spieksma, F.T.M., Corden, J.M., Detandt, M., Millington, W.M., Nikkels, H., Nolard, N., Schoenmakers, C.H.H., Wachter, R., de Weger, L.A., Willems, R., Emberlin, J., 2003. Quantitative trends in annual totals of five common airborne pollen types (Betula, Quercus, Poaceae, Urtica, and Artemisia), at five pollen-monitoring stations in western Europe. Aerobiologia (Bologna). 19, 171–184. doi:10.1023/B:AERO.0000006528.37447.15

Stach, A., Emberlin, J., Smith, M., Adams-Groom, B., Myszkowska, D., 2008. Factors that determine the severity of Betula spp. pollen seasons in Poland (Poznań and Krakow) and the United Kingdom (Worcester and London). Int. J. Biometeorol. 52, 311–21. doi:10.1007/s00484-007-0127-2

Stach, A., Smith, M., Prieto Baena, J.C., Emberlin, J., 2008. Long-term and short-term forecast models for Poaceae (grass) pollen in Poznań, Poland, constructed using regression analysis. Environ. Exp. Bot. 62, 323–332. doi:10.1016/j.envexpbot.2007.10.005

Storkey, J., Stratonovitch, P., Storkey, J., Stratonovitch, P., Chapman, D.S., Vidotto, F., Semenov, M.A., 2014. A Process-Based Approach to Predicting the Effect of Climate Change on the Distribution of an Invasive Allergenic Plant in Europe A Process-Based Approach to Predicting the Effect of Climate Change on the Distribution of an Invasive Allergenic Plant in Eu. PLoS One 1–7. doi:10.1371/journal.pone.0088156

Sunyer, J., Jarvis, D., Pekkanen, J., Chinn, S., Janson, C., Leynaert, B., 2004. Geographic variations in the effect of atopy on asthma in the European Community Respiratory Health Study. J. Allergy Clin. Immunol. 114, 1033–1039.

Toro, F., Recio, M., Trigo, M.D.M., Cabezudo, B., 1998. Predictive models in aerobiology: data transformation. Aerobiologia (Bologna). 14, 179–184.

Veriankaitė, L., Siljamo, P., Sofiev, M., Šaulienė, I., Kukkonen, J., 2009. Modelling analysis of source regions of long-range transported birch pollen that influences allergenic seasons in Lithuania. Aerobiologia (Bologna). 26, 47–62. doi:10.1007/s10453-009-9142-6

Yli-Panula, E., Fekedulegn, D.B., Green, B.J., Ranta, H., 2009. Analysis of Airborne Betula Pollen in Finland; a 31-Year Perspective. Int. J. Environ. Res. Public Heal. 6, 1706–1723. doi:10.3390/ijerph6061706

Zhang, Y., Bielory, L., Georgopoulos, P.G., 2014. Climate change effect on Betula (birch) and Quercus (oak) pollen seasons in the United States. Int. J. Biometeorol. 58, 909–919. doi:10.1007/s00484-013-0674-7

Zhang, Y., Isukapalli, S., Bielory, L., Georgopoulos, P., 2013. Bayesian Analysis of Climate Change Effects on Observed and Projected Airborne Levels of Birch Pollen. Atmos. Environ. (1994). 68, 64–73. doi:10.1016/j.atmosenv.2012.11.028

Ziello, C., Sparks, T.H., Estrella, N., Belmonte, J., Bergmann, K.C., Bucher, E., Brighetti, M.A., Damialis, A., Dedandt, M., Galan, C., Gehrig, R., Grewling, L., Guiterrez Bustillo, A.M., Hallsdottir, M., Kockhans-Bieda, C., Linares, C., Myskowska, D., Paldy, A., Sanches, A., Smith, M., Thibaudon, M., Travaglini, A., Uruska, A., Valencia Berra, R.M., Vokou, D., Wacher, R., de Weger, L.A., Menzel, A., 2012. Changes to Airborne Pollen Counts across Europe. PLoS One 7 (4). doi:10.1371/journal.pone.0034076

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. doi:10.5194/gmd-6-1961-2013

Zink, K., Vogel, H., Vogel, B., Magyar, D., Kottmeier, C., 2012. Modeling the dispersion of Ambrosia artemisiifolia L. pollen with the model system COSMO-ART. Int. J. Biometeorol. 56, 669–80. doi:10.1007/s00484-011-0468-8

I

Ritenberga, O., Sofiev, M., Kirillova, V., Kalnina, L.,
Genikhovich, E.

Statistical modelling of non-stationary processes of atmospheric pollution from natural sources: example of birch pollen

Agricultural and Forest Meteorology 226–227 (2016) 96–107

Agricultural and Forest Meteorology 226–227 (2016) 96–107

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Statistical modelling of non-stationary processes of atmospheric pollution from natural sources: example of birch pollen

Olga Ritenberga[a,*], Mikhail Sofiev[b], Victoria Kirillova[c], Laimdota Kalnina[a], Eugene Genikhovich[c]

[a] University of Latvia Faculty of Geography and Earth Sciences, Rainis bvld 19, Riga LV-1586, Latvia
[b] Finnish Meteorological Institute, Erik Palmenin aukio 1, 00560, Helsinki, Finland
[c] Voeikov Main Geophysical Observatory, Karbysheva street 7, St.Petersburg, 194021, Russia

ARTICLE INFO

Article history:
Received 16 November 2015
Received in revised form 19 May 2016
Accepted 22 May 2016

Keywords:
Statistical models
Non-stationary processes
Pollen forecasting
Betula spp

ABSTRACT

A statistical model for predicting daily mean pollen concentrations during the flowering season is constructed and its parameterization and application to birch pollen in Riga (Latvia) are discussed. The model involves several steps of transformations of both meteorological data and pollen observations, aiming at a normally distributed homogeneous stationary dataset with linearized dependencies between the transformed meteorological predictors and pollen concentrations. The data transformation includes normalization of daily mean birch pollen concentrations, a switch of the independent axis from time to heat sum, a projection of governing parameters to pollen concentrations, and a reduction of non-stationarity via removal of the mean pollen season curve. These transformations resulted in a substantial improvement of statistical features of the data and, consequently, a higher efficiency of statistical procedures and better scores of the model. The transformed datasets are used for the model construction via multi-linear regression. For the application in Riga, the model coefficients were calculated using 9 years of birch pollen observations. The model was evaluated using years withheld from the training dataset. The evaluation showed robust model performance with the overall Model Accuracy exceeding 80% and Odds Ratio = 30.

1. Introduction

Prevalence of sensitization to aeroallergens in Europe has continuously risen in recent decades (Ring et al., 2012) and presently exceeds 20% (Bauchau and Durham, 2004; Newson et al., 2014). The main aeroallergens in northern Europe are birch and grasses (D'Amato et al., 2007; Huynen et al., 2003) but hazel, alder, and mugwort, are also important (Akdis et al., 2014; D'Amato et al., 2007; Gadermaier et al., 2008). Adverse health effects of allergens can be significantly reduced by pre-emptive medication and behavioural adaptation. However, their planning requires reliable forecasts of expected pollen concentrations a few days ahead (Huynen et al., 2003).

Monitoring of atmospheric concentrations of pollen usually provides, with one-to-two weeks' delay, information about pollen in the air, which is used for pollen information and forecasting services in many European countries. Integrated pollen samplers allow collection of near-real-time data but they are expensive and so far not commonly used (Scheifinger et al., 2013).

Arguably, the most important parameter of the pollen season, from a practical point of view, is its starting date. It is followed by the season end date and the season-long sum of daily mean pollen concentrations – the seasonal pollen index (SPI). Determination of the start/end dates is not straightforward and several criteria were formulated (Jato et al., 2006). For instance, the season start (end) can be defined as a date when cumulative daily mean pollen concentration reaches 5% (95%) of the SPI (Taylor and Andersen, 2009). Specific numbers vary between the studies (Andersen, 1991; Emberlin et al., 2007; Pathirane, 1975; Smith et al., 2009; Stach et al., 2008b). However, this approach does not allow determination of the season start date until it already ends and the SPI is known, which makes it unsuitable for forecasting and real-time assessment purposes.

Two types of forecasting models are the most popular: regional-to-continental dispersion models and local-scale statistical models. Dispersion models (Helbig et al., 2004; Prank et al., 2013; Sofiev et al., 2015, 2012, 2006; Zink et al., 2013, 2012) are capable of predicting the pollen distribution over large areas but their accuracy strongly varies in space and depends on available information on plant distribution (Siljamo et al., 2012; Sofiev et al., 2015).

* Corresponding author.
E-mail addresses: olga.ritenberga@lu.lv (O. Ritenberga), mikhail.sofiev@fmi.fi (M. Sofiev), ego@main.mgo.rssi.ru (E. Genikhovich).

http://dx.doi.org/10.1016/j.agrformet.2016.05.016

The local-scale statistical models exploit empirically established relations between the predicted quantity (predictant, such as pollen concentration) and independent predictors (meteorological factors and historical pollen concentrations) (Rodriguez-Rajo 2000, as referred by Castellano-Méndez et al., 2005). The ways of establishing these relations vary widely and include: (i) *artificial neural networks* (Castellano-Méndez et al., 2005; Puc, 2012); (ii) *discriminant linear analysis* (Sánchez Mesa et al., 2005); (iii) *multiple regression analysis* (Inatsu et al., 2014); (iv) *autoregressive integrated moving average* (Rodríguez-Rajo et al., 2006; García-Mozo et al., 2014); (v) *Gamma, Gaussian*, or *logistic distribution models* (Kasprzyk and Walanus, 2014).

Such statistical models commonly predict the start of season (Emberlin et al., 2002; Frei and Gassner, 2008; Laaidi, 2001; Laaidi et al., 2003; Siniscalco et al., 2014), its peak and duration (Ribeiro et al., 2007), 10-day mean concentrations (Makra et al., 2011), etc. The "classical" task of forecasting the daily/hourly pollen concentrations a few days ahead is less common (Chapter 7 of Sofiev and Bergman, 2013).

A common methodological difficulty of statistical models is stringent requirements to features of the analysed data: (i) the least-square-error and correlation quality criteria are justified only for normally distributed stochastic processes; (ii) averaging- and correlation- based methods require the processes to be stationary and ergodic, so that the averaging over a statistical ensemble of realizations can be substituted with averaging over time; (iii) (multi-) linear regressions imply near-linear dependencies of the predicted quantity (predictant) and independent predictors. None of these assumptions is fulfilled in case of pollen modelling. Indeed, the mere existence of the start and the end of the season makes the process both non-stationary and non-ergodic. The distribution function of daily mean pollen concentrations is closer to a log-normal than to a normal distribution (Limpert et al., 2008; Toro et al., 1998). And relations between the main controlling parameters (temperature, humidity, precipitation, etc) and daily mean pollen concentrations are by no means linear. Finally, models are usually developed with observed meteorological data as an input, whereas their forecasting applications have to use weather model predictions. Consequences of such substitution are rarely analysed despite known limited accuracy of the meteorological models.

The most common method for data transformation applied in the literature is the log-transform of pollen concentrations (Masaka, 2001; Méndez et al., 2005) as a precaution against log-normally distributed data. A similar effect, albeit with thinner theoretical ground, is sometimes obtained via square root function (Toro et al., 1998) or employing the SPI as shown by Moseholm et al. (1987). Other difficulties have not been considered in the studies we are aware of.

The current study aims to construct a local statistical forecasting model that takes the above peculiarities of the pollen time series into account. The objective is *to predict the daily mean pollen concentrations using only basic meteorological parameters*. For this purpose, we shall modify the methodology developed at Voeikov Main Geophysical Observatory for urban air pollution (Berlyand, 1991; Genikhovich et al., 2004). The model will be applicable to any monitoring site location and any taxa, whose flowering is controlled by accumulated heat. For illustrative purposes, we shall use birch pollen observations obtained in Riga (Latvia).

2. Materials and methods

2.1. Study area

Pollen monitoring was conducted in the central part of Riga (N56°57′02", E24°06′57", Fig. 1). Due to its location next to the Gulf of Riga (Baltic Sea), the city has temperate (humid continental) climate with frequent rain. Mean annual temperature of air in Riga is 6.9 °C and annual precipitation is 708 mm. The monitoring site is located in the centre of the city and surrounded by parks. The coastline is 12 km NW of the site.

2.2. Pollen sampling

Birch pollen monitoring was performed with a *Burkard* 7-day pollen spore trap of the Hirst design (Hirst, 1954) from March to September during the period 2003–2014. The sampler was situated at a height of 23 m agl. Pollen was collected with an airflow rate of 10 l min^{-1}, airflow rate controlled with an external flow meter *G 1.6. BK Premagas*. Pollen counting was done by using Primo Star light microscope with a magnification of × 400 over 12 full vertical traverses. The method with 12 vertical traverses produces comparable results to other commonly used counting methods, such as 4 horizontal traverses (Carinanos et al., 2000), and can produce both daily average and bi-hourly values. It also examines the whole traverses of the slide rather than the central parts where most of the pollen is deposited, thereby avoiding overestimation (Carinanos et al., 2000; Käpylä and Penttinen, 1981). However, it can, in theory, miss short peaks in pollen concentrations if they fall between the counted transects (Carinanos et al., 2000; Comtois et al., 1999)

Birch (*Betula spp*) pollen in Latvia comes from more than 31 species, however widely distributed are only four of them – *B. nana L., B. pendula Roth, B.humilis Schrank and B.pubescens Ehrh*. Birch tree contribution within the forests near Riga is about 14.6%, it is the second largest taxon after Scots pine tree (73.3%). The major forests are located about 40 km NE of Riga (Fig. 1).

Due to similar characteristics, birch pollen from different species are not distinguishable by light microscopy. Therefore, all pollen grains were counted jointly as one general birch group *Betula spp*.

From the 12-year-long data set, we have randomly picked 9 years for the model construction and withheld 2009 (typical-to-low pollen season), 2012, and 2014 (both high pollen seasons) for its evaluation (Figs. 2 and 3 left-hand panel)

2.3. Meteorological information

Meteorological data for the years 2006–2014 included daily mean and maximum values of air temperature, relative humidity, wind speed and direction, atmospheric pressure a.s.l., cloud fraction, visibility, and daily sum of precipitation. They were extracted from the meteorological station "Riga- LU", as provided by Latvian Environment, Geology and Meteorology Centre. Meteorological observations were divided into two parts:

(i) data on wind speed and direction, visibility, precipitation and cloud fraction came from the same place where the aerobiological monitoring was performed;
(ii) data on air temperature, relative humidity of air, atmospheric pressure came from the second part of the monitoring station located at about 1 km distance but also in downtown Riga.

The meteorological parameters for the early years 2003–2005 came from another meteorological station in Riga, located about 9 km from the aerobiological monitoring site.

For modelling purposes, wind speed and direction were recalculated to longitudinal wind component U and latitudinal one V.

2.4. Sensitivity study and comparison with other approaches

To mimic the application of the developed model in the forecasting regime, a sensitivity study has been performed: the meteorological observations were replaced with the forecasted

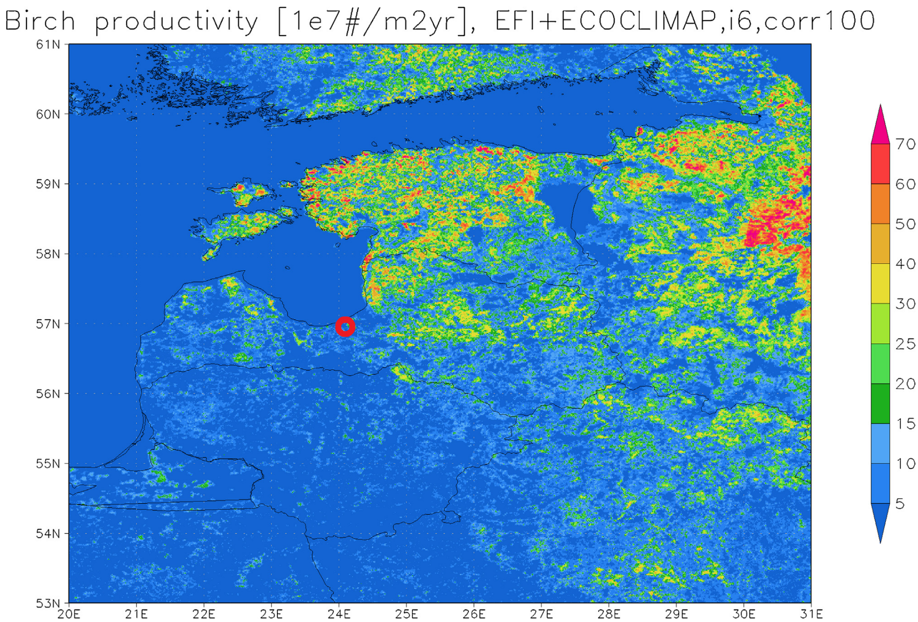


Fig. 1. Regional birch forest productivity map [10^7 pollen m^{-2} $year^{-1}$]. Riga is marked with a red circle. Updated map of Sofiev et al. (2006, 2012). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

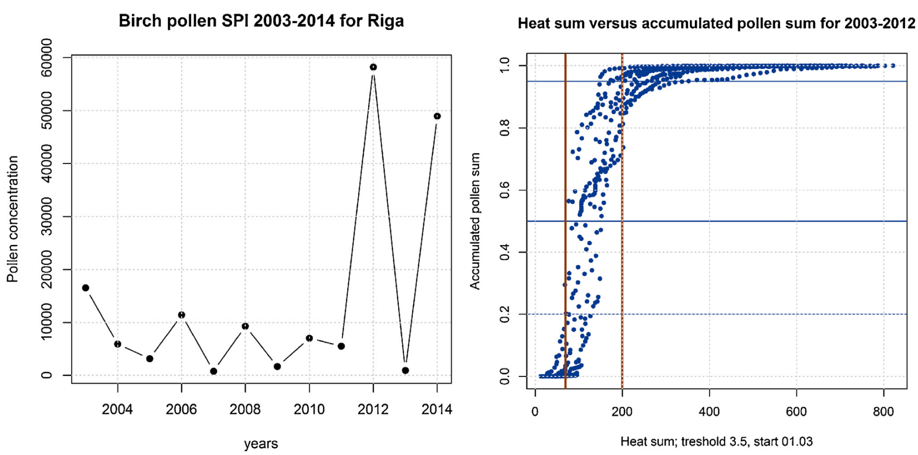


Fig. 2. Left-hand panel: Birch seasonal pollen index (SPI) for Riga 2003–2014. Right-hand panel: Seasons 2003–2012 presented in heat sum vs cumulative relative daily pollen concentration axes. The example is for cut-off temperature 3.5 °C and start of accumulation on 1 March (optimal parameters for Riga). Vertical lines correspond to 70 and 200 degree days (DD).

data. The weather forecasts for the years 2006–2010 were taken from operational archives of European Centre for Medium-Range Weather Forecast ECMWF. Importantly, the statistical model itself was not refitted, only the pollen forecasts were re-created.

For comparison with a deterministic pollen-forecasting model, we used the Europe-wide predictions of System for Integrated modeLling of Atmospheric coMposition (SILAM, Sofiev et al., 2012).

3. The model construction

3.1. The main components of the local-scale statistical forecasting model

We distinguished three separate parts of the local-scale short-term pollen forecasting model:

(i) Season timing model that includes calculation of the start and end of the season, and the season propagation from start to end. Outside the season, the daily mean pollen concentrations are assumed to be negligibly low.
(ii) Intra-seasonal model for daily mean pollen concentrations. The model predicts the daily mean pollen concentrations normalized with SPI.
(iii) Seasonal birch productivity model for predicting the SPI.

This division corresponds to a significant temporal separation of the processes and driving factors controlling each of the parts: the season start is almost completely decided by the current year's pre-season meteorology; the intra-season concentration variation is completely determined by the actual environmental conditions; the SPI is largely (although not completely) controlled by the previous year situation (Sofiev, submitted).

Below, we shall develop parts (i) and (ii), which are comparatively tightly connected, and outline the direction towards the more independent model part (iii) leaving its realization outside the current paper.

3.2. Season timing model

The current model for the season timing is based on the thermal-time approach, more specifically, on the double-threshold temperature model of Linkosalo et al. (2010). This model assumes a directly proportional relation between accumulated temperature sum and season stage. This function in a differential form has been used for construction of birch and olive source terms in the European scale SILAM model (Sofiev et al., 2012):

$$H(d) = \sum_{d=d_s}^{D} [\overline{T(d)} - T_{c-o}]_+ \qquad (1)$$

Here, H is temperature sum (heat sum), d is day, d_s is starting day of the heat accumulation, $T(d)$ is daily temperature, T_{c-o} is cut-off temperature (temperatures below this threshold are not summed up), $[x]_+$ equals 0 for $x<0$ and x for $x \geq 0$ (it excludes the temperatures below the cut-off level).

Eq. (1) has two adjustable parameters, which have to be identified for every specific location: cut-off temperature T_{c-o} and start day of accumulation D_s. Two heat sum thresholds describe the start and end of the flowering season: H_s, H_e. Specific values of these four parameters for Riga were selected in two steps.

At the first step, the cut-off temperature and accumulation start were identified from the requirement that the difference between the seasons taken as functions of heat sum must be minimal. In particular, the season propagation in different years should be similar, when considered as pollen count vs heat sum axes. In practice, the parameters were found simply by testing the different dates – 20.02, 01.03, 10.03 – and the cut-off temperatures from 0 °C to 5 °C with a step of 0.5°. The criterion was the smallest standard deviation at three levels of the accumulated pollen sum (a fraction of SPI): 0.2, 0.5 and 0.95 of the SPI for all years (horizontal lines in Fig. 2 right-hand panel). As a result, the start date of the 1st of March and the cut-off temperature of 3.5 °C were identified as the best combination for the heat sum calculation – for Riga. Interestingly, these values are identical to the parameters calculated from phenological data for the Europe-wide birch source term of SILAM.

Having the parameters of the heat sum Formula (1) identified, the thresholds for the start and end of the season were estimated from the SPI 5%-95% criterion (Nilsson and Persson, 1981): 70 degree days (DD) and 200 DD day, respectively, in Riga. Interestingly, these thresholds were not optimal for some of the years, with one of the explanations being the impact of long-distance pollen transport or unusually early or late flowering of birch in the region.

3.3. Season propagation model

The intra-seasonal forecasting model was parameterized using a multi-linear regression procedure, which relates the independent meteorological variables to the dependent daily mean pollen concentrations. The main efforts have been put into preparation of the input data and pollen concentrations, in order to eliminate or reduce the violations of the assumptions behind the fitting procedure (data inhomogeneity, non-ergodicity, non-stationarity, and non-linearity of the dependencies). These transformations are considered further one-by-one (Fig. 3).

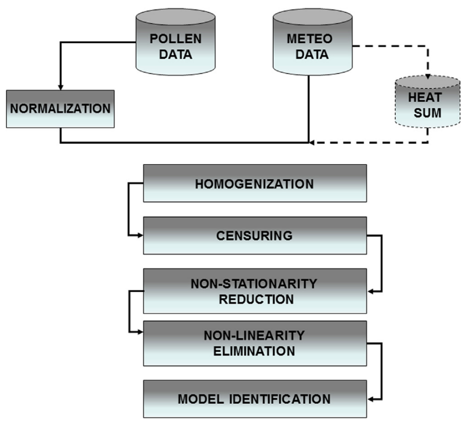


Fig. 3. Steps of the data transformation procedure used for constructing the intra-seasonal forecasting model.

3.3.1. Separation of productivity and intra-seasonal models: normalization of pollen concentration data

Productivity of the plants strongly varies from year to year (Fig. 2 left-hand panel) depending on several long-term and large-scale parameters, such as the previous-year intensity of flowering, the environmental conditions in winter, pre-seasonal spring conditions, etc (Dahl and Strandhede, 1996; Linkosalo et al., 2006; Stach et al., 2008a).

Characteristic temporal scales of these processes are completely different from the local short-term meteorological and biological processessuch as the intensity of flowering the previous year. Such scale separation allows for splitting of the problem: (i) determination of the general flowering intensity characterized by the SPI, (ii) intra-seasonal development of daily mean pollen concentrations.

In practice, these problems can be separated by a simple normalization of the daily concentration $C_i(d)$ (Fig. 4 left-hand panel) with the SPI of the corresponding year, for each considered year i,

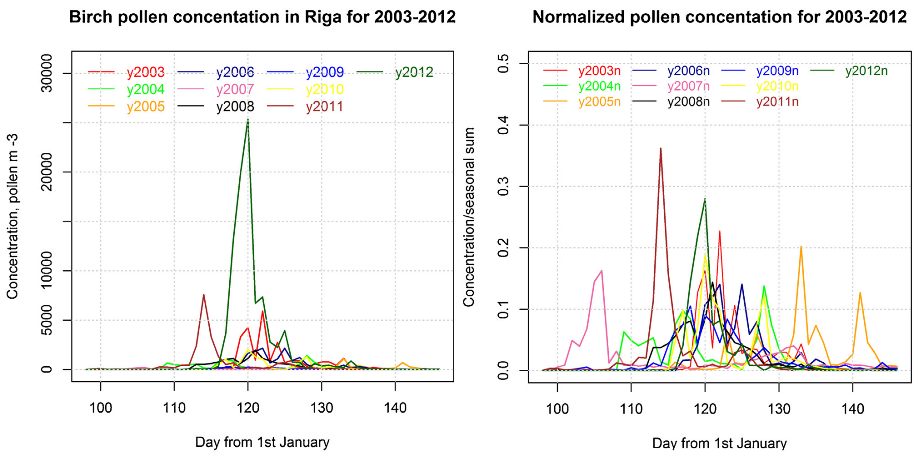


Fig. 4. Pollen seasons 2003–2012. Left-hand panel: Daily mean pollen concentrations, [pollen m^{-3}]. Right-hand panel: Daily mean pollen concentrations normalized with the SPI, relative unit.

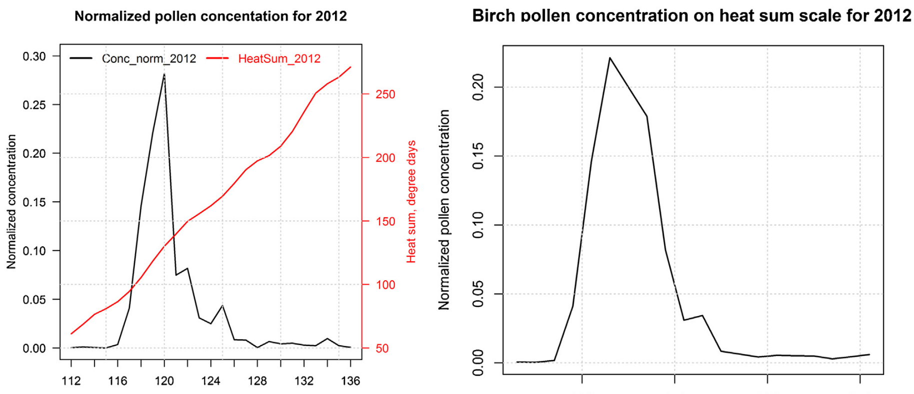


Fig. 5. Example of the time axis transformation of the pollen time series in Riga for 2012. Left-hand panel: original pollen data and heat sum as functions of time. Right-hand panel: same pollen time series as a function of the heat sum.

thus obtaining multi-annual concentration time series $c_i(d)$, which sum up to unity for each season (Fig. 4 right-hand panel):

$$c_i(d) = \frac{C_i(d)}{\sum_{d \in i} C_i(d)} \tag{2}$$

3.3.2. Homogenization of the annual data sets: switching the time axis to heat sum

One of the problems of the pollen data is the temporal inhomogeneity of the multi-annual time series: the season start, duration, and end are all different in different years (Fig. 4 right-hand panel). Therefore, the calendar time (day of year) is an inconvenient variable. It should be replaced with a variable, which: (i) is unambiguously connected with calendar time, (ii) is easy to calculate, (iii) encapsulates the main processes driving the flowering, so that the season start, end, and duration would be the same in all years.

The heat sum Eq. (1) satisfies all these requirements. It is a monotonic function of time, i.e. its inverse exists (except for very cold days, see the Discussion section). It can be trivially calculated from daily temperatures. As shown by Linkosalo et al. (2010), it controls the season propagation, whereas the connection to birch phenological phases was noted more than a century ago (Linsser, 1867).

From now on, all meteorological and pollen data are presented as functions of heat sum. For instance, all seasons have the same start and end: H_s and H_e. (for Riga, 70 and 200 DD-days, respectively).

An example of this transformation for the 2012 season is shown for Riga in Fig. 5.

3.3.3. Reducing the non-stationarity in pollen data and smoothing the intra-seasonal fluctuations

The primary source of non-stationarity of the pollen time series is the development of daily mean pollen concentrations from the start of the season towards its maximum and gradual fading out towards the end, as well as the tails before the start and after the end of the season.

Limiting the analysis within the $H_s - H_e$ range removes the tails. The remaining non-stationarity was removed by subtracting the multi-annual mean seasonal curve from the individual time series. There were three methods tested for generating this mean curve: arithmetic average, maximum, and median. The corresponding propagation curves were calculated from 9 years of pollen data. After that, each of the calculated mean-season curves was smoothed by (i) an "envelope" approximation, i.e. a convex curve enveloping the season curve, (ii) a beta-distribution approximation. An example for the arithmetic-average mean-season curve and its envelope is shown in Fig. 6, left-hand panel (red and green curves, respectively). The mean-season envelopes were subtracted from the actual concentrations, each year resulting in three data sets. The obtained pollen time series contained the pollen concentrations as deviations from the corresponding mean seasonal curves, as functions of heat sum (Fig. 6, right hand panel). These deviations are to be predicted from the meteorological parameters.

3.3.4. Linearization of relations between the independent and dependent variables

The last step of the data transformation before applying the multi-linear regression analysis was related to connection between pollen and meteorological data sets. Their relationship was linearized by projecting all meteorological data on pollen concentration:

$$m \to m_c : \quad m_c(m_i) = \overline{c(H, m \in m_i)} \tag{3}$$

Here m is the input meteorological quantity (temperature, wind speed, etc), m_i is a set of discrete ranges of m, for which the mean pollen concentration m_c is calculated. Each range m_i of actual values of m is associated with the pollen concentrations that were observed for meteorological conditions falling in this range. As a result, the set of $m_c(m_i)$ values represent the same meteorological quantity in terms of pollen concentrations (Fig. 7). As an example, in Riga the range of temperature 10–12 °C corresponds to −0.27 of deviation of the normalized daily mean pollen concentration from the multi-annual mean season, i.e. this temperature range corresponds to cold conditions and the pollen concentration during such days are usually low. Similarly, relative humidity ranging from 45% up to 70% (Fig. 7, right-hand panel) also corresponds to lower pollen concentrations.

Such transformation accommodates the bulk of non-linearity in the dependence of the pollen concentrations on the particular meteorological variable.

3.3.5. Computation of the regression coefficients

At the final phase of the model construction, the multiple regression analysis was made between the dependent variable (transformed pollen data) and the independent predictants (meteorological data projected to pollen concentration).

To limit the impact of the possible problem of multicollinearity, stepwise multiple regression analysis was used. This method was preferred over the standard multiple regression due to its inherent robustness to correlated and non-informative predictors, which are not included in the final regression equation. This method also automatically accounts for the local peculiarities of the predictor-predictant relationships, which might be overshadowed by noise if all predictors are involved at once. As any automated procedure, it also has its limitation: the parameters are selected on a formal basis with no relation to their physical meaning or dependencies.

For the transformed variables, the multiple regression model reads:

$$c(M_i) = a_0 + \sum_{j=1}^{n} a_j m_{cj}(m_i), \tag{4}$$

where c is normalized pollen concentration; a_0, a_i are regression coefficients, $M_i = \left\{m_{ij}, j = \overline{1, n}\right\}$ is a set of n meteorological variables considered at the regression step, i is the range index of each variable according to Eq. (3).

We used a step-wise forward method of multi-linear regression, which starts with no variables in the model and sequentially adds the variables that improve the model result by the largest increment at each step. The process ends when all possible variables are tested and no expansion gives further model improvement.

3.4. Inverse transformation of the predicted quantity

The above data transformations and the multi-linear regression create the forecasting model, which predicts the deviations of the normalized pollen concentrations from the multi-annual mean seasonal curve, as a function of the forecasted heat sum. The inverse transformation to the time axis and the absolute pollen concentrations includes three steps:

- Step 1. Conversion of the predicted deviations to the normalized concentrations.
- Step 2. Switch the argument from heat sum back to time.
- Step 3. De-normalization of the pollen concentrations

Step 1 is just a summation of the predicted deviations and the mean season propagation curve derived in section 3.3.3.

Step 2 requires conversion from heat sum to time, which is a discrete inverse of the Eq. (1) defined by the list of calendar days during the specific season and the heat sum value for each day. We are interested exclusively in periods of active flowering, in the case of Riga, the heat sum range from 70 to 200 DD. Outside this interval, the predicted pollen concentrations are set to zero. For the days when mean temperature is above the cut-off limit, the task is to find the day when the heat sum was below the given value at 00:00 and turned above it by 24:00. For cold days with the mean temperature below the cut-off limit, the model does not provide any predictions: their contribution to heat sum growth is zero and the strict monotonicity of the time – heat sum relation breaks. As a simple common-sense patch, one can put zero pollen predicted for such days. A more accurate way is to create the below-cut-off temperature range m_0 and project it to pollen concentrations following the Eq. (3). The obtained mean concentration will be the unbiased forecast for such days.

As the last step of the inverse transformation, the normalized concentration predictions should be multiplied with the predicted SPI for the pollen season, finally obtaining the absolute pollen concentrations predicted for the specified day.

4. Calculations

In this section, we finalize the application of the above generic procedure to the birch pollen observations in Riga and evaluate the obtained model against the control dataset of three years that were withheld from the model training.

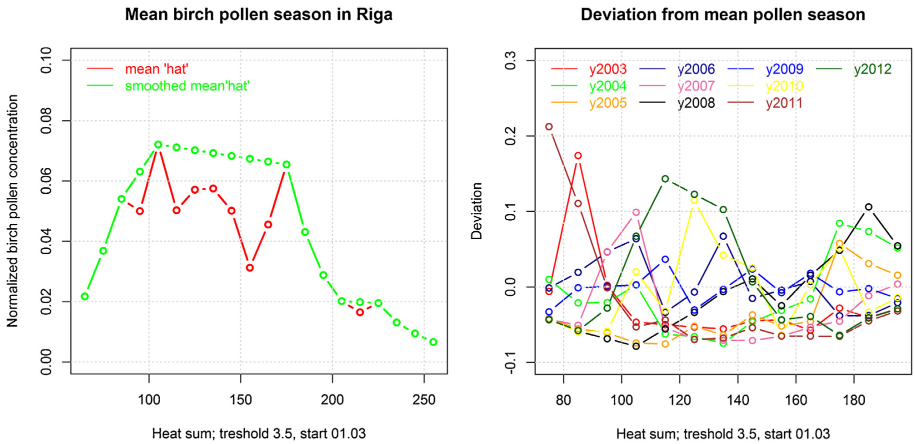


Fig. 6. Mean dependence of daily mean pollen concentration in Riga on heat sum during the season, mean over 2003–2012, and its envelope (left panel) and the deviation of the actual daily mean pollen concentrations from the mean envelope curve (right panel).

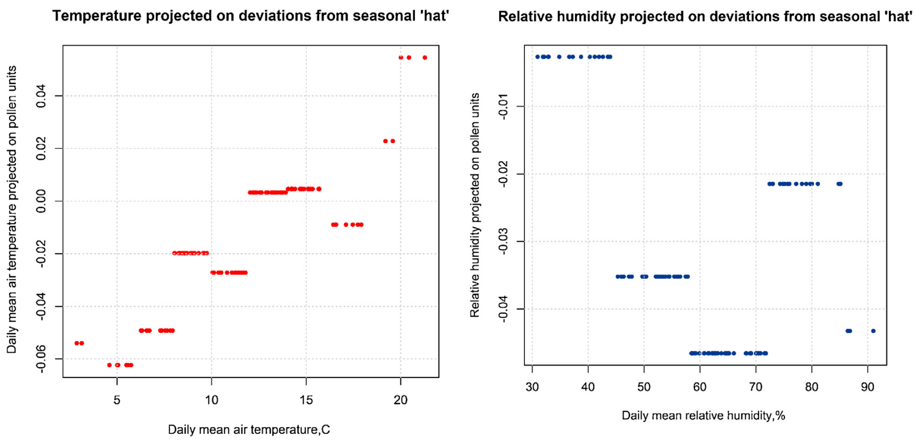


Fig. 7. Examples of linearization of the concentration dependence on meteorological variables: mean pollen concentrations as functions of temperature (left) and relative humidity (right).

4.1. Outcome of multi-linear regression

As stated in the section 3.3.3, we tested three possible ways of generating the mean-season propagation curve: arithmetic average with envelope smoothing, median with envelope smoothing, and beta-function. The best-performing approach for Riga appeared to be the arithmetic average, which showed the highest coefficient of determination of 0.44 for the 9 years used for the model training (Table 1).

Statistically significant ($p < 0.05$) influencing factors defined by linear regression are daily mean air temperature, daily mean cloud cover, daily sum of precipitation, and u-component of wind direction. The coefficients of the multi-linear regression for these parameters are shown in Table 2.

4.2. Evaluation against the control years

The model evaluation consists of two parts: (i) verification of the model ability to show the main seasonal characteristics – start, peaks, and end of pollen season, intra-seasonal variability, days of particularly high/low concentrations, etc; (ii) ability of the model to reproduce the exceedances of a practically relevant concentration threshold following Siljamo et al. (2012) – see Appendix A for the selected statistics.

During the model development, two randomly selected years were withheld from the model training procedure (2009 and 2012). In addition, the data of the 2014 season became available during the study – and were also used for evaluation.

Table 1
Results of the step-wise forward multi-linear regression for Riga.

Residuals					
Min	1Q	Median	3Q	Max	
−0.102049	−0.026	−0.0037	0.024	0.130	
Coefficients					
	Estimate	Std.Error	t value	Pr (> |t|)	
(Intercept)	0.073	0.023	3.21	0.002	**
Cloud cover	0.937	0.302	3.10	0.002	**
Dew point temperature	0.302	0.283	1.07	0.289	
Pressure tendency	0.936	0.908	1.03	0.305	
Sum of precipitation	0.580	0.170	3.42	0.0009	***
Relative humidity	0.358	0.299	1.20	0.235	
Daily mean temperature	0.454	0.211	2.15	0.034	*
Visibility	0.405	0.485	0.84	0.406	
Wind dir (U component)	0.501	0.203	2.47	0.015	*
Wind dir (V component)	0.078	0.588	0.13	0.895	
Daily mean wind speed	0.443	0.293	1.51	0.134	

Significance: '***' < 0.001 '**' < 0.01 '*' < 0.05 '.' < 0.1. Residual standard error: 0.043 on 105 degrees of freedom; Multiple R^2 = 0.44, Adjusted R^2: = 0.38; F-statistics 8.18 on 10 and 105 DF, p-value: 1.14e-09.

Table 2
The regression coefficient Eq. (4) for the transformed independent variables.

	variable/significance	coefficient
a_0	Intercept	+0.034
$a_1c(X_1)$	Cloud cover projected on pollen	+1.057
$a_2c(X_2)$	Sum of precipitation projected on pollen	+0.535
$a_3c(X_3)$	Mean daily temperature projected on pollen	+0.658
$a_4c(X_4)$	U wind component projected on pollen	+0.680

A summary of the model scores for the training and evaluation sets are shown in Table 3. It also contains the scores of the SILAM model (Sofiev et al., 2012, 2015), which was run for the same years over Europe, with subsequent extraction of the model predictions for Riga. The scores can also be compared with the European SILAM evaluation for 2006 (Siljamo et al., 2012).

As seen from Table 3, the model scores do not deteriorate from the learning set of years to the control one. The overall Model Accuracy (MA) is ~5% lower for the control dataset, whereas the Odds Ratio (OR) and other parameters were even slightly better. These differences are within the uncertainty of the evaluation procedure: the whole control dataset included only 70 days with concentration exceeding 50 pollen m^{-3}, about half of them reproduced correctly. About 170 days were below the threshold, 97% reproduced correctly.

The predicted and observed time series for 2009 and 2012 are shown in Fig. 8. The year 2009 appears to be a typical year, which is reproduced very well: with all the ambiguity of correlation coefficient for pollen time series, r = 0.94 still confirms accurate prediction of both season timing and intra-seasonal developments. For other years – 2012 and 2014–the correlation coefficient was 0.64 in both cases, which rather illustrates the model robustness to unusual years: both 2012 and 2014 were exceptionally high with regard to the SPI.

5. Discussion

5.1. Comparison with other models

Comparison of the scores of the new model with the existing instruments is not straightforward: this is the first local model for Riga. Comparing the models made for different locations makes little sense due to local specifics of the pollen season and large variation of the model accuracy from year to year (Fig. 8). The second difficulty is that only few models aim to forecast pollen concentrations day by day. One can, however, note some important similarities and differences of the new model and other approaches.

A quite common practice in short-term forecasting models is to "*nudge*" them towards observations via auto-regression methods: the forecast is taken as a function of the observed previous-day concentrations, see Inatsu et al. (2014) for birch in Sapporo, Japan, and Stach et al. (2008b) for grass in Poznan, Poland. The current model does not use the nudging for two reasons: (i) the birch season in Riga is short with very strong day-to-day variability; (ii) pollen data for *"today"* are never available when the forecast for "*tomorrow*" is to be generated. However, despite the exclusion of this strong predictor, the temporal correlation of the new model – with all reservations against this parameter – is within the same range from 0.64 to 0.94 as the scores of Inatsu et al. (2014) – from 0.6 up to 0.9–and Stach et al. (2008b) – from 0.6 to 0.7.

One can also compare the current scores with those of SILAM (Table 3). Expectedly, the new Riga model showed noticeably better performance – as one would expect for the strictly localized statistical development. The difference between the models can be illustrated via their comparison for 2014, one of the control years (Fig. 9). One can see that the season was long, with the tail originating from long-range transport and naturally missed by the local model but reproduced by SILAM. Conversely, the local model captured the main season with two strong peaks (albeit under-predicting both), whereas SILAM results were not so accurate: only one peak was predicted, shifted by one day from the actual one. Since the models are based on completely different principles, their joint application to the forecasting has a certain potential for further improvement of the forecast accuracy.

Table 3
Model scores in dichotomous classification task (probability of exceeding a threshold).[a]

Data set/model	MA	POD	FAR	POFD	OR
Training years (2003–2008, 2010- 2011), observed meteorology	0.86	0.77	0.23	0.05	13.32
Control years (2009, 2012, 2014), observed meteorology	0.83	0.87	0.13	0.03	29.81
SILAM (2003–2012, 2014), forecasted meteorology	0.86	0.72	0.28	0.09	7.73
Forecasted meteorology (years 2006–2010)	0.89	0.70	0.30	0.08	8.21

[a] Model accuracy (MA), Probability of Detection (POD), False Alarm Ratio (FAR), Probability of False Detection (POFD), Odds Ratio (OR).

Table 4
Comparison of scores of the model formally built using the non-transformed or partly transformed pollen time series for Riga.

Data used for MLR analysis	R^2	p-value	Parameters for model
Pollen and meteo data without any transformation	0.24	3e-06	T mean*, wind speed **
Homogenized data set (heat sum scaled data)	0.27	1e-03	Heat sum *, wind U **
Data set after full transformation (section 4)	0.44	1e-09	Cloud cover**, precipitation***, T mean*, wind U*

Signif. Codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.' 0.1 ' '.

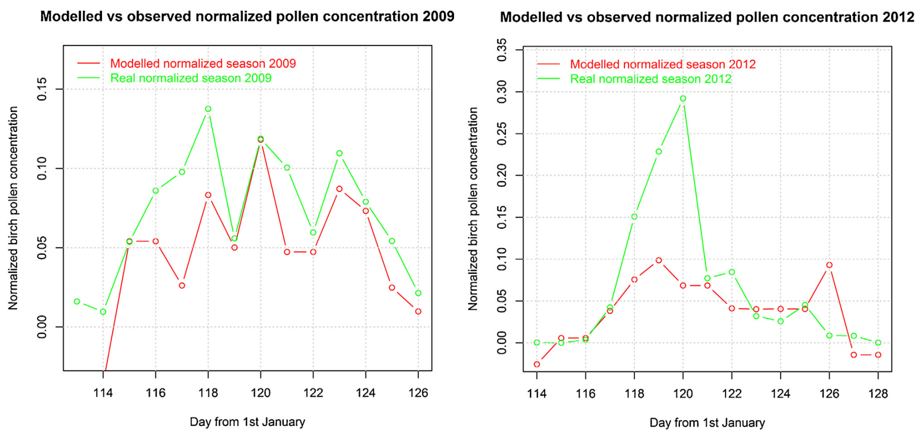


Fig. 8. Time series of normalized pollen concentrations are shown for 2009 (left) and 2012 (right).

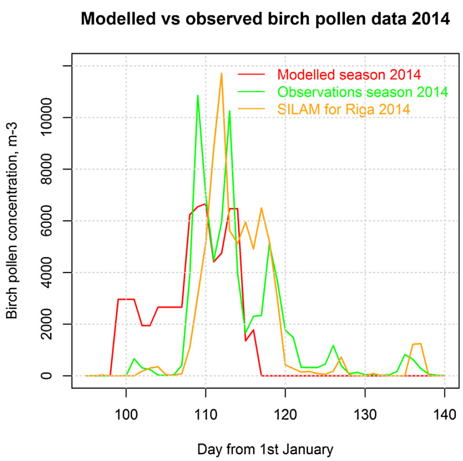


Fig. 9. Time series of birch pollen observations and predictions of the current statistical model and SILAM.

5.2. *Do we need this complexity?*

Complicated data transformations implemented in the current study is a requirement of statistical methods used for training and evaluation of the model. The normality of the data distribution, stationarity of the time series, ergodicity and near-linear dependencies of the variables are necessary for the model construction – in theory. One can, of course, formally apply the multi-linear regression to non-transformed datasets and still obtain a "technically-working" model with some forecasting skills. Its performance, however, will be substantially worse, penalized for every transformation step skipped (Table 4).

A separate issue is that the inverse transformations, being inevitable for practical use of the forecasts, also introduce errors. Indeed, the regression model is built and optimized in the transformed space of both predictors and predictants. Returning to the physical space for concentrations is a non-linear transformation, which may not preserve the optimality of the solution. Therefore, the scores for the transformed concentrations are higher than for concentrations after the inverse transformations. A similar effect was noticed by Toro et al. (1998).

5.3. *Forecasted meteorological parameters*

The model presented in the above sections was parameterised using the actual meteorological observations while its aim is to forecast the pollen concentrations. It puts a major limitation: the model cannot use any observations from the day to predict. In the case of pollen observations, this has been resolved by completely disregarding the pollen data from the regression model. For meteorology, it has to rely on Numerical Weather Prediction (NWP) forecasts. This requirement can potentially be a roadblock because the uncertainty of the weather forecast will add to the uncertainty of the pollen model itself and, in case of significant issues, disqualify the pollen prediction.

A test of model robustness with regard to the input data type was performed by providing the ECMWF meteorological forecasts as meteorological input for the years 2006–2010. Using the same model coefficients (Table 2) and transformations, these meteorological forecasts were used for predicting pollen concentrations following the above procedure (sections 3.3.4 and 3.4).

The test shows certain degradation of performance caused by replacing the observed meteorological data with the modelled ones (Fig. 10). This is also confirmed by the formal scores (Table 3, last row).

Comparison of the scores shown by various models and model setups in Table 3, as well as qualitative analysis of the time series Fig. 10, suggests that the main contributor to the somewhat lower scores of the model with the forecasted meteorology come from the slightly different predictions of the season timing. Indeed, even a half-degree bias between the predicted grid-cell average temperature and its point observation downtown of Riga would result in a few days of a shift of the season. Since the model was calibrated with the observed temperature, this shift manifests itself via higher uncertainties for the forecasted meteorology case. Importantly, the overall accuracy stays essentially the same (MA > 0.8 for all cases)

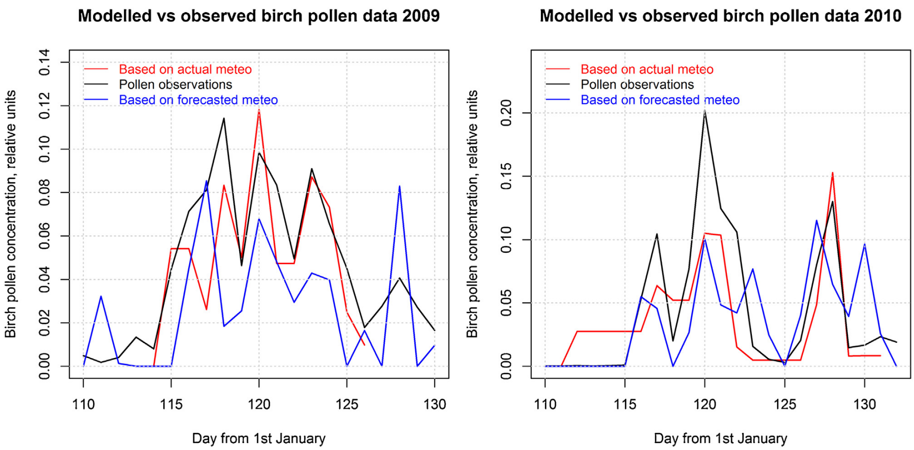


Fig. 10. Predicted pollen concentration based on actual and forecasted meteorological data from the years 2009 (left panel) and 2010 (right panel).

and only the False Alarm Ratio grows, by ~50%, reflecting the out-of-season time mis-predicted to be inside the flowering period. This also affects the OR, which is still high (OR > 8) and comparable with the scores for the learning years (OR > 13).

One has to keep in mind that utilization of point-wise meteorological observations leads to a disarray in the spatial scales of the observations and the pollen model. Indeed, in case of Riga, coastal and urban effects limit the representativeness of the downtown meteorological observations by just a few km (i.e., size of the city and distance to the shore). Conversely, pollen concentrations are largely decided by the birch forests around the city, at a distance up to a few tens of km or even more. The point meteorological observations downtown do not conform to such a non-local problem. In that sense, meteorological forecasts with spatial resolutions of ~15 km (the IFS resolution in 2006–2010) may be better.

The above model configurations are evidently not the only ones imaginable. For instance, one can perform a model fit using the above NWP model dataset, which would allow the regression to compensate for some of the meteorological model uncertainties. Alternatively, one can consider the lagging-data model, which would use today's meteorological observations as input for tomorrow's pollen forecast. It is also possible to include SILAM pollen forecast as one of predictors, and/or the lagging pollen observations (even week-old pollen data might still appear useful). One can also generate the so-called "hindcast" by predicting the pollen concentrations for the current day using either meteorological analysis or short-term forecasts. Each of these configurations has its own area of application and a set of uncertainties that will be inherited by the model. The list can be continued and it is not the purpose of this paper to explore all of these possibilities. The goal here is to present the methodology, which makes the day-by-day pollen forecast possible, and to provide an example of the most-basic case when the input meteorological predictors are taken from the local observations of the same day as the predicted pollen concentration. Other combinations of input data and predicted variables can be explored following the procedure presented in the previous sections.

5.4. Towards the SPI prediction model

The only part of the model left outside this paper is the prediction of the multi-annual variability of the seasonal pollen index. According to the publication mainstream, such models can be successfully developed following similar statistical procedures applied to monthly and annual data. One such formula was developed in Japan (Masaka, 2001) and adapted to Finland by Ranta et al. (2005).

Unfortunately, the 12 years of data available for Riga is not sufficient for training the model: it requires a longer time series. As an illustration, the SPI in Riga (Fig. 2) suggests a sudden step-change during the last years. However, consideration of longer series for Finland shows that, apart from 2012 and 2014, years with very high SPI were also observed before 2000, so that 2012 and 2014 highs are by no means a principal shift in the pollen pattern – but rather the very inter-annual variability that needs to be predicted.

With long time series existing in several European countries, the work will be continued as a joint effort of a multi-national team.

6. Conclusions

The current paper presents a local-scale statistical model for pollen forecasting. The model is formulated in generic terms and its methodology is applicable at any location. Specific application is demonstrated for Riga, Latvia.

The model construction is based on the multi-step transformations of the input meteorological data. This ensures compliance with the requirements of the applied statistical methods: normally distributed homogeneous stationary data with linearized dependencies between the transformed meteorological predictors and pollen concentrations. Being important for success of the regression coefficients identification, this procedure has never been developed for aerobiological data previously.

The pollen data transformations include normalization, homogenization, censoring, and reduction of the non-stationarity. The meteorological data were also projected to the pre-selected pollen concentration ranges, thus sharply reducing the non-linearity of their dependencies. The transformed datasets were supplied to the multi-step regression procedure, which determined the regression coefficients, thus concluding the model training.

The developed model was applied in Riga in several configurations and its scores were compared with those of other approaches reported in the literature for other locations, as well as with the SILAM performance for Riga. The model demonstrated solid performance and stability in the control years, with the model accuracy MA exceeding 80% and the odds ratio OR = 30.

Replacement of the observed data with weather forecasts resulted in a limited deterioration of the scores, primarily due to a few days of shift in the season timing (MA > 80% and OR > 8). This experiment highlighted the importance of the model calibration with regard to the input data representing the real-life application conditions.

Acknowledgements

The study has been performed within the scope of Academy of Finland APTA and TEKES SmartPollen projects. Support of FP7-MACC project is kindly acknowledged. This study was also supported by Science-based funding of Latvian Ministry of Education and Science via "*Attraction of Human Resources to Development of Scientific Study in the area of Earth and Environmental Sciences*" programme.

We are very thankful to Jānis Jātnieks and Didzis Elferts for technical help during the model development.

Appendix A.

Following Siljamo et al. (2012), the model accuracy was quantified using two ranks: concentrations below and above 50 pollen m^{-3}. For these two ranks, N_{M_low} and N_{M_high} are the number of daily observations with mean model-predicted concentration below and above 50 grains m^{-3}, respectively. Similarly, N_{O_low} and N_{O_high} are the number of daily observed concentrations below and above 50 pollen m^{-3}, respectively. Finally, $N_{M\_low\_O\_low}$, $N_{M\_low\_O\_high}$, $N_{M\_high\_O\_low}$, $N_{M\_high\_O\_high}$ represent all combinations of the relation between the model predictions and observations, and N_{total} is the total number of observed days.

Based on the above notations, the performance of the model was quantified using the following quality criteria:

An overall model accuracy (*MA*) represents the fraction of correct forecasts:

$$MA = \frac{N_{M\_low\_O\_low} + N_{M\_high\_O\_high}}{N_{total}} \quad \text{(A1)}$$

Hit Rate (also called Probability of Detection *POD*) is the fraction of "high" forecasts appeared to be correct:

$$POD, HR = \frac{N_{M\_high\_O\_high}}{N_{M\_high\_O\_low} + N_{M\_high\_O\_high}} \quad \text{(A2)}$$

False alarms can be represented by two quantities: the False Alarm Ratio (*FAR*), i.e., the fraction of the "high" forecasts that appear to be incorrect:

$$FAR = \frac{N_{M\_high\_O\_low}}{N_{M\_high\_O\_low} + N_{M\_high\_O\_high}}, \quad \text{(A3)}$$

and a Probability of False Detection, which shows the fraction of low-concentration days predicted as "high":

$$POFD = \frac{N_{M\_high\_O\_low}}{N_{M\_high\_O\_low} + N_{M\_low\_O\_low}} \quad \text{(A4)}$$

Finally, the Odds Ratio shows how much higher are the chances to get the "high" than "low" day if the model prediction is "high":

$$OR = \frac{POD}{POFD} \quad \text{(A5)}$$

A similar quantity is the difference *POD-POFD* known also as the Hansen-Kuiper or True Skill Score. However, its meaning is the same and below we shall use the *OR*.

References

Akdis, A., Agache, I., Asher, M.I., Genuneit, J., Jarvis, D., Flohr, C., Burney, P., Holt, P.G., Smith, H. (Eds.), 2014. EAACI, p. 179.
Andersen, T.B., 1991. A model to predict the beginning of the pollen season. Grana 30, 269–275, http://dx.doi.org/10.1080/00173139109427810.
Bauchau, V., Durham, S.R., 2004. Prevalence and rate of diagnosis of allergic rhinitis in Europe. Eur. Respir. J. 24, 758–764, http://dx.doi.org/10.1183/09031936.04.00013904.
Berlyand, M.E., 1991. Prediction and Regulation of Air Pollution. Springer, Netherlands, http://dx.doi.org/10.1007/978-94-011-3768-3.
Carinanos, P., Emberlin, J., Galan, C., Dominguez-Vilches, E., 2000. Comparison of two pollen counting methods of slides from a hirst type volumetric trap. Aerobiologia (Bologna) 16, 339–346.
Castellano-Méndez, M., Aira, M.J., Iglesias, I., Jato, V., González-Manteiga, W., 2005. Artificial neural networks as a useful tool to predict the risk level of Betula pollen in the air. Int. J. Biometeorol. 49, 310–316, http://dx.doi.org/10.1007/s00484-004-0247-x.
Comtois, P., Alcazar, P., Daniel, N., 1999. Pollen counts statistics and its relevance to precision. Methodology, 19–28.
D'Amato, G., Cecchi, L., Bonini, S., Nunes, C., Liccardi, G., Popov, T., Cauwenberge, P., Van, 2007. Allergenic pollen and pollen allergy in Europe. Allergy, 976–990, http://dx.doi.org/10.1111/j.1398-9995.2007.01393.x.
Dahl, A., Strandhede, S., 1996. Aerobiologia Predicting the intensity of the birch pollen season. Aerobiologia (Bologna) 12, 97–106.
Emberlin, J., Dedandt, M., Gehrig, R., Jaeger, S., Nolard, N., Rantio-Lehtimäki, A., 2002. Responses in the start of Betula (birch) pollen seasons to recent changes in spring temperatures across Europe. Int. J. Biometeorol. 46, 159–170, http://dx.doi.org/10.1007/s00484-002-0139-x.
Emberlin, J., Smith, M., Close, R., Adams-Groom, B., 2007. Changes in the pollen seasons of the early flowering trees Alnus spp. and Corylus spp. in Worcester United Kingdom. 1996-2005. Int. J. Biometeorol. 51, 181–191, http://dx.doi.org/10.1007/s00484-006-0059-2.
Frei, T., Gassner, E., 2008. Climate change and its impact on birch pollen quantities and the start of the pollen season an example from Switzerland for the period 1969–2006. Int. J. Biometeorol. 52, 667–674, http://dx.doi.org/10.1007/s00484-008-0159-2.
Gadermaier, G., Wopfner, N., Wallner, M., Egger, M., Didierlaurent, A., Regl, G., Aberger, F., Lang, R., Ferreira, F., Hawranek, T., 2008. Array-based profiling of ragweed and mugwort pollen allergens, 1543–1549, http://dx.doi.org/10.1111/j.1398-9995.2008.01780.x.
García-Mozo, H., Yaezel, L., Oteros, J., Galán, C., 2014. Statistical approach to the analysis of olive long-term pollen season trends in southern Spain. Sci. Total Environ. 473–474, 103–109, http://dx.doi.org/10.1016/j.scitotenv.2013.11.142.
Genikhovich, E.L., Sonkin, L.R., Kirillova, V.I., 2004. A statistical prognostic model for daily maxima of concentrations of urban air pollutants. 9th International Conference on Harmonisation Within Atmospheric Dispersion Modelling for Regulatory Purposes, 34–39.
Helbig, N., Vogel, B., Vogel, H., Fiedler, F., 2004. Numerical modelling of pollen dispersion on the regional scale. Aerobiologia (Bologna) 3, 3–19.
Hirst, J.M., 1954. An automatic volumetric spore trap. Ann. Appl. Biol. 39, 257–265.
Huynen, M., Menne, B., Behrendt, H., Bertollini, R., Bonini, S., Brandao, R., Clot, B., D'Ambrosio, C., De Nuntiis, P., Ebi, K., Emberlin, J., Orbanne, E., Galán, C., Jäger, S., Kovats, S., Mandrioli, P., Martens, P., Menzel, A., Nyenzi, B., Rantio-Lehtimäki, A., Ring, J., Rybnicek, O., Traidl-Hoffmann, C., Van Vliet, A., Voigt, T., Weiland, S., Wickman, M., 2003. Phenology and Human Health: Allergic Disorders. World Health Organisation.
Inatsu, M., Kobayashi, S., Takeuchi, S., Ohmori, A., 2014. Statistical analysis on daily variations of birch pollen amount with climatic variables in Sapporo. Sola 10, 172–175, http://dx.doi.org/10.2151/sola.2014-036.
Jato, V., Rodríguez-Rajo, F.J., Alcázar, P., De Nuntiis, P., Galán, C., Mandrioli, P., 2006. May the definition of pollen season influence aerobiological results? Aerobiologia (Bologna) 22, 13–25, http://dx.doi.org/10.1007/s10453-005-9011-x.
Käpylä, M., Penttinen, A., 1981. An evaluation of the microscopical counting methods of the tape in hirst-burkard pollen and spore trap. Grana 20, 131–141, http://dx.doi.org/10.1080/00173138109427653.
Kasprzyk, I., Walanus, A., 2014. Gamma, Gaussian and logistic distribution models for airborne pollen grains and fungal spore season dynamics. Aerobiologia (Bologna), http://dx.doi.org/10.1007/s10453-014-9332-8.
Laaidi, M., Thibaudon, M., Besancenot, J.-P., 2003. Two statistical approaches to forecasting the start and duration of the pollen season of Ambrosia in the area of Lyon (France). Int. J. Biometeorol. 48, 65–73, http://dx.doi.org/10.1007/s00484-003-0182-2.
Laaidi, M., 2001. Regional variations in the pollen season of Betula in Burgundy: two models for predicting the start of the pollination. Aerobiologia (Bologna), 247–254.
Limpert, E., Burke, J., Galan, C., Trigo, M.D.M., West, J.S., Stahel, W. a., 2008. Data, not only in aerobiology: how normal is the normal distribution? Aerobiologia (Bologna) 24, 121–124, http://dx.doi.org/10.1007/s10453-008-9092-4.
Linkosalo, T., Häkkinen, R., Hänninen, H., 2006. Models of the spring phenology of boreal and temperate trees: is there something missing? Tree Physiol. 26, 1165–1172, http://dx.doi.org/10.1093/treephys/26.9.1165.
Linkosalo, T., Ranta, H., Oksanen, A., Siljamo, P., Luomajoki, A., Kukkonen, J., Sofiev, M., 2010. A double-threshold temperature sum model for predicting the

flowering duration and relative intensity of Betula pendula and B. pubescens. Agric. For. Meteorol., 6–11, http://dx.doi.org/10.1016/j.agrformet.2010.08.007.

Linsser, C., 1867. Die periodischen Ercheinungen des Pflanzenlebens in ihrem Verhältniss zu den Wärmeerscheinungen, VIIE Serie. Mem. L'Académie Impériale des Sci., St.-Pétersberg, pp. 44.

Méndez, J., Comtois, P., Iglesias, I., 2005. Betula pollen: One of the most important aeroallergens in Ourense, Spain. Aerobiological studies from 1993 to 2000. Aerobiologia (Bologna) 21, 115–124, http://dx.doi.org/10.1007/s10453-005-4181-0.

Makra, L., Matyasovszky, I., Thiubadon, M., Bonini, M., 2011. Forecasting ragweed pollen characteristics with nonparametric regression methods over the most polluted areas in Europe Forecasting ragweed pollen characteristics with nonparametric regression methods over the most polluted areas in Europe. Int. J. Biometeorol., 361–371, http://dx.doi.org/10.1007/s00484-010-0346-9.

Masaka, K., 2001. Modelling the Masting Behaviour of Betula platyphylla var. japonica using the Resource Budget Model. Ann. Bot. 88, 1049–1055, http://dx.doi.org/10.1006/anbo.2001.1547.

Moseholm, L., Weeke, E.R., Petersen, B.N., 1987. Forecast of pollen concentrations of Poaceae (grasses) in the air by time series analysis. Poll.et Spores 19, 131–141.

Newson, R.B., van Ree, R., Forsberg, B., Janson, C., Lötvall, J., Dahlén, S.-E., Toskala, E.M., Baelum, J., Brożek, G.M., Kasper, L., Kowalski, M.L., Howarth, P.H., Fokkens, W.J., Bachert, C., Keil, T., Krämer, U., Bislimovska, J., Gjomarkaj, M., Loureiro, C., Burney, P.G.J., Jarvis, D., 2014. Geographical variation in the prevalence of sensitization to common aeroallergens in adults: the GA(2) LEN survey. Allergy, http://dx.doi.org/10.1111/all.12397.

Nilsson, S., Persson, S., 1981. Tree pollen spectra in the stockholm region (sweden), 1973–1980. Grana 20, 179–182, http://dx.doi.org/10.1080/00173138109427661.

Pathirane, L., 1975. Graphical determination of the main pollen season. Grana 15, 145–147.

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, http://dx.doi.org/10.1016/j.agrformet.2013.08.003.

Puc, M., 2012. Artificial neural network model of the relationship between Betula pollen and meteorological factors in Szczecin (Poland). Int. J. Biometeorol. 56, 395–401, http://dx.doi.org/10.1007/s00484-011-0446-1.

Ranta, H., Oksanen, A., Hokkanen, T., Bondestam, K., Heino, S., 2005. Masting by Betula-species; applying the resource budget model to north European data sets. Int. J. Biometeorol. 49, 146–151, http://dx.doi.org/10.1007/s00484-004-0228-0.

Ribeiro, H., Cunha, M., Abreu, I., 2007. Definition of main pollen season using a logistic model. Ann. Agric. Environ. Med. 14, 259–264.

Ring, J., Akdis, C., Behrendt, H., Lauener, R.P., Schäppi, G., Akdis, M., Ammann, W., de Beaumont, O., Bieber, T., Bienenstock, J., Blaser, K., Bochner, B., Bousquet, J., Crameri, R., a, Custovic, Czerkinsky, C., Darsow, U., Denburg, J., Drazen, J., de Villiers, E.M., A, Fire, Galli, S., Haahtela, T., zur Hausen, H., Hildemann, S., Holgate, S., Holt, P., Jakob, T., A, Jung, Kemeny, M., Koren, H., Leung, D., Lockey, R., Marone, G., Mempel, M., Menné, B., Menz, G., Mueller, U., von Mutius, E., Ollert, M., O'Mahony, L., Pawankar, R., Renz, H., Platts-Mills, T., Roduit, C., Schmidt-Weber, C., Traidl-Hoffmann, C., Wahn, U., Rietschel, E., 2012. Davos declaration: allergy as a global problem. Allergy 67, 141–143, http://dx.doi.org/10.1111/j.1398-9995.2011.02770.x.

Rodríguez-Rajo, F.J., Valencia-barrera, R.M., Vega-maray, A.M., Suárez, F.J., Fernández-gonzález, D., Jato, V., 2006. Prediction Of Airborne Alnus Pollen Concentration, 25–32.

Sánchez Mesa, J.A., Galán, C., Hervás, C., 2005. The use of discriminant analysis and neural networks to forecast the severity of the Poaceae pollen season in a region with a typical Mediterranean climate. Int. J. Biometeorol. 49, 355–362, http://dx.doi.org/10.1007/s00484-005-0260-8.

Scheifinger, H., Belmonte, J., Buters, J., Celenk, S., Damialis, A., Dechamp, C., Garcia-Mozo, H., Gehrig, R., Grewling, L., Halley, J.M., Hogda, K.-A., Jager, S., Karatzas, K., Karlsen, S.-R., Koch, E., Pauling, A., Peel, R., Sikoparija, B., Smith, M., Galan-Soldevilla, C., Thibaudon, M., Vokou, D., de Weger, L., 2013. Monitoring, Modelling and Forecasting of the Pollen Season, in: Sofiev, M., Bergmann, K.-C. (Eds.). Allergenic Pollen., p247.

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. 57, 125–136, http://dx.doi.org/10.1007/s00484-012-0539-5.

Siniscalco, C., Migliavacca, M., Mercalli, L., Richardson, A.D., 2014. Models to predict the start of the airborne pollen season. Int. J. Biometeorol., http://dx.doi.org/10.1007/s00484-014-0901-x.

Smith, M., Emberlin, J., Stach, A., Rantio-Lehtimäki, A., Caulton, E., Thibaudon, M., Sindt, C., Jäger, S., Gehrig, R., Frenguelli, G., Jato, V., Rajo, F.J.R., Alcázar, P., Galán, C., 2009. Influence of the North Atlantic Oscillation on grass pollen counts in Europe. Aerobiologia (Bologna) 25, 321–332, http://dx.doi.org/10.1007/s10453-009-9136-4.

Sofiev, M., Siljamo, P., Ranta, H., Rantio-Lehtimaki, A., 2006. Towards numerical forecasting of long-range air transport of birch pollen: theoretical considerations and a feasibility study. Int. J. Biometeorol. 50, 392–402, http://dx.doi.org/10.1007/s00484-006-0027-x.

Sofiev, M., Bergmann, K.-C. (Eds.), 2013. Allergenic Pollen. A Review of the Production, Release, Distribution and Health Impacts. Springer Science+ Buisness Media Dordrecht, http://dx.doi.org/10.1007/978-94-007-4881-1.

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., http://dx.doi.org/10.1007/s00484-012-0532-z.

Sofiev, M., Berger, U., Prank, M., 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., 2015. MACC regional multi-model ensemble simulations of birch pollen dispersion in Europe. Atmos. Chem. Phys. 15, 8115–8130, http://dx.doi.org/10.5194/acp-15-8115-2015.

Stach, A., Emberlin, J., Smith, M., Adams-Groom, B., Myszkowska, D., 2008a. Factors that determine the severity of Betula spp. pollen seasons in Poland (Poznań and Krakow) and the United Kingdom (Worcester and London). Int. J. Biometeorol. 52, 311–321, http://dx.doi.org/10.1007/s00484-007-0127-2.

Stach, A., Smith, M., Baena, J.C.P., Emberlin, J., 2008b. Long-term and short-term forecast models for Poaceae (grass) pollen in Poznan Poland, constructed using regression analysis. Environ. Exp. Bot. 62, 323–332, http://dx.doi.org/10.1016/j.envexpbot.2007.10.005.

Taylor, P., Andersen, T.B., 2009. Grana A model to predict the beginning of the pollen season, http://dx.doi.org/10.1080/00173139109427810.

Toro, F., Recio, M., Trigo, M.D.M., Cabezudo, B., 1998. Predictive models in aerobiology: data transformation. Aerobiologia (Bologna) 14, 179–184.

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, http://dx.doi.org/10.5194/gmd-6-1961-2013.

Zink, K., Vogel, H., Vogel, B., Magyar, D., Kottmeier, C., 2012. Modeling the dispersion of Ambrosia artemisiifolia L. pollen with the model system COSMO-ART. Int. J. Biometeorol. 56, 669–680, http://dx.doi.org/10.1007/s00484-011-0468-8.

II

Ritenberga, O., Sofiev, M., Siljamo, P., Saarto, A., Dahl, A., Ekebom, A., Šauliene, I., Shalaboda, V., Severova, E., Hoebeke, L., Ramfjord, H.

A statistical model for predicting the inter-annual variability of birch pollen abundance in Northern and North-Eastern Europe

Science of the Total Environment 615 (2018) 228–239

Science of the Total Environment 615 (2018) 228–239

Contents lists available at ScienceDirect

Science of the Total Environment

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

A statistical model for predicting the inter-annual variability of birch pollen abundance in Northern and North-Eastern Europe

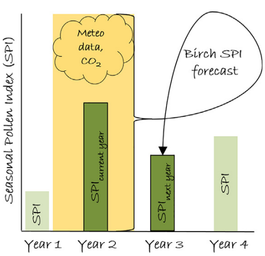
Olga Ritenberga [a,*], Mikhail Sofiev [b], Pilvi Siljamo [b], Annika Saarto [c], Aslog Dahl [d], Agneta Ekebom [e], Ingrida Sauliene [f], Valentina Shalaboda [g], Elena Severova [h], Lucie Hoebeke [i], Hallvard Ramfjord [j]

[a] *University of Latvia Faculty of Geography and Earth Sciences, Rainis bvld 19, Riga, LV -1586, Latvia*
[b] *Finnish Meteorological Institute, Erik Palmenin aukio 1, 00560 Helsinki, Finland*
[c] *Unit of Aerobiology, University of Turku, Finland*
[d] *Department of Biological and Environmental Sciences, University of Gothenburg, Sweden*
[e] *Palynological Laboratory, Swedish Museum of Natural History, Stockholm, Sweden*
[f] *Research Institute, Siauliai University, Siauliai, Lithuania*
[g] *Institute for Experimental Botany of the NAS of Belarus, Minsk, Belarus*
[h] *Moscow State University, Moscow, Russian*
[i] *Belgian Aerobiological Network, Mycology and Aerobiology service, Scientific Institute of Public Health, Brussels, Belgium*
[j] *Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway*

HIGHLIGHTS

- New model for predicting seasonal pollen index for large regions is developed.
- Procedure of cluster analysis-based region selection is proposed.
- A single universal equation describes the next year seasonal pollen index.
- Combination biological and meteorological factors shows the best predicting capacity.
- The model was tested for Russia and Belgium to identify the limits of the method.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 22 June 2017
Received in revised form 7 September 2017
Accepted 7 September 2017
Available online xxxx

Editor: P Elena Paoletti

Keywords:
Seasonal pollen index
Birch pollen
Inter-annual variability
Pollen forecasting

ABSTRACT

The paper suggests a methodology for predicting next-year seasonal pollen index (SPI, a sum of daily-mean pollen concentrations) over large regions and demonstrates its performance for birch in Northern and North-Eastern Europe. A statistical model is constructed using meteorological, geophysical and biological characteristics of the previous year). A cluster analysis of multi-annual data of European Aeroallergen Network (EAN) revealed several large regions in Europe, where the observed SPI exhibits similar patterns of the multi-annual variability. We built the model for the northern cluster of stations, which covers Finland, Sweden, Baltic States, part of Belarus, and, probably, Russia and Norway, where the lack of data did not allow for conclusive analysis. The constructed model was capable of predicting the SPI with correlation coefficient reaching up to 0.9 for some stations, odds ratio is infinitely high for 50% of sites inside the region and the fraction of prediction falling within factor of 2 from observations, stays within 40–70%. In particular, model successfully reproduced both the bi-annual cycle of the SPI and years when this cycle breaks down.

* Corresponding author.
E-mail addresses: olga.ritenberga@lu.lv (O. Ritenberga), mikhail.sofiev@fmi.fi (M. Sofiev), pilvi.siljamo@fmi.fi (P. Siljamo), annika.saarto@utu.fi (A. Saarto), aslog.dahl@bioenv.gu.se (A. Dahl), agneta.ekebom@nrm.se (A. Ekebom), ingrida.sauliene@su.lt (I. Sauliene), shalaboda_valya@mail.ru (V. Shalaboda), elena.severova@mail.ru (E. Severova), lucie.hoebeke@wiv-isp.be (L. Hoebeke), hallvard.ramfjord@ntnu.no (H. Ramfjord).

https://doi.org/10.1016/j.scitotenv.2017.09.061
0048-9697/

1. Introduction

One of the most-important parameters quantifying the strength of an allergenic pollen season is a Seasonal Pollen Index, SPI, which is defined as a sum of all daily-mean pollen concentrations, i.e. a season-long integral of pollen concentrations. It was related to severity of human allergy (Bastl et al., 2016; D'Amato et al., 2007; Huynen et al., 2003), used as an indicator of the productivity of trees, such as olives (Galán et al., 2014; Myszkowska, 2013; Orlandi et al., 2005; Oteros et al., 2013; Prasad and Craufurd, 1999), as well as predictive parameter for the wine (Cunha and Ribeiro, 2015) or olives (Dhiab et al., 2016) production and as a bio indicator of plant reaction to the on-going climate change (Hatfield and Prueger, 2015; Hedhly et al., 2009; Storkey et al., 2014; Zhang et al., 2014). Apart from that, the SPI is used in numerous pollen forecasting models as a scaling factor determining the predicted pollen concentrations (Helbig et al., 2004; Prank et al., 2013; Puc, 2012; Ranta et al., 2008; Ritenberga et al., 2016; Siljamo et al., 2012; Sofiev et al., 2012; Stach et al., 2008; Toro et al., 1998; Veriankaitė et al., 2009; Zhang et al., 2013; Ziello et al., 2012).

The SPI is known to change substantially from year to year depending on combination of meteorological factors and physiology of the plant (Masaka, 2001; Ranta and Satri, 2007), see also a review of Dahl et al. (2013). Such variability, for some trees (e.g., birch), exhibits a quasi-bi-annual behaviour (a strong year is followed by a weak one and vice versa), which however, is broken in some years (Dahl and Strandhede, 1996; Detandt and Nolard, 2000; Grewling et al., 2012; Hättestrand et al., 2008; Jato et al., 2007; Latałowa et al., 2002). This behaviour was attributed by Dahl and Strandhede (1996) to a combination of meteorological and physiological factors, who suggested that neither meteorology nor the innate biannual behaviour is decisive, but rather a combination of both. Indeed, since catkin development is expensive, their abundance takes a large toll of carbohydrates from the annual shoot and impedes the expansion of leaves, thus limiting the amount of photosynthesis products available for development of the next year catkins. This cycle can be interrupted if weather is, for instance, strongly favourable allowing the smaller leaves to assimilate efficiently. Then flowering can be strong for two years in a row.

The same variability tends to occur synchronously in several plants, such as birch, alder and hazel, and over large regions (Ranta and Satri, 2007; Šaulienė et al., 2014). Such synchronization also suggests strong influence of meteorology as the only common factor for different plants distributed over large areas.

Apart from the strong inter-annual variability, several wind pollinated trees, such as birch, olives, oak, etc. have positive long-term trend of the SPI (Garcia-Mozo et al., 2014; Yli-Panula et al., 2009; Prank et al., 2013; Severova and Volkova, 2016; Spieksma et al., 2003, 1995). These trends were attributed to changing climatic conditions and/or to increasing abundance of the plants. Among measurable indicators of these factors, one can consider the growing level of CO_2 in the atmosphere and trends in the regional leaf area index. Thus, several studies showed that with higher level of CO_2, plants tend to produce more pollen (Ladeau and Clark, 2006; Zhang et al., 2015; Ziska et al., 2001; Ziska and Beggs, 2012).

Among the meteorological factors affecting the intensity of the trees flowering, temperature and precipitation amount of the preceding year are the most-commonly mentioned (Yli-Panula et al., 2009). One can therefore expect that the regional synchronization of the SPI behaviour takes place at least at synoptic scale, i.e. 10^2–10^3 km and this very scale should be considered when developing models for the SPI. However, most of studies consider it at shorter scales using one or few closely-located sites (Corden et al., 2002; Dahl and Strandhede, 1996; Grewling et al., 2012; Severova and Volkova, 2016).

Many of the above studies completed the analysis of the SPI inter-annual behaviour with statistical models aiming at predicting the next-year SPI using the previous-year SPI and meteorology. The procedure of constructing such models usually started from log- or square-root (sqrt-) transform of the SPI followed by multi-linear regression fitting. The input parameters usually included in the analysis are mean temperature of various time intervals (one study took heat sum over certain periods) and the previous-year SPI. However, no systematic pre-processing of the input parameters was performed, and all models were built for individual locations, even if the study was considering several sites.

The current study addresses the above-outlined omissions and aims at construction of a predictive model for birch SPI over large regions in Europe. We will propose a simple procedure for delineating such areas and build the model for the northern region.

2. The working hypotheses

We assume that the SPI is a regional parameter determined by the synoptic-scale meteorological processes, i.e. a few hundreds of kilometres. It should therefore be possible to identify the regions that react synchronously and demonstrate similar patterns of the SPI year-to-year variations. The corresponding temporal scale is from several days up to 1–2 weeks – this will be the maximum temporal resolution of the input data, i.e. we shall not be interested in individual meteorological events.

Secondly, absolute values of the SPI are of essentially no importance: they are decided by vegetation density in proximity to the station, which is a static parameter. Therefore, spatial and temporal variations inside these regions are separable.

With the above assumptions, it should be possible to construct a statistical model for the SPI variation over these regions taken as "boxes", i.e. not resolving individual stations but taking each region as a single entity with the normalised SPI averaged over the region.

3. Materials and methods

3.1. Study area

The study was focused over the region between the latitudes 50°N and 70°N and longitudes from 5°E to 40°E, Birch pollen data from 15 aerobiological sites were included in the analysis.

The analysed region has moderate maritime climate with cold winters and moderate summers. Birch fraction reaches up to 30% of the total forest coverage (see maps in (Ritenberga et al., 2016; Sofiev et al., 2006)).

3.2. Airborne pollen concentrations

Birch pollen data were extracted from the database of European Aeroallergen Network (EAN). Pollen sampling was performed using Burkard or Lanzoni 7-days volumetric trap (Hirst, 1954). Pollen was identified with optical microscopy, using country-specific (random, vertical, horizontal traverses) counting technique but generally following the aerobiological standards (Galán et al., 2014). For each year, the SPI was computed as an integral of concentrations over the whole observed period. The length of time series varied from station to station, reaching 40 years (1974–2015) for a few sites. The varying length of observations created evident challenges for processing and interpretation of the obtained results. However, under the assumption of internal homogeneity of the SPI variation within the region, verified in the next section, all stations are normalised and averaged into the regional mean, thus reducing the problem to a different sampling volume in different years. Additionally, all time series shorter than 11 years were filtered out (Table 1). No other thinning of the dataset was possible due to the limited number of sites (15, as shown in Table 1 and Fig. 1).

3.3. Meteorological data, CO_2 and LAI

A wide list of parameters potentially governing the regional birch SPI was created based on literature analysis and general considerations.

3.3.1. Meteorological information

The most-frequently mentioned meteorological factors influencing the seasonal pollen index are temperature and precipitation rates (Dahl and Strandhede, 1996; Yli-Panula et al., 2009; Latałowa et al., 2002). It was decided to use also related parameters, such as solar short wave radiation and accumulated heat amount (heat sum), with care taken against mutual correlation of these parameters as descried in the next section. The meteorological data were taken from the European Re-analysis ERA-Interim dataset (Dee et al., 2011; Simmons et al., 2010). Temporal resolution of the data was 3 h with spatial grid of 0.72° × 0.72°. The meteorological data covered the period from 1980 to 2015, which thus limited the analysed period to these 36 years. As seen from Table 1, the sampling network of the region was limited during 1980s to 4 locations, expanding over 10 sites only after 2000.

Heat sum (1) is calculated from daily mean temperature following the approach of the SILAM model (Sofiev et al., 2012), which suitability for local-scale computations was confirmed by modelling in Riga (Ritenberga et al., 2016).

$$H(d) = \sum_{d=d_s}^{D} \left[\overline{T(d)} - T_{c-o}\right]_+ \tag{1}$$

Here, H is temperature sum (heat sum), with units of degree day (C°/day); d is the day of H calculation; d_s – starting day of the heat accumulation – 1st of March in this case; $T(d)$ – daily temperature, °C; T_{c-o} – cut-off temperature (temperatures below this threshold are not summed up), °C, $[x]_+$ is the function that equals 0 for $x < 0$ and 1 for $x \geq 0$ (it excludes the temperatures below the cut-off level).

The same cut-off temperature of 3.5 °C was used for the heat sum calculations for all monitoring sites in the region (Sofiev et al., 2012), justification for accepting a value of 3.5 °C is described by Ritenberga et al. (2016) and Siljamo et al. (2012).

3.3.2. Leaf area index

Leaf area index (LAI) was included in the data analysis as one of possible predictors for the SPI. It was hypothesised that there could be a relationship (straight or inverse) between the annual production of pollen and the LAI. An indirect support of this consideration is that the pollen data are used for the LAI reconstruction in palynological studies (Williams et al., 2008). Data of the LAI were taken from a global reanalysis of the vegetation phenology (Stckli et al., 2011), which contains daily 1°×1° LAI maps for the whole study period. For our purposes, they were aggregated to monthly level.

3.3.3. Carbon dioxide

Carbon dioxide (CO_2) is one of the factors, possibly influencing the annual amount of pollen from various taxa (Albertine et al., 2014). Carbon dioxide has a well-expressed positive trend during at least last 40 years, somewhat resembling the trends in the birch pollen abundance. The CO_2 data used for the analysis were downloaded from the NOAA Earth System Research Laboratory (ESRL) public archive (*Ed Dlugokencky and Pieter Tans, NOAA/ESRL* (www.esrl.noaa.gov/gmd/ccgg/trends/) for the years from 1980 to 2015.

3.3.4. Seasonal pollen index

The *SPI (Y-1)* of the preceding year was included to account for the plant state and behaviour in the past.

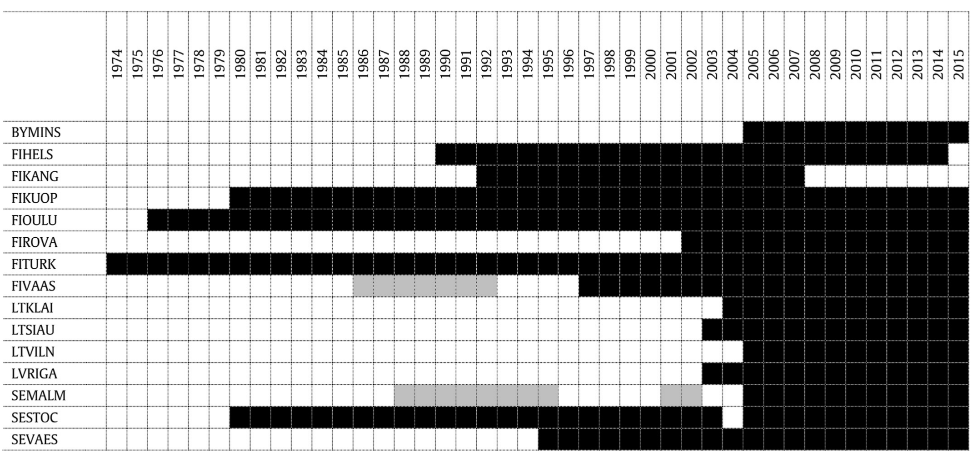
4. Model construction

4.1. Region selection

The SPI time series observed at EAN stations with >11 years of data (taken both from within the target region and from the stations outside) were correlated with each other and clustered with regard to the pairwise correlation coefficient r. Despite not satisfying the rigorous definition of distance, this parameter has two vital features: (i) it does not depend on absolute SPI values at the stations, and (ii) it is of correct monotonicity, i.e. shows the stations with similar inter-annual variations as close ones.

The clustering procedure also requires definition of the reference point, against which all other stations are positioned. We used the Turku station in Finland, which: (i) possesses one of the longest time

Table 1
Availability of birch pollen data for analysis of Northern, North-Eastern Europe regional data set sites (grey - data available; black - data suitable for analysis).



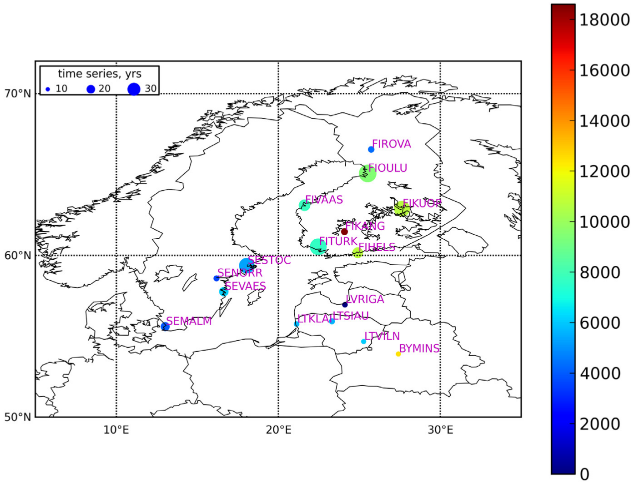


Fig. 1. A map of station characteristics for the Northern, North-Eastern Europe regional cluster. Color represents the geometrical mean of the SPI, the dot size is proportional to the length of the time series, provided in Table 1.

series in EAN fully covering the study period, (ii) is conveniently located in the middle of the region with comparatively dense pollen network, (iii) shows strong SPI variability, (iv) shows high absolute birch pollen concentrations, thus making the problem important from allergological point of view, (v) is located in the climate-sensitive area: the temperature rise in Finland due to climate change is much higher than the global average (Hartmann et al., 2013), which can have its fingerprints in the SPI trends.

The correlation matrix was analysed and only site pairs with $r > 0.85$ and $p < 0.01$ were retained in the cluster (Table 2). The resulting cluster of stations, hereinafter referred as the Northern, North-Eastern Europe regional data set, included aerobiological stations of Finland (Helsinki (FIHELS), Kangasala (FIKANG), Kuopio (FIKUOP), Oulu (FIOULU), Rovaniemi (FIROVA), Turku (FITURKU), Vaasa (FIVAAS)), Lithuania (Klaipeda (LTKALI), Siauliai (LTSIAU), Vilnius (LTVILN)), Latvia (Riga (LVRIGA)), Belarus (Minsk (BYMINS)), Sweden (Stockholm (SESTOC), Malmö (SEMALM), Vaestervik (SEVAES)), totally 15 stations with >11 years of data (Fig. 1). Long- term observations are available from Finland (data from 7 stations) and Sweden (3 stations), whereas the stations from Belarus, Lithuania, and Latvia have from 11 up to 20 years of data.

As a result of this procedure, the area of the region for analysis gets clear: it covers the above countries and possibly extends further to the east, where no stations are available all the way down to Moscow, which is outside the cluster. Western border is also somewhat unclear: in Norway, only NOTRON in central part of the country satisfies the requirement on the time series length – and is outside the cluster, probably because of strong dominance of Atlantic Ocean in the weather pattern.

As shown in Table 1, sampling of the region evolves with time, with its southern part practically not represented in early years. However, the existing time series are sufficient for establishing the statistically

Table 2
Correlation matrix of SPI for the **Northern, North-Eastern Europe** regional data set.

	BYMINS	FIHELS	FIKANG	FIKUOP	FIOULU	FIROVA	FITURK	FIVAAS	LTKLAI	LTSIAU	LTVILN	LVRIGA	SEMALM	SESTOC	SEVAES
BYMINS	1	0.66	0.95	0.61	0.74	0.84	0.8	0.69	0.78	0.70	0.69	0.49	0.79	0.87	0.95
FIHELS		1	0.91	0.87	0.90	0.83	0.93	0.89	0.91	0.97	0.89	0.83	0.87	0.85	0.69
FIKANG			1	0.92	0.91	0.84	0.96	0.79	0.99	0.78	0.97	0.67	0.80	0.92	0.84
FIKUOP				1	0.85	0.85	0.76	0.86	0.87	0.91	0.88	0.94	0.75	0.77	0.66
FIOULU					1	0.89	0.8	0.79	0.94	0.86	0.95	0.83	0.78	0.79	0.71
FIROVA						1	0.90	0.81	0.82	0.79	0.80	0.80	0.73	0.77	0.75
FITURK							1	0.92	0.86	0.89	0.86	0.79	0.84	0.86	0.85
FIVAAS								1	0.86	0.94	0.85	0.87	0.67	0.83	0.74
LTKLAI									1	0.96	0.96	0.82	0.92	0.94	0.79
LTSIAU										1	0.94	0.85	0.88	0.89	0.72
LTVILN											1	0.83	0.91	0.93	0.78
LVRIGA												1	0.60	0.70	0.57
SEMALM													1	0.82	0.72
SENORR														0.91	0.78
SESTOC														1	0.93
SEVAES															1

significant high correlation between all sites in the region, which justifies the consideration of the "regional mean SPI variability", unified over the whole period. The next section describes the transformations used to obtain the single 36-years-long regional time series set for further analysis.

4.2. Target variable transformation: Normalization of the SPI distribution, regionalization

The raw pollen data have several features, which make construction of an efficient statistical model problematic – see discussion in Ritenberga et al. (2016). Integration over the season towards SPI reduces some of the problems, e.g., the non-stationarity of the time series. However, several data pre-processing steps are still required prior to the model construction.

The SPI_i of each station i for the particular year Y should be normalised with its multi-annual geometric mean $SPI_i^{geomean}$ taken over the whole observed period of the station (represented by color in the Fig. 1), thus eliminating the dependence on local birch abundance and making all sites within the region comparable (Fig. 2, panels a and b):

$$SPI_i^{norm}(Y) = \frac{SPI_i(Y)}{SPI_i^{geomean}} \tag{2}$$

Since the distribution density of the SPI_i^{norm} is closer to log-normal than to normal, we applied logarithmic transformation, finally arriving at the deviation $\Delta SPI_i^N(Y)$ of the normalised annual SPI_i for each year Y from its long-term geometric mean value (Fig. 2, panel c).

According to the working hypothesis supported by the correlation analysis above, the same processes control the SPI behaviour at all stations within the selected region. Therefore, one should consider their SPI series $\Delta SPI_i^N(Y)$ as realizations of the single stochastic process – the regional time series $\Delta SPI^{reg}(Y)$, — which effective and unbiased estimation is arithmetic mean over the cluster.

4.3. Input data transformation

Upon turning the SPI time series into normally-distributed deviations from the long-term mean levels, corresponding transformations should be applied to the input data as well: they should be converted into deviations with near-normal distribution type, the approach analogous to (Ritenberga et al., 2016) but with the SPI as the target variable.

4.3.1. Selection of the time axis

According to the above-mentioned studies, only meteorological conditions during specific phenological phases affect the next-year SPI. One needs therefore to delineate the corresponding periods. Astronomical time used in those studies is an inconvenient variable since it ignores difference between the years and thus tends to mix-up the phenological phases. It also ignores the difference in the flowering time across the region. Therefore, again following (Ritenberga et al., 2016) approach, we selected the heat sum as the phenologically-relevant variable for the time axis.

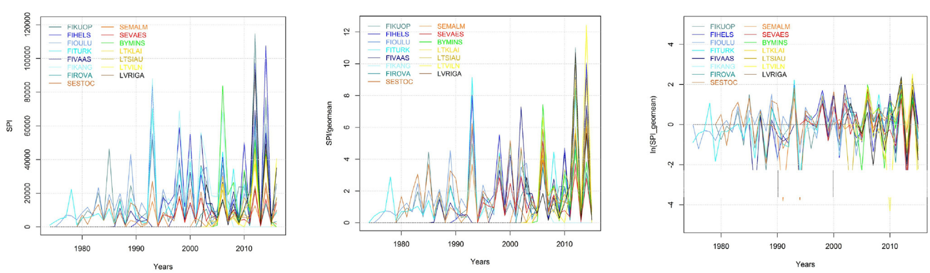
The challenge of the regional consideration is that even expressed via heat sum, the phenological phases still occur at different moments: the further to the north the less heat is needed for the season progression (Sofiev et al., 2012). To overcome this difficulty, we express the time scale in % of the total heat accumulated during the whole year, normalised to its long-term mean value, at each station. Annual heat sum differs by up to a factor of two between the aerobiological stations within the region, whereas, e.g., the heat sum threshold for flowering expressed in relative terms is nearly constant - see (Supplementary material 1 - "Annual_cumulative_heat_sum_in_2010") of (Sofiev et al., 2012) with the data source of ERA-Interim and degree day as unit. The idea of this transformation goes back to (Linsser, 1867), who demonstrated it for pine.

For the normalised heat sum, the year was split to 6 periods representing different fractions of the annual heat sum: 0–0.05 (pre-season before the flowering starts); 0.05–0.13 (flowering season); the three post-season intervals: 0.13–0.25; 0.25–0.45; the whole post-flowering period with shorter end 0.13–0.4; and the whole period from start of heat sum accumulation until the end of active vegetation was taken as a single time slot 0.0–0.45. The selection of the periods was based on preliminary calculations of the heat scale for Turku and Riga.

4.3.2. Conversion of the input meteorological data into deviations

The deviation of the regional $\Delta SPI^{reg}(Y)$ from its long-term mean should be related to the normalised meteorological variables, also taken as deviations. The transformation was made station- and variable- wise and included:

(i) calculation of variable deviation from the long-term mean value, for each year, each site, during each of the above heat sum intervals; Since this step does not involve pollen data there is no limitations due to the data availability. The meteorological data are taken from ERA-Interim and the whole period 1980–2015 is used without omissions;



(a) initial time series, ∑ pollen /yearm-3]

(b) normalized station-wise with multi-annual geometric mean, [relative unit]

(c) SPI normalized deviation: logarithm of the time series panel (b), [relative unit]

Fig. 2. Transformation of the SPI observations for the stations in the Northern, North-Eastern Europe regional data set.

(ii) spatial averaging of deviations across the region, for each year. As in the step (i), all years 1980–2015 were taken into consideration. As a result, we obtain the regional set of predictors: regional deviations of the environmental parameters from their long-term average, for each of the six heat sum intervals during each year.

4.3.3. Physical parameters describing the long-term trends

No transformations were applied to of CO_2 and LAI in order to preserve their properties: CO_2 has a well-expressed trend during the last 40 years and LAI reflects the annual changes in the vegetation of the region (Supplementary material 2, Supplementary material 3). Both parameters monthly data were simply averaged for each year.

4.4. Construction of predictive SPI model

Very limited size of the dataset (only one regional time series comprised of 15 individual time series of varying length) did not allow splitting the stations to learning and control datasets. Therefore, the regional-model coefficients were obtained using *bootstrap* method and multi linear regression (MLR).

At the first step, MLR was made against a complete set of meteorological parameters listed above, taken for all 6 time intervals during the year *Y* (the year to predict) and the year *Y-1* (previous year), in order to identify statistically significant predictors. Expectedly, predictors from the previous year *Y-1* showed larger influence on the SPI of the year *Y* than those from the year *Y*. From 21 predictors (6 different periods for temperature, solar radiation, precipitation; CO_2; LAI; SPI – all for the year Y-1), the following parameters from *Y-1* were identified as potentially useful:

- precipitation $Pr_{0.13}$ for the period 0.13–0.25 of the annual heat sum accumulation
- temperature $T_{0.13}$ for the period 0.13–0.25, of the annual heat sum accumulation
- temperature $T_{0.25}$ for the period 0.25 to 0.45 of the annual heat sum accumulation
- short wave solar radiation SW_0 for the period from 0 to 0.05 of the annual heat sum accumulation
- $SPI_{Y\text{-}1}$
- $CO_{2\ Y\text{-}1}$

All further analysis uses only these 6 parameters.

The bootstrap included 15 iterations, one for each of 15 station datasets withheld from both regional SPI and regional meteorological averaging. The SPI data and meteorological parameters for the remaining stations were averaged over the region. After that, the multilinear regression was calculated obtaining the model coefficients for the regional time series with one site excluded.

The fitting coefficients were averaged across the iterations, finally obtaining the mean regression coefficients and their standard deviation.

The procedure was repeated with and without previous-year SPI_{Y-1} in the list of possible predictors.

5. Results: models for SPI in Northern, North-Eastern Europe

5.1. Model with meteo-only input

During the data analysis, we realized that it is possible to build the model using *only available meteorological data* and the trend-describing variable $CO_{2,Y\text{-}1}$. This opportunity allows predicting the SPI relative deviation from the long-term mean knowing neither this mean level nor the previous-year SPI. The model becomes detached from the SPI observations and applicable also in places with no aerobiological observations whatsoever.

The regional formula for $\Delta SPI^{reg}(Y)$ of the year *Y* based on meteorological data and data of CO_2 of the previous year *Y-1* is:

$$\Delta SPI^{reg}(Y) = a_0 + a_{CO_2} CO_{2,Y-1} + a_{pr} Pr_{0.13} + a_{sw} sw_0 + a_{T_{0.13}} T_{0.13} + a_{T_{0.25}} T_{0.25} \tag{3}$$

The mean values of the intercept a_0 and other coefficients are shown in Table 3, first row.

The obtained model is not perfect (see evaluation section) but its independence from pollen observations makes it a very important instrument for assessment of the SPI variability.

5.2. Model with full set of variables; accounting for the different statistical significance

Adding $SPI_{Y\text{-}1}$ as a predictor significantly increases the quality of the model. The final formula includes the same meteorological parameters as the meteo-only model, CO_2 and SPI from the previous year (Table 3, second row).

There are two parameters of Eq. (3) with noticeably lower statistical significance than the others (*p-value* 0.05–0.1): precipitation *Pr* and $T_{0.25}$. Their importance was evaluated by removing them from the data set and constructing the model based only on predictors of high statistical significance (Table 3, last row).

5.3. Evaluation of the constructed models

The evaluation of the three models of the Table 3 followed several pathways and was performed for both entire region (i.e. the learning dataset) and for each station in the cluster. There are several parameters used for the evaluation, including both "standard" statistics and threshold-based ones. Efficiency of the standard statistics was however hampered by the limited size of the dataset and strongly non-normal distribution of the values. The evaluation approach also took into account the way these models will be applied: for forecasting the next-year SPI, in particular determining whether it will be high or low year and for forecasting the absolute SPI needed for the modelling.

5.3.1. Statistics over the whole observed periods

We consider: (i) Odds Ratio (OR, Fig. 3a) as a measure of differentiating between the high- and low- seasons, (ii) Model Accuracy (MA, Fig. 3b) and (iii) a fraction of the SPI predictions that fall within the factor 2 from the observations (F2, Fig. 3c).

Behaviour of individual stations inside the region differs somewhat but the regional formula tends to work better for the stations with longer time series, i.e. those, which contributed mostly to the model identification (Turku, Kuopio, etc.). The limited size of the dataset and high scores of the models resulted in just few cases of the predictions being wrong. For several sites, the low-high differentiation appeared always correct during the observed years. The sample OR for such sites is infinitely high, shown as OR = 30 in Fig. 3a.

The fraction of the predictions falling within the factor of 2 from absolute observations (F2, Fig. 3c) allowed formulating the static persistence-based estimate: the mean SPI at each station (yellow bars in Fig. 3c). The F2 generally stays within 40%–70% being lower only close to the southern edge of the domain. Similarly to other quality measures, the full-parameters and significant-only models showed the highest F2 levels, with meteo-only model staying somewhat behind. The suggested models demonstrated similar or better skills than the persistence approach. One can also see that the northern-most sites FIKEVO and FIROVA tend to be exceptional: at FIROVA, the meteorological information alone was sufficient for good predictions, while FIKEVO was the only site with the highest skills showed by the persistence approach. The reasons for the unusual behaviour of the models are discussed in the next section.

Table 3
Coefficients (with standard deviation) for regional formula of SPI forecast based on SPI_{-1}, CO_2 and meteorological set.

	a_0	a_{CO2}	a_{Pr}	a_{sw}	$a_{T0.13}$	$a_{T0.25}$	a_{SPI}
Only meteo and CO_2	-2 $(\pm 8\times10^{-1})$	5×10^{-3} $(\pm 2\times10^{-3})$	6×10^{-2} $(\pm 7\times10^{-3})$	2×10^{-5} $(\pm 6\times10^{-7})$	3×10^{-1} $(\pm 3\times10^{-2})$	3×10^{-1} $(\pm 3\times10^{-2})$	–
All predictors	-4 $(\pm 7\times10^{-1})$	9×10^{-3} $(\pm 1\times10^{-3})$	-5×10^{-2} $(\pm 5\times10^{-3})$	2×10^{-5} $(\pm 7\times10^{-6})$	2×10^{-1} $(\pm 1\times10^{-2})$	1×10^{-1} $(\pm 1\times10^{-2}$	-5×10^{-1} $(\pm 3\times10^{-2})$
Only highly significant predictors	-4 $(\pm 7\times10^{-1})$	9×10^{-3} $(\pm 9\times10^{-3})$	–	2×10^{-5} $(\pm 7\times10^{-6})$	2×10^{-1} $(\pm 1\times10^{-2})$	–	-5×10^{-1} $(\pm 3\times10^{-2})$

– was not included in model.

5.3.2. Prediction for 2016

The year 2016 was not included into the model fitting, thus serving for additional, albeit limited, evaluation mimicking the way the model will actually be used: to forecast the next-season severity. The predictions of the models for this year are shown in Fig. 4 (blue, red and yellow columns) together with the long-term mean for each station (shaded area, geometric mean of observed SPI at the site over al years available) and the observed SPI in 2016 (dark green columns). A few stations were excluded from the analysis: FIHELS was moved a few km inland from its historical location, thus making absolute-SPI analysis impossible, whereas the data of FIKANG were not available.

As one can see, 2016 was mostly high-or-normal year over the region: low SPI was recorded only at SEMALM (south-western edge of the region) and LTKLAI (southern edge, near the coast). The meteo-only model was predominantly suggesting moderately-low year, whereas the models including the previous-year SPI suggested high year, in agreement with the observations. The long-term average was a worse predictor than the full-parameters models but evidently was a better than the meteo-only model at the sites where that one predicted low season.

6. Discussion

The primary goal of the study was to build the unified model suitable for applications over large regions, thus demonstrating the feasibility of spatial generalization of the SPI predictions using large-scale meteorological features as the controlling parameters. As pointed out in the introduction, the bulk of previous studies concentrated on a single or a few closely-located stations.

The second principal difference of the suggested approach is that we applied a series of non-linear transformations of the input data and changed the governing variable from time to normalised heat sum. These transformations aimed at two targets: (i) sharply reduce the year-to-year variability of temperature accumulation, thus making the seasonal propagation and flower/leaves/seed maturation curves similar for all years; (ii) eliminate also the spatial variability of these curves, thus making them similar for all locations over the considered domain. After that, construction of the predictive models followed standard procedures for multi-component optimal model fitting.

6.1. Comparison with other approaches

Since the suggested models are first in their class, direct comparison with published studies is quite difficult. However, certain conclusions can be derived. The most-direct comparison is possible with the study of (Ranta and Satri, 2007), who built three local models for Turku, Kuopio and Oulu stations. Temporal correlation of these models however trails significantly behind the current unified model: it was 0.59 vs 0.82 for Turku and 0.61 vs 0.77 in Oulu. Only in Kuopio, both models scored to 0.65. The most-probable reason for the higher scores of the current approach is the more comprehensive set of transformations of the data before applying regression.

Similar predicting capacity to that of (Ranta and Satri, 2007) was reported by (Dahl and Strandhede, 1996) but direct comparison is not possible due to sqrt-transformed values reported in that paper. An important difference of that work was that the predictors were taken as heat sums rather than mean temperature over some period, which potentially can further improve the model accuracy.

Quite high correlations of the local SPI and several meteorological drivers were reported for Poznan by (Grewling et al., 2012) but no predictive model were built. The authors also pointed out that they found no parameters capable of explaining the break points in the bi-annual cycle. This study has succeeded but Poznan is outside the current region and the developed models are not directly applicable there.

6.2. Selection of input parameters

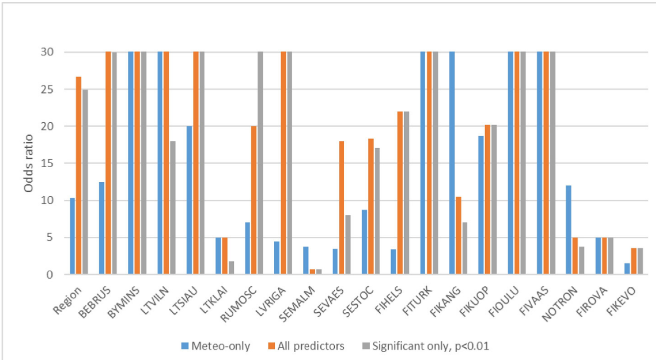
Among the parameters influencing the SPI, the northern part of Europe is mostly limited by the available heat and solar radiation, so it is not surprising that heat is the most important predictor for our region (Table 3). This is supported by very high scores of the model in the north of Finland (up to 90%, Fig. 5) and also pointed out by (Ranta et al., 2005). In the south of Europe the stress will probably shift towards water and nutrient availability (Dahl et al., 2013), with Central Europe being arguably the most-difficult to model.

A noticeable degradation of the model skills at the northern-most site FIKEVO is probably caused by the ignored pollen atmospheric transport. Indeed, the characteristic level of the SPI in that region is much lower than in Central Finland and Russia, which makes the local pollen production less important than the impact of the long-distance transport. The transport contribution was estimated by (Sofiev, 2016), who showed that the transport-induced SPI variability grows sharply towards remote regions. This points out at a certain shortcut of the current approach: the developed models should predict the total amount of pollen released from catkins rather than the observed concentrations. The proportionality of these two quantities is not exact and is disturbed by both local pollen dispersion conditions and by the long-range transport (Sofiev, 2016). Averaged over the whole season, the disturbance is comparatively small in the self-polluting regions but can be large if strong remote sources affect the area. The other possible explanation is the genetic differentiation between the birches in far north (mainly *Betula nana*) and other parts of the domain. The significance of such differentiation was shown for olives by (Chuine and Belmonte, 2004).

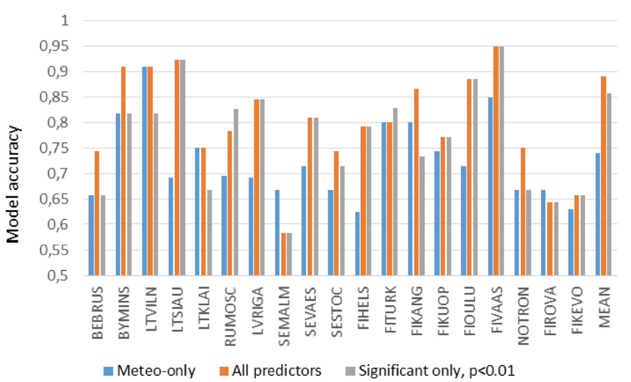
6.2.1. Meteo only model

An intriguing feature of the current approach differentiating it from the previous studies is that it had produced a comparatively well-performing model using only meteorological parameters without any reference to the previous-year pollen observations. This is a confirmation of the working hypothesis: meteorology alone controls a large fraction of the SPI variability, synchronously over the whole region.

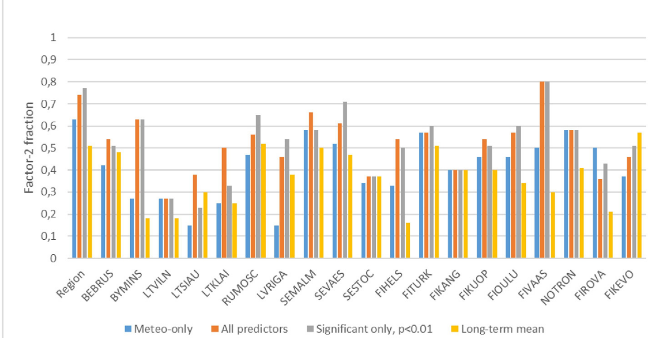
Performance of the meteo-only model still trails behind the full-input model, thus suggesting that the biological processes also significantly contribute to the SPI – in agreement with (Dahl and Strandhede, 1996). In particular, the dynamic range of the SPI variability is the largest in the observations, closely followed by the full-input model, whereas the meteo-only model is more conservative (Fig. 5). One can argue then that the plant response serves as an amplifier for the meteorological signals. This also corroborates with conclusions of



a)Odds ratio for the low-high season predictions by the three models presented in Table 3 (full-set; meteo only, full set with only significant (p<0.01) predictors). Sites are ordered along the latitude**.**



b) Model accuracy (MA) for the low-high season predictions by the three models presented in Table 3 (full- set; meteo only, full set with only significant (p<0.01) predictors). Sites are ordered along the latitude**.**



c) **Fraction of the SPI predictions within the factor of 2 (F2). Same notations as in c)3b.**

Fig. 3. Statistics for the constructed models evaluation.

the biological model (Dahl et al., 2013) that stressed the importance of the combination of meteorological and biological parameters for adequate next-season prediction.

Finally, the meteo-only model reproduces both the bi-annual cycle and the years when it breaks down (Fig. 5), also strongly suggesting that such behaviour of birch is at least inspired by the regional-scale weather phenomena.

6.2.2. SPI only model

The second "simplified" model can be built using only SPI but taking it from more than one preceding years. In-essence, it exploits the bi-annual cycle as the foundation (Ranta et al., 2005; Ranta and Satri, 2007) but can, to a certain extent, adapt to several years without such cycle in sight. In the Baltic countries and Southern Sweden (area 15°E, 25°E, 55°N, 65°N), the SPI correlation between forecasted year Y and previous year (Y-1) is −0.47 whereas the correlation between the forecasted Y and two years earlier birch SPI (Y-2) is 0.41. One of possible ways to formalise the relations is the following:

$$\Delta SPI^{reg}(Y) = \ln(b) + a\left(\Delta SPI^{reg}(Y-2) - \Delta SPI^{reg}(Y-1)\right) \quad (4)$$

Fitting the coefficients *a* and *b* for Finland (obs: not for the whole region!), one obtains a = 0.21 and b = 0.97. The optimal *a* and *b* values differ from area to area and country to country but *b* is usually around 1 and *a* varies between 0.1 and 0.3. Areas where mean SPI is low have proven to be difficult. As a result, $b = 1.0$ and $a = 0.21$ may be taken for the whole Europe, thus providing a rough screening algorithm for the SPI predictions if a few years of pollen observations are available but meteorological data are not straightforward to obtain (a rare occasion).

Performance of the SPI-only model Eq. (4) is usually worse than the scores of the meteo-only model – see examples for a few Finnish stations in (Fig. 5). An exception is the Vaasa station (Fig. 5), which is one of the northern-most sites in the region. That far in the north, the limitations of the plant productivity are the most-severe, so the biology-related amplification of the external stress become so strong that SPI itself contains sufficient amount of information to predict the next-year flowering. As a result, both models that use SPI as (one of) predictor(s) show very high correlation whereas meteo-only model stays behind.

6.3. Possibilities for further generalization of the models

The suggested models were made for a large but still limited region in the north of Europe. Since the climatic conditions and plant response to the stress change gradually along north-south and west-east directions, the models score lower in Lithuania and southern Sweden (Fig. 6), which delineate the southern border of the region.

6.3.1. Performance outside the domain

Outside the region, the model gives poor predictions if temporal correlation is considered (Fig. 7) – but the low-high- season differentiation is still very good (see OR ans MA, Fig. 3). The time series are still quite good for the years with pronounced bi-annual cycle but large errors show up when this cycle breaks down. Such "unusual" years also become more frequent, especially in Brussels, where the bi-annual cycle is practically non-existent. It indicates a presence of some unaccounted factors, which control the SPI in temperate climate.

6.3.2. Steps towards European model

Considering the expansion of the models, one needs to involve other parameters, at least the amount of precipitation and winter-time chilling. The principal difficulty will be that the interplay between the governing processes will, most probably, be strongly non-linear and non-additive. In-essence, this study covered the area where the relations between the governing parameters and the SPI are well represented as linear and additive. Noteworthy, the SPI variations correlate between different locations within the region but not with places outside of it. It again points out at non-linear dependencies of the SPI on governing parameters if a wider region is considered.

At the current stage, there are two possibilities for constructing the model for the whole Europe: (i) increase the number of parameters and allow for strong non-linearities in the dependencies (possible, as shown by Ritenberga et al., 2016), (ii) construct different model(s) for Central and Eastern Europe as, for instance, done for the olive season timing forecasting by (Aguilera et al., 2014). Each approach has its pros and contras.

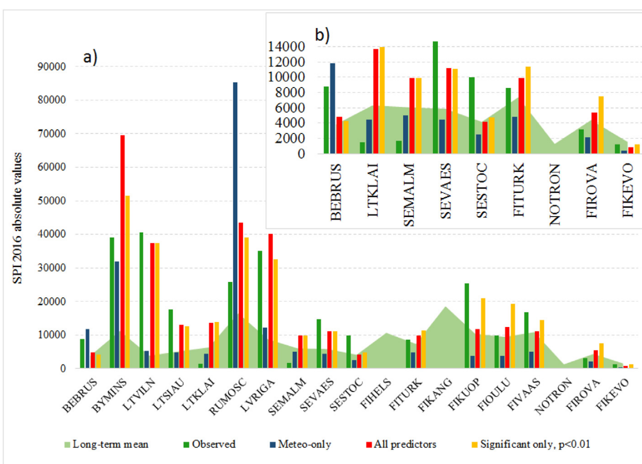


Fig. 4. Comparison of the predicted and observed SPI for the three models presented in Table 3 in 2016 ($\sum$ pollen/year m^{-3}), a) – full data set; b) low SPI sites.

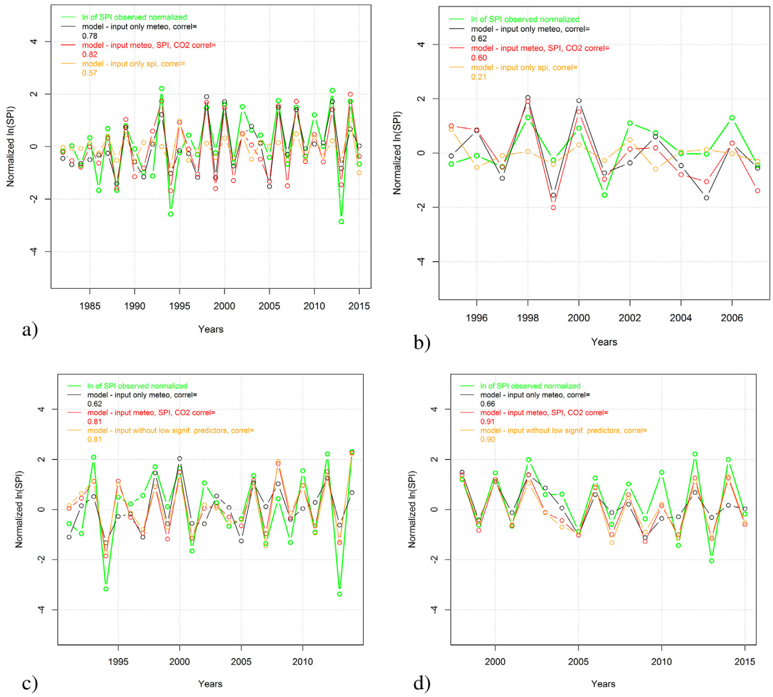


Fig. 5. Comparison of the three models for Finnish stations. Panel: a) Turku; b) Kangasala; c) Helsinki; d) Vaasa.

The first approach, being clearly preferable from the user standpoint, faces a wide variety of interplay of the governing parameters, which have to be identified and quantified. For instance, lack of chilling can both reduce the SPI and shift the season (Hänninen, 1990), whereas very cold winters may have no effect at all (saturation above certain threshold, (Hänninen, 1990)), provoke more intense flowering and seeding (reproduction push), or damage some trees and reduce SPI for several years (damage of the habitat). These and other processes can dominate or be negligible depending on climatic zones, so that their unified consideration will be cumbersome. Finally, expansion will require different treatment of meteorology since the area will be much larger than the synoptic scale, i.e. the homogeneity assumption with regard to the meteorological input will no longer be valid.

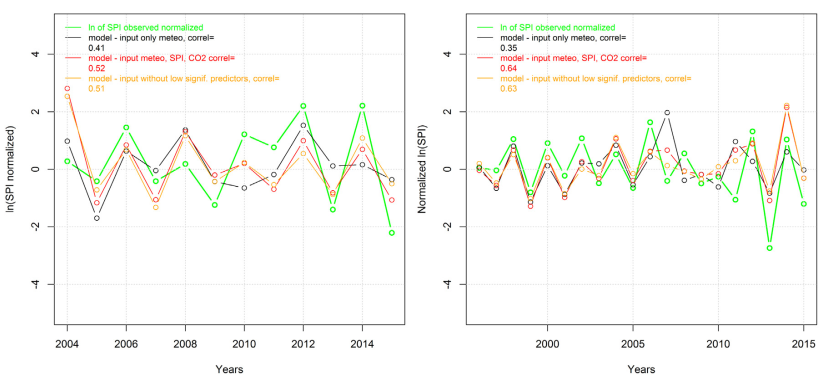


Fig. 6. Model performance in Lithuania (Siauliai) (left-hand panel) and southern Sweden (Vaestervik) (right-hand panel).

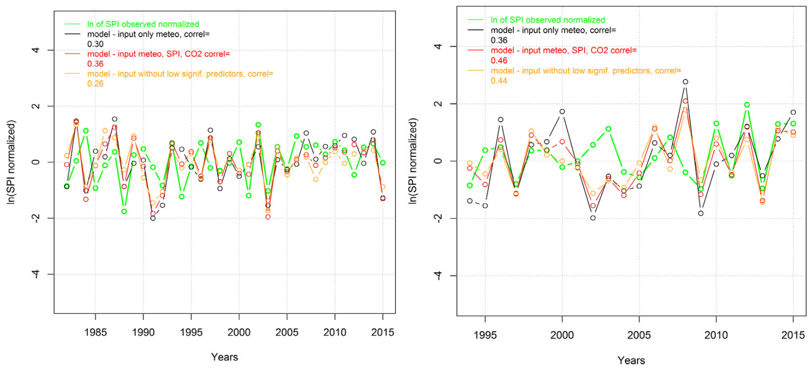


Fig. 7. Examples of the model performance outside the identified region: Brussels (left-hand panel) and Moscow (right-hand panel).

The multi-model approach, albeit being easier in each region, leads to discontinuities of the parameterizations at the borders of the delineated areas, where also none of the models is good. Therefore, it can be recommended only as an intermediate step revealing the governing parameters in each region and suggesting their (linearized) relations, which are then used for development of the unified model.

7. Conclusions

We demonstrated that the inter-annual variability of the birch seasonal pollen index SPI is synchronized over large regions of Europe significant correlations at the distances associated with the synoptic spatial scale.

A predictive model was constructed for the region covering Finland, Sweden, and Baltic States, part of Belarus, and, probably, part of Russia and Norway, where the lack of data did not allow for conclusive analysis.

Three models were constructed based on: (i) previous-year meteorology and CO_2, (ii) several previous years of the SPI, (iii) combination of the previous-year meteorology, CO_2 and the SPI.

The best-performing model based on combination of the meteorological and aerobiological data from the preceding years showed correlation coefficient reaching up to 0.9 and successfully reproduced both bi-annual cycle of the SPI as well as the years when this cycle breaks down. Odds ratio is infinitely high for 50% of sites inside the region and the fraction of prediction falling within factor of 2 from observations, stays within 40–70%.

Meteo-only model also showed remarkably good prediction skills, which allow its usage in the areas with sparse or no pollen observational network. The model captured the bi-annual cycles and its breaking years, which highlights the key role of meteorology in formation of this cycle. The dynamic range of the variations is however understated by this model, pointing out the importance of the plant response to the meteorological stress.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2017.09.061.

Acknowledgements

The study was supported by: performance-based funding of University of Latvia Nr. AAP2016/B041//ZD2016/AZ03 within the “Climate change and sustainable use of natural resources” programme; National Research Program of Latvia LATENERGI 2014.10-4/VPP-1/27; Finnish Academy project APTA 266314 and Copernicus Atmospheric Monitoring Service CAMS; Finish Funding Agency for Innovations (Tekes) project Smart Pollen and Finnish Academy project SAPID.

Authors thank Nickolay Chashnikov for technical help with the model development.

References

Aguilera, F., Fornaciari, M., Ruiz-Valenzuela, L., Galan, C., Msallem, M., Dhiab, A. Ben, de la Guardia, C.D., del Mar Trigo, M., Bonofiglio, T., Orlandi, F., 2014. Phenological models to predict the main flowering phases of olive (*Olea europaea* L.) along a latitudinal and longitudinal gradient across the Mediterranean region. Int. J. Biometeorol. 59:629–641. https://doi.org/10.1007/s00484-014-0876-7.

Albertine, J.M., Manning, W.J., Da Costa, M., Stinson, K.A., Muilenberg, M.L., Rogers, C.A., 2014. Projected carbon dioxide to increase grass pollen and allergen exposure despite higher ozone levels. PLoS One 9:1–6. https://doi.org/10.1371/journal.pone.0111712.

Bastl, K., Kmenta, M., Pessi, A.M., Prank, M., Saarto, A., Sofiev, M., Bergmann, K.C., Buters, J.T.M., Thibaudon, M., Jäger, S., Berger, U., 2016. First comparison of symptom data with allergen content (Bet v 1 and Phl p 5 measurements) and pollen data from four European regions during 2009–2011. Sci. Total Environ. 548–549:229–235. https://doi.org/10.1016/j.scitotenv.2016.01.014.

Chuine, I., Belmonte, J., 2004. Improving Prophylaxis for Pollen allergies: Predicting the Time Course of the Pollen Load of the Atmosphere of Major Allergenic Plants in France and Spain 43, 65–80. https://doi.org/10.1080/00173130410019163.

Corden, J.M., Stach, A., Millington, W.M., 2002. A comparison of Betula pollen seasons at two European sites; Derby, United Kingdom and Poznan, Poland (1995–1999). Aerobiologia 18, 45–53.

Cunha, M., Ribeiro, H., 2015. Pollen-based predictive modelling of wine production: application to an arid region. Eur. J. Agron. https://doi.org/10.1016/j.eja.2015.10.008.

Dahl, A., Strandhede, S.-O., 1996. Predicting the intensity of the birch pollen season. Aerobiologia 12, 97–106.

Dahl, Å., Galan, C., Hajkova, L., Pauling, A., Sikoparija, B., Smith, M., Vokou, D., 2013. The onset, course and intensity of the pollen season. In: Sofiev, M., Bergmann, K.-C. (Eds.), Allergenic Pollen. A Review of the Production, Release, Distribution and Health Impacts. Springer, Dordrecht, Heidelberg, New York, London:pp. 29–70 https://doi.org/10.1007/978-94-007-4881-1.

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

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137:553–597. https://doi.org/10.1002/qj.828.

Detandt, M., Nolard, N., 2000. The fluctuations of the allergenic pollen content of the air in Brussels (1982 to 1997). Aerobiologia 16, 55–61.

Dhiab, A. Ben, Dominguez-vilches, E., Dhiab, A. Ben, Mimoun, M. Ben, Oteros, J., Garcia-mozo, H., 2016. Modeling olive-crop forecasting in Tunisia. https://doi.org/10.1007/s00704-015-1726-1.

Galán, C., Smith, M., Thibaudon, M., Frenguelli, G., Oteros, J., Gehrig, R., Berger, U., Clot, B., Brandao, R., 2014. Pollen monitoring: minimum requirements and reproducibility of analysis. Aerobiologia 30:385–395. https://doi.org/10.1007/s10453-014-9335-5.

Garcia-Mozo, H., Yaezel, L., Oteros, J., Galan, C., 2014. Statistical approach to the analysis of olive long-term pollen season trends in southern Spain. Sci. Total Environ. 473–474:103–109. https://doi.org/10.1016/j.scitotenv.2013.11.142.

Grewling, Ł., Jackowiak, B., Nowak, M., Uruska, A., Smith, M., 2012. Variations and trends of birch pollen seasons during 15 years (1996–2010) in relation to weather conditions in Poznań (Western Poland). Grana 51 (4):280–292. https://doi.org/10.1080/00173134.2012.700727.

Hänninen, H., 1990. Modeling bud dormancy release in trees from cool and temperate regions. Acta For. Fenn. 213, 1–47.

Hartmann, D.L., Tank, A.M.G.K., Rusticucci, M., 2013. IPCC Fifth Assessment Report, Climatie Change 2013: The Physical Science Basis. Ipcc AR5. pp. 31–39.

Hatfield, J.L., Prueger, J.H., 2015. Temperature extremes: effect on plant growth and development. Weather Clim. Extrem. 10:4–10. https://doi.org/10.1016/j.wace.2015.08.001.
Hättestrand, M., Jensen, C., Hallsdóttir, M., Vorren, K.-D., 2008. Modern pollen accumulation rates at the north-western fringe of the European boreal. Rev. Palaeobot. Palynol. 151: 90–109. https://doi.org/10.1016/j.revpalbo.2008.03.001.
Hedhly, A., Hormaza, J.I., Herrero, M., 2009. Global warming and sexual plant reproduction. Trends Plant Sci. 14:30–36. https://doi.org/10.1016/j.tplants.2008.11.001.
Helbig, N., Vogel, B., Vogel, H., Fiedler, F., 2004. Numerical modelling of pollen dispersion on the regional scale. Aerobiologia 3, 3–19.
Hirst, J.M., 1954. An automatic volumetric spore trap. Ann. Appl. Biol. 39, 257–265.
Huynen, M., Menne, B., Behrendt, H., Bertollini, R., Bonini, S., Brandao, R., Clot, B., D'Ambrosio, C., De Nuntiis, P., Ebi, K., Emberlin, J., Orbanne, E., Galán, C., Jäger, S., Kovats, S., Mandrioli, P., Martens, P., Menzel, A., Nyenzi, B., Rantio-Lehtimäki, A., Ring, J., Rybnicek, O., Traidl-Hoffmann, C., Van Vliet, A., Voigt, T., Weiland, S., Wickman, M., 2003. Phenology and Human Health: Allergic Disorders. World Health Organisation.
Jato, V., Rodríguez-Rajo, F.J., Aira, M.J., 2007. Use of phenological and pollen-production data for interpreting atmospheric birch pollen curves. Ann. Agric. Environ. Med. AAEM 14, 271–280.
Ladeau, S.L., Clark, J.S., 2006. Pollen production by *Pinus taeda* growing in elevated atmospheric CO_2. Funct. Ecol. 20:541–547. https://doi.org/10.1111/j.1365-2435.2006.01133.x.
Latałowa, M., Mietus, M., Uruska, A., 2002. Seasonal variations in the atmospheric Betula pollen count in Gdansk (southern Baltic coast) in relation to meteorological parameters. Aerobiologia 18, 33–43.
Linsser, G., 1867. Die periodischen Erscheinungen des Pflanzenlebens in ihrem Verhältniss zu den Wärmeerscheinungen. Mem. l'Acad. Imp. d. sc. St Petersbg. XI. pp. 1–44.
Masaka, K., 2001. Modelling the masting behaviour of *Betula platyphylla* var. japonica using the resource budget model. Ann. Bot. 88:1049–1055. https://doi.org/10.1006/anbo.2001.1547.
Myszkowska, D., 2013. Prediction of the birch pollen season characteristics in Cracow, Poland using an 18-year data series. Aerobiologia 29:31–44. https://doi.org/10.1007/s10453-012-9260-4.
Orlandi, F., Romano, B., Fornaciari, M., 2005. Relationship between pollen emission and fruit production in olive (*Olea europaea* L.). Grana 44:98–103. https://doi.org/10.1080/00173130510010440.
Oteros, J., Garcia-Mozo, H., Hervas, C., Galan, C., 2013. Biometeorological and autoregressive indices for predicting olive pollen intensity biometeorological and autoregressive indices for predicting olive pollen intensity. Int. J. Biometeorol. 307–316. https://doi.org/10.1007/s00484-012-0555-5.
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.
Prasad, P.V.V., Craufurd, P.Q., Summerfield, 1999. Fruit number in relation to pollen production and viability in groundnut exposed to short episodes of heat stress. Ann. Bot. 84:381–386. https://doi.org/10.1006/anbo.1999.0926.
Puc, M., 2012. Artificial neural network model of the relationship between Betula pollen and meteorological factors in Szczecin (Poland). Int. J. Biometeorol. 56:395–401. https://doi.org/10.1007/s00484-011-0446-1.
Ranta, H., Satri, P., 2007. Synchronized inter-annual fluctuation of flowering intensity affects the exposure to allergenic tree pollen in North Europe. Grana 46:274–284. https://doi.org/10.1080/00173130701653079.
Ranta, H., Oksanen, A., Hokkanen, T., Bondestam, K., Heino, S., 2005. Masting by Betula-species; applying the resource budget model to north European data sets. Int. J. Biometeorol. 49: 146–151. https://doi.org/10.1007/s00484-004-0228-0.
Ranta, H., Hokkanen, T., Linkosalo, T., Laukkanen, L., Bondestam, K., Oksanen, A., 2008. Male flowering of birch: spatial synchronization, year-to-year variation and relation of catkin numbers and airborne pollen counts. For. Ecol. Manag. 255:643–650. https://doi.org/10.1016/j.foreco.2007.09.040.
Ritenberga, O., Sofiev, M., Kirillova, V., Kalnina, L., Genikhovich, E., 2016. Statistical modelling of non-stationary processes of atmospheric pollution from natural sources: example of birch pollen. Agric. For. Meteorol. 226–227:96–107. https://doi.org/10.1016/j.agrformet.2016.05.016.
Šaulienė, I., Šukienė, L., Severova, E., Kalnina, L., 2014. Comparison of *Alnus*, *Corylus*, *Betula* pollen seasons in Riga, Moscow and Vilnius. Aerobiologia 30:423–433. https://doi.org/10.1007/s10453-014-9341-7.
Severova, E., Volkova, O., 2016. Variations and trends of *Betula* pollen seasons in Moscow (Russia) in relation to meteorological parameters. Aerobiologia https://doi.org/10.1007/s10453-016-9460-4.
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. https://doi.org/10.1007/s00484-012-0539-5.
Simmons, A.J., Willett, K.M., Jones, P.D., Thorne, P.W., Dee, D.P., 2010. Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: inferences from reanalyses and monthly gridded observational data sets. J. Geophys. Res. Atmos. 115: 1–21. https://doi.org/10.1029/2009JD012442.
Sofiev, M., 2016. On impact of transport conditions on variability of the seasonal pollen index. Aerobiologia:1–13 https://doi.org/10.1007/s10453-016-9459-x.
Sofiev, M., Siljamo, P., Ranta, H., Rantio-Lehtimaki, A., 2006. Towards numerical forecasting of long-range air transport of birch pollen: theoretical considerations and a feasibility study. Int. J. Biometeorol. 50:392–402. https://doi.org/10.1007/s00484-006-0027-x.
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. https://doi.org/10.1007/s00484-012-0532-z.
Spieksma, F.T.M., Emberlin, J.C., Hjelmroos, M., Jäger, S., 1995. Atmospheric birch (Betula) pollen in Europe: trends and fluctuations in annual quantities and the starting dates of the seasons atmospheric birch (*Betula*) pollen in Europe: trends and fluctuations in annual quantities and the starting dates of the. Grana 34:51–57. https://doi.org/10.1080/00173139509429033.
Spieksma, F.T.M., Corden, J.M., Detandt, M., Millington, W.M., Nikkels, H., Nolard, N., Schoenmakers, C.H.H., Wachter, R., de Weger, L.A., Willems, R., Emberlin, J., 2003. Quantitative trends in annual totals of five common airborne pollen types (*Betula*, *Quercus*, *Poaceae*, *Urtica*, and *Artemisia*), at five pollen-monitoring stations in western Europe. Aerobiologia 19:171–184. https://doi.org/10.1023/B:AERO.0000006528.37447.15.
Stach, A., Smith, M., Prieto Baena, J.C., Emberlin, J., 2008. Long-term and short-term forecast models for Poaceae (grass) pollen in Poznań, Poland, constructed using regression analysis. Environ. Exp. Bot. 62:323–332. https://doi.org/10.1016/j.envexpbot.2007.10.005.
Stckli, R., Rutishauser, T., Baker, I., Liniger, M.A., Denning, A.S., 2011. A global reanalysis of vegetation phenology. J. Geophys. Res. Biogeosci. 116:1–19. https://doi.org/10.1029/2010JG001545.
Storkey, J., Stratonovitch, P., Storkey, J., Stratonovitch, P., Chapman, D.S., Vidotto, F., Semenov, M.A., 2014. A process-based approach to predicting the effect of climate change on the distribution of an invasive allergenic plant in Europe a process-based approach to predicting the effect of climate change on the distribution of an invasive allergenic plant in Eu. PLoS One:1–7 https://doi.org/10.1371/journal.pone.0088156.
Toro, F., Recio, M., Trigo, M.D.M., Cabezudo, B., 1998. Predictive models in aerobiology: data transformation. Aerobiologia 14, 179–184.
Veriankaitė, L., Siljamo, P., Sofiev, M., Šaulienė, I., Kukkonen, J., 2009. 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.
Williams, J.W., Gonzales, L.M., Kaplan, J.O., 2008. Leaf area index for northern and eastern North America at the Last Glacial Maximum: a data-model comparison. Glob. Ecol. Biogeogr. 17: 122–134. https://doi.org/10.1111/j.1466-8238.2007.00349.x.
Yli-Panula, E., Fekedulegn, D.B., Green, B.J., Ranta, H., 2009. Analysis of airborne *Betula* pollen in Finland; a 31-year perspective. Int. J. Environ. Res. Publ. Health 6:1706–1723. https://doi.org/10.3390/ijerph6061706.
Zhang, Y., Isukapalli, S., Bielory, L., Georgopoulos, P., 2013. Bayesian analysis of climate change effects on observed and projected airborne levels of *Birch* pollen. Atmos. Environ. 68: 64–73. https://doi.org/10.1016/j.atmosenv.2012.11.028.
Zhang, Y., Bielory, L., Georgopoulos, P.G., 2014. Climate change effect on *Betula* (birch) and *Quercus* (oak) pollen seasons in the United States. Int. J. Biometeorol. 58:909–919. https://doi.org/10.1007/s00484-013-0674-7.
Zhang, Y., Bielory, L., Mi, Zh., Cai, T., Robock, A., Georgopoulos, P., 2015. Allergelnic pollen season variations in the past two decades under changing climate in the United States. Glob. Chang. Biol. 21:1581–1589. https://doi.org/10.1111/geb.12755.
Ziello, C., Sparks, T.H., Estrella, N., Belmonte, J., Bergmann, K.C., Bucher, E., Brighetti, M.A., Damialis, A., Dedandt, M., Galan, C., Gehrig, R., Grewling, L., Guiterrez Bustillo, A.M., Hallsdottir, M., Kockhans-Bieda, C., Linares, C., Myskowska, D., Paldy, A., Sanches, A., Smith, M., Thibaudon, M., Travaglini, A., Uruska, A., Valencia Berra, R.M., Vokou, D., Wacher, R., de Weger, L.A., Menzel, A., 2012. Changes to airborne pollen counts across Europe. PLoS One 7 (4). https://doi.org/10.1371/journal.pone.0034076.
Ziska, L.H., Beggs, P.J., 2012. Anthropogenic climate change and allergen exposure: the role of plant biology. J. Allergy Clin. Immunol. 129:27–32. https://doi.org/10.1016/j.jaci.2011.10.032.
Ziska, L.H., Ghannoum, O., Baker, J.T., Conroy, J., Bunce, J.A., 2001. A global perspective of ground level, ambient carbon dioxide for assessing the response of plants to atmospheric CO_2. Glob. Chang. Biol. 7, 789–796.

III

Sofiev, M., Ritenberga, O., Siljamo, P., Albertini, R., Arteta, J., Belmonte, J., Bonini, M., Damialis, T., Elbern, H., Friese, E., Galan, C., Hrga, I., Kouznetsov, R., Plu, M., Prank, M., Robertson, L., Selenc, S., Thibaudon, M., Segers, A., Stepanovich, B., Valdebentino, A.M., Vira, J., Vokou, D.

Multi-model ensemble simulations of olive pollen distribution in Europe in 2014: current status and outlook

AAtmospheric Chemistry and Physics, 17, 12341–12360, 2017

Atmos. Chem. Phys., 17, 12341–12360, 2017
https://doi.org/10.5194/acp-17-12341-2017

Atmospheric Chemistry and Physics
Open Access
EGU

Multi-model ensemble simulations of olive pollen distribution in Europe in 2014: current status and outlook

Mikhail Sofiev[1], Olga Ritenberga[2], Roberto Albertini[3], Joaquim Arteta[4], Jordina Belmonte[5,6], Carmi Geller Bernstein[7], Maira Bonini[8], Sevcan Celenk[9], Athanasios Damialis[10,11], John Douros[12], Hendrik Elbern[13], Elmar Friese[13], Carmen Galan[14], Gilles Oliver[15], Ivana Hrga[16], Rostislav Kouznetsov[1,22], Kai Krajsek[17], Donat Magyar[18], Jonathan Parmentier[4], Matthieu Plu[4], Marje Prank[1], Lennart Robertson[19], Birthe Marie Steensen[20], Michel Thibaudon[15], Arjo Segers[21], Barbara Stepanovich[16], Alvaro M. Valdebenito[20], Julius Vira[1], and Despoina Vokou[11]

[1]Finnish Meteorological Institute, Erik Palmenin Aukio 1, Helsinki, Finland
[2]University of Latvia, Latvia
[3]Department of Medicine and Surgery, University of Parma, Italy
[4]CNRM UMR 3589, Météo-France/CNRS, Toulouse, France
[5]Institute of Environmental Sciences and Technology (ICTA), Universitat Autònoma de Barcelona, Spain
[6]Depatment of Animal Biology, Plant Biology and Ecology, Universitat Autònoma de Barcelona, Spain
[7]Sheba Medical Center, Ramat Gan Zabludowicz Center for Autoimmune Diseases, Israel
[8]Agenzia Tutela della Salute della Città Metropolitana di Milano/LHA ATS Città Metropolitana Milano, Italy
[9]Biology department, Uludag University, Turkey
[10]Chair and Institute of Environmental Medicine, UNIKA-T, Technical University of Munich and Helmholtz Zentrum München – German Research Center for Environmental Health, Augsburg, Germany
[11]Department of Ecology, School of Biology, Aristotle University of Thessaloniki, Greece
[12]Royal Netherlands Meteorological Institute, De Bilt, the Netherlands
[13]Rhenish Institute for Environmental Research at the University of Cologne, Germany
[14]Dpto. Botánica, Ecología y Fisiol. Vegetal, University of Cordoba, Spain
[15]RNSA, Brussieu, France
[16]Andrija Stampar Teaching Institute of Public Health, Croatia
[17]Institute of Energy and Climate Research (IEK-8), Forschungszentrum Jülich, Germany
[18]National Centre of Public Health, Hungary
[19]Swedish Meteorological and Hydrological Institute SMHI, Sweden
[20]MET Norway
[21]TNO, the Netherlands
[22]IAPh, Russian Academy of Sciences, Moscow, Russia

Correspondence to: Mikhail Sofiev (mikhail.sofiev@fmi.fi)

Received: 31 December 2016 – Discussion started: 2 February 2017
Revised: 26 August 2017 – Accepted: 31 August 2017 – Published: 17 October 2017

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract. The paper presents the first modelling experiment of the European-scale olive pollen dispersion, analyses the quality of the predictions, and outlines the research needs. A 6-model strong ensemble of Copernicus Atmospheric Monitoring Service (CAMS) was run throughout the olive season of 2014, computing the olive pollen distribution. The simulations have been compared with observations in eight countries, which are members of the European Aeroallergen Network (EAN). Analysis was performed for individual models, the ensemble mean and median, and for a dynamically optimised combination of the ensemble members obtained via fusion of the model predictions with observations. The models, generally reproducing the olive season of 2014, showed noticeable deviations from both observations and each other. In particular, the season was reported to start too early by 8 days, but for some models the error mounted to almost 2 weeks. For the end of the season, the disagreement between the models and the observations varied from a nearly perfect match up to 2 weeks too late. A series of sensitivity studies carried out to understand the origin of the disagreements revealed the crucial role of ambient temperature and consistency of its representation by the meteorological models and heat-sum-based phenological model. In particular, a simple correction to the heat-sum threshold eliminated the shift of the start of the season but its validity in other years remains to be checked. The short-term features of the concentration time series were reproduced better, suggesting that the precipitation events and cold/warm spells, as well as the large-scale transport, were represented rather well. Ensemble averaging led to more robust results. The best skill scores were obtained with data fusion, which used the previous days' observations to identify the optimal weighting coefficients of the individual model forecasts. Such combinations were tested for the forecasting period up to 4 days and shown to remain nearly optimal throughout the whole period.

1 Introduction

Biogenic aerosols, such as pollen and spores, constitute a substantial fraction of particulate matter mass in the air during the vegetation flowering season and can have strong health effects, causing allergenic rhinitis and asthma (D'Amato et al., 2007).

Olive is one of the most extensive crops and its oil is one of the major economic resources in southern Europe. The bulk of olive habitation (95 % of the total area worldwide) is concentrated in the Mediterranean basin (Barranco et al., 2008). Andalusia has by far the world's largest area of olive plantations: 62 % of the total olive land of Spain and 15 % of the world's plantations (Gómez et al., 2014).

Olive pollen is also one of the greatest causes of respiratory allergies in the Mediterranean basin (D'Amato et al., 2007), and in Andalusia it is considered the main cause of allergy. In Córdoba (southern Spain), 71–73 % of pollen-allergy sufferers are sensitive to olive pollen (Sánchez-Mesa et al., 2005; Cebrino et al., 2017). High rates of sensitization to olive pollen have been documented in Mediterranean countries: 44 % in Spain and 20 % in Portugal (Pereira et al., 2006), 31.8 % in Greece (Gioulekas et al., 2004), 27.5 % in Portugal (Loureiro et al., 2005), 24 % in Italy (Negrini et al., 1992), 21.6 % in Turkey (Kalyoncu et al., 1995), and 15 % in France (Spieksma, 1990). At the same time, relations between allergy and pollen concentrations are person- and case-specific: allergen content of the pollen grains varies from year to year and day to day, as well as the individual sensitivity of allergy sufferers (de Weger et al., 2013; Galan et al., 2013).

Olive is an entomophilous species that presents a secondary anemophily, favoured by the agricultural management during the last few centuries. This tree is very well adapted to the Mediterranean climate and tolerates the high summer and the low winter temperatures, as well as the summer drought, which is characteristic for this climate.

Olive floral phenology is characterised by bud formation during summer, dormancy during autumn, budburst in late winter, and flowering in late spring (Fernandez-Escobar et al., 1992; Galán et al., 2005; García-mozo et al., 2006). Similarly to some other trees, olive flowering intensity shows alternating years with high and low or even no pollen production. The characteristic quasi-biannual cycles are easily visible in the observations (Ben Dhiab et al., 2017; Garcia-Mozo et al., 2014). This cycle, similarly to that of other trees, e.g. birch, is not strict and is frequently interrupted, showing several years with similar flowering intensity (Garcia-Mozo et al., 2014). Such cyclic behaviour is related to the reproductive development, which is completed in two consecutive years. In the first year, the bud vegetative or reproductive character is determined by the current harvest level, since this is the main factor responsible for the interannual variation of flowering. In the second year, after the winter rest, the potentially reproductive buds that have fulfilled their chilling requirements develop into inflorescences (Barranco et al., 2008).

After budbreak, certain biothermic units are required for the development of the inflorescences. Both the onset of the heat accumulation period and the temperature threshold for the number of positive heat units might vary according to the climate of a determined geographical area. The threshold level was also reported to decrease towards the north (Aguilera et al., 2013). Altitude is the topographical factor that most influences olive local phenology and the major weather factors are temperature, rainfall, and solar radiation, which control plant evapotranspiration (Oteros et al., 2013, 2014).

Several studies used airborne pollen as a predictor variable for determining the potential sources of olive pollen emission, e.g. concentric ring method (Oteros et al., 2015a; Rojo et al., 2016), geostatistical techniques (Rojo and Pérez-

Badia, 2015), and the spatio-temporal airborne pollen maps (Aguilera et al., 2015).

There is substantial variability of the biological characteristics of olive and its responses to environmental stresses. In particular, the allergen content was shown to be strongly different in pollen from different parts of the Iberian Peninsula (Galan et al., 2013).

Forecasting efforts of the olive pollen season were mainly concentrated on statistical models predicting the start of the season and peak using various meteorological predictors. The bulk of studies is based on information from one or a few stations within a limited region (e.g. Orlandi et al., 2006; Moriondo et al., 2001; Alba and Diaz De La Guardia, 1998; Frenguelli et al., 1989; Galán et al., 2005; Fornaciari et al., 1998). Several wider-area studies were also undertaken to aim at more general statistical characteristics of the season, e.g. Aguilera et al. (2014, 2013), Galan et al. (2016).

Numerical modelling of olive pollen transport is very limited. In fact, the only regular regional-scale computations since 2008 were made by the SILAM model (http://silam.fmi.fi), but the methodology was only scarcely outlined in Galan et al. (2013).

The Copernicus Atmospheric Monitoring Service (CAMS; http://atmosphere.copernicus.eu) is one of the services of the EU Copernicus programme, and it addresses various global and regional aspects of atmospheric state and composition. The CAMS European air quality ensemble (Marécal et al., 2015) provides high-resolution forecasts and reanalysis of the atmospheric composition over Europe. Olive pollen is one of the components which is being introduced to the CAMS European ensemble in co-operation with the European Aeroallergen Network (EAN; https://www.polleninfo.org/country-choose.html).

One of the possible ways of improving the quality of model predictions without direct application of data assimilation is to combine them with observations via ensemble-based data fusion methods (Potempski and Galmarini, 2009). Their efficiency has been demonstrated for air quality problems (Johansson et al., 2015 and references therein) and climatological models (Genikhovich et al., 2010), but the technology has never been applied to pollen.

The aim of the current publication is to present the first Europe-wide ensemble-based evaluation of the olive pollen dispersion during the season of 2014. The study followed the approach of the multi-model simulations for birch (Sofiev et al., 2015a) with several amendments reflecting the peculiarity of olive pollen distribution in Europe. We also made further steps towards fusion model predictions and observations and demonstrate their value in the forecasting regime.

The next section will present the participating models and set-up of the simulations, the observation data used for evaluation of the model predictions, an approach for constructing an optimised multi-model ensemble, and a list of sensitivity computations. The Results section will present the outcome of the simulations and the quality scores of the individual models and the ensemble. Section 4 will be dedicated to analysis of the results, considerations of the efficiency of the multi-model ensemble for olive pollen, and identification of the development needs.

2 Materials and methods

This section presents the regional models used in the study, outlines the olive pollen source term implemented in all of them, and describes pollen observations used for evaluation of the model predictions.

2.1 Dispersion models

The dispersion models used in the study comprise the CAMS European ensemble, which is described in detail by Marécal et al. (2015) and Sofiev et al. (2015a). Below, only the model features relevant for the olive pollen atmospheric transport calculations are described.

The ensemble consisted of six models.

The EMEP model of EMEP MSC–West (European Monitoring and Evaluation Programme Meteorological Synthesizing Centre – West) is a chemical transport model developed at the Norwegian Meteorological Institute and described in Simpson et al. (2012). It is flexible with respect to the choice of projection and grid resolution. Dry deposition is handled in the lowest model layer. A resistance analogy formulation is used to describe dry deposition of gases, whereas for aerosols the mass-conservative equation is adopted from Venkatram (1978) with the dry deposition velocities dependent on the land-use type. Wet scavenging is dependent on precipitation intensity and is treated differently within and below cloud. The below-cloud scavenging rates for particles are based on Scott (1979). The rates are size-dependent, growing for larger particles.

EURAD-IM (http://www.eurad.uni-koeln.de) is an Eulerian mesoscale chemistry transport model involving advection, diffusion, chemical transformation, wet and dry deposition, and sedimentation of tropospheric trace gases and aerosols (Hass et al., 1995; Memmesheimer et al., 2004). It includes 3D-VAR and 4D-VAR chemical data assimilation (Elbern et al., 2007) and is able to run in nesting mode. The positive definite advection scheme of Bott (1989) is used to solve the advective transport and the aerosol sedimentation. An eddy diffusion approach is applied to parameterise the vertical subgrid-scale turbulent transport (Holtslag and Nieuwstadt, 1986). Dry deposition of aerosol species is treated size-dependent using the resistance model of Petroff and Zhang (2010). Wet deposition of pollen is parameterised according to Baklanov and Sorensen (2001).

LOTOS-EUROS (http://www.lotos-euros.nl/) is an Eulerian chemical transport model (Schaap et al., 2008). The advection scheme follows Walcek and Aleksic (1998). The dry deposition scheme of Zhang et al. (2001) is used to describe the surface uptake of aerosols. Below-cloud scavenging is described using simple scavenging coefficients for particles (Simpson et al., 2003).

MATCH (http://www.smhi.se/en/research/research-departments/air-quality/match-transport-and-chemistry-model-1.6831) is an Eulerian multiscale chemical transport model with mass-conservative transport and diffusion based on a Bott-type advection scheme (Langner et al., 1998; Robertson and Langner, 1999). For olive pollen, dry deposition is mainly treated by sedimentation and a simplified wet scavenging scheme is applied. The temperature sum, which drives pollen emission, is computed offline from January onwards and is fed into the emission module.

MOCAGE (http://www.cnrm.meteo.fr/gmgec-old/site_engl/mocage/mocage_en.html) is a multiscale dispersion model with grid-nesting capability (Josse et al., 2004; Martet et al., 2009). The semi-Lagrangian advection scheme of Williamson and Rasch (1989) is used for the grid-scale transport. The convective transport is based on the parameterisation proposed by Bechtold et al. (2001), whereas the turbulent diffusion follows the parameterisation of Louis (1979). Dry deposition including the sedimentation scheme follows Seinfeld and Pandis (1998). The wet deposition caused by convective and stratiform precipitation is based on Giorgi and Chameides (1986).

SILAM (http://silam.fmi.fi) is a meso- to global-scale dispersion model (Sofiev et al., 2015b), also described in the review of Kukkonen et al. (2012). Its dry deposition scheme (Kouznetsov and Sofiev, 2012) is applicable for a wide range of particle sizes including coarse aerosols, which are primarily removed by sedimentation. The wet deposition parameterisation distinguishes between sub- and in-cloud scavenging by both rain and snow (Sofiev et al., 2006). For coarse particles, impaction scavenging, parameterised following Kouznetsov and Sofiev (2012), is dominant below the cloud. The model includes emission modules for six pollen types: birch, olive, grass, ragweed, mugwort, and alder, albeit only birch, ragweed, and grass sources are so far described in the literature (Prank et al., 2013; Sofiev, 2017; Sofiev et al., 2012).

Three ENSEMBLE models were generated by (i) the arithmetic average, (ii) the median and (iii) an optimal combination of the six model fields. The average and median were taken on hourly basis, whereas optimisation was applied at daily level following the temporal resolution of the observational data. For the current work, we used a simple linear combination c_{opt} of the models c_m, $m = 1 \ldots M$, minimising the regularised RMSE J of the optimal field:

$$c_{\mathrm{opt}}(i,j,k,t,\tau,A) = a_0(\tau) + \sum_{m=1}^{M} a_m(\tau)\, c_m(i,j,k,t),$$
$$A = [a_1 \ldots a_M],\ \ a_m \geq 0 \forall m \tag{1}$$

$$J(t,\tau) = \sqrt{\frac{1}{O}\sum_{o=1}^{O}\left(c_{\mathrm{opt}}(i_\mathrm{o}, j_\mathrm{o}, k_\mathrm{o}, t, \tau, A) - c_\mathrm{o}(t)\right)^2} \tag{2}$$

$$+\alpha \sum_{m=1}^{M}\left(a_m(\tau) - \frac{1}{M}\right)^2 + \beta \sum_{m=1}^{M}\left(a_m(\tau-1) - a_m(\tau)\right)^2$$
$$\tau = \{d_{-k}, d_0\}. \tag{3}$$

Here, i, j, k, t are indices along the x, y, z, and time axes, M is the number of models in the ensemble, O is the number of observation stations, $\tau = \{d_{-k} : d_0\}$ is the time period of $k+1$ days covered by the analysis window, starting from d_{-k} to d_0, $\tau - 1$ is the previous-day analysis period $\tau - 1 = \{d_{-k-1} : d_{-1}\}$, c_m is the concentration of pollen predicted by the model m, c_o is observed pollen concentration, a_m is the time-dependent weight coefficient of the model m in the ensemble, a_0 is the time-dependent bias correction. In Eq. (2), the first term represents the RMSE of the assimilated period τ, the second term limits the departure of the coefficients from the homogeneous weight distribution, the third one limits the speed of evolution of the a_m coefficients in time. The scaling values α and β decide on the strength of regularisation imposed by these two terms.

The ensemble was constructed to mimic the forecasting mode. Firstly, the analysis is made using data from the analysis period τ. The obtained weighting coefficients a_i are used over several days from day d_0: from d_1 to d_{nf}, which constitute the forecasting steps. The performance of the ensemble is evaluated for each length of the forecast, from 1 to n_f days.

2.2 Olive pollen source term

All models of this study are equipped with the same olive pollen source term, which has not yet been described in the scientific literature. However, it follows the same concept as the birch source (Sofiev et al., 2012) that was used for the birch ensemble simulations (Sofiev et al., 2015a). The formulations and input data are available at http://silam.fmi.fi/MACC. The main input data set is the annual olive pollen production map based on the ECOCLIMAP data set (Champeaux et al., 2005; Masson et al., 2003), Fig. 1.

ECOCLIMAP incorporates the CORINE land-cover data for most western European countries with olive plantations as an explicit land-use type (CEC, 1993). For Africa and countries missing from CORINE, the empty areas were filled manually, assuming that 10 % of all tree-like land-use types are olives. This way, Tunisian, Egyptian, and Algerian olive plantations were recovered and included in the inventory. In some areas, such as France (Fig. 1), the olive habitat looks unrealistically low, probably because the large olive plantations are rare but the trees are planted in private gardens, city park areas, streets, etc. Since these distributed sources are not reflected in the existing land-use inventories, they are not included in the current pollen production map.

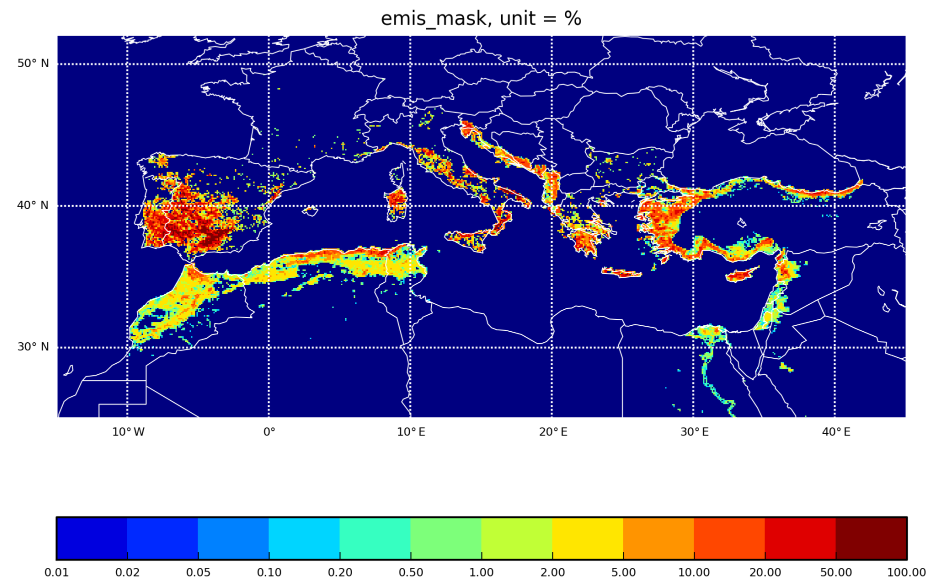


Figure 1. Olive habitat map, showing the percentage of the area occupied by the trees [%]. Productivity of an area with 100 % olive coverage is assumed to be 10^{10} pollen grain m^{-2} season^{-1}.

Similarly to birch, the flowering description follows the concept of thermal time phenological models and, in particular, the double-threshold air temperature-sum approach of Linkosalo et al. (2010) modified by Sofiev et al. (2012). Within that approach, the heat accumulation starts on a prescribed day in spring (1 January in the current set-up – after Spano et al., 1999; Moriondo et al., 2001; Orlandi et al. 2005a, b) and continues throughout spring. The cut-off daily temperature below which no summation occurs is 0 °C, in contrast to 3.5 °C for birch. It was obtained from the multi-annual fitting of the start of the season. Flowering starts when the accumulated heat reaches the start threshold (Fig. 2) and continues until the heat reaches the end threshold (in the current set-up, which is equal to the starting threshold +275 degree day). The rate of heat accumulation is the main controlling parameter for pollen emission: the model assumes direct proportionality between the flowering stage and fraction of the heat sum accumulated to date.

Similarly to the birch parameterisation in Sofiev et al. (2012), the model distinguishes between pollen maturation, which is solely controlled by the heat accumulation described above, and pollen release, which depends on other parameters. Higher relative humidity (RH) and rain reduce the release, completely stopping it for RH > 80 % and/or rain > 0.1 mm h^{-1}. Strong wind promotes it by up to 50 %. Atmospheric turbulence is taken into account via the turbulent velocity scale and thus becomes important only in cases close to free convection. In stable or neutral stratification and calm conditions the release is suppressed by 50 %. The interplay between pollen maturation and release is controlled by an intermediate ready pollen buffer, which is filled in by maturation and emptied by the release flows.

Local-scale variability of flowering requires a probabilistic description of its propagation (Siljamo et al., 2008). In the simplest form, the probability of an individual tree entering the flowering stage can be considered via the uncertainty of the temperature-sum threshold determining the start of flowering for the grid cell – 10 % in the current simulations. The end of the season is described via the open-pocket principle: the flowering continues until the initial available amount of pollen is completely released. The uncertainty of this number is taken to be 10 % as well.

2.3 Pollen observations

The observations for the model evaluation in 2014 have been provided by the following eight national networks, members of the EAN: Croatia, Greece, France, Hungary, Israel, Italy, Spain, and Turkey. The data were screened for completeness

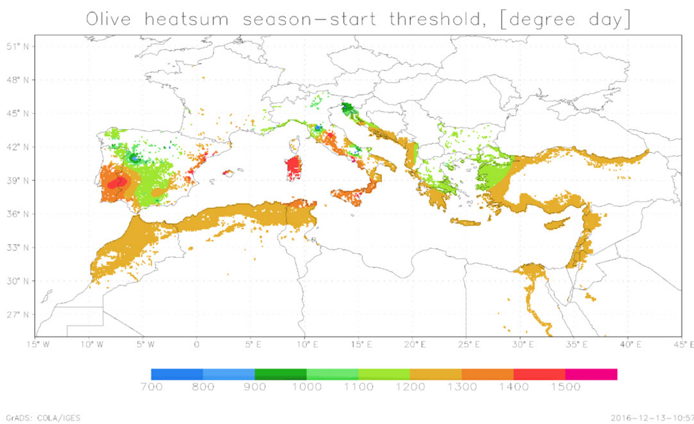


Figure 2. Heat-sum threshold for the start of the season. Unit = [° day].

and the existence of a significant olive season: (i) the time series should have at least 30 valid observations, (ii) at least 10 daily values during the season should exceed 3 pollen m^{-3}, and (iii) the seasonal pollen index (SPI, an integral of the concentrations over the whole season) should be at least 25 pollen day m^{-3}. After this screening, information from 62 sites was used in the intercomparison. Data from Hungary referred to 2016 and required dedicated computations for evaluating the long-range transport events.

Pollen monitoring was performed with Burkard 7 day and Lanzoni 2000 pollen traps based on the Hirst design (Hirst, 1952). The pollen grains were collected at an airflow rate of 10 L min^{-1}. The observations covered the period from March to September, with some variations between the countries. Daily pollen concentrations were used. Following the EAS-EAN requirements (Galán et al., 2014; Jäger et al., 1995), most samplers were located at heights between 10 and 30 m on the roofs of suitable buildings. The places were frequently downtown of the cities, i.e. largely representing the urban background conditions (but not always). With regard to microscopic analysis, the EAS-EAN requirement is to count at least 10 % of the sample using horizontal or vertical strips (Galán et al., 2014). The actual procedures vary between the countries but generally comply. The counting in 2014 was mainly along four horizontal traverses as suggested by Mandrioli et al. (1998). In all cases, the data were expressed as mean daily concentrations (pollen m^{-3}).

2.4 Set-up of the simulations

Simulations followed the standards of the CAMS European ensemble (Marécal et al., 2015). The domain spanned from 25° W to 45° E and from 30 to 70° N. Each of the six models was run with its own horizontal and vertical resolutions, which varied from 0.1 to 0.25° of the horizontal grid cell size and had from 3 up to 52 vertical layers within the troposphere (Table 1). This range of resolutions is not designed to reproduce local aspects of pollen distribution; instead it covers the whole continent and describes large-scale transport. The 10 km grid cells reach the sub-city scale but are still insufficient to resolve the valleys and individual mountain ridges. The limited number of vertical dispersion layers used by some models is a compromise, allowing for high horizontal resolution. Thick layers are not a major limitation as long as the full vertical resolution of the input meteorological data is used for evaluation of dispersion parameters (Sofiev, 2002).

The simulations were made retrospectively for the season of 2014, from 1 January (the beginning of the heat-sum accumulation) to 30 June. All models produced hourly output maps with concentrations at eight vertical levels (near the surface, 50, 250, 500, 1000, 2000, 3000 and 5000 m above the surface), as well as dry and wet deposition maps.

All models considered pollen as an inert water-insoluble particle, 28 µm in diameter, and with a density of 800 kg m^{-3}.

3 Results of the pollen season of 2014

3.1 Observed peculiarities of the season

At the French Mediterranean stations (Aix-en-Provence, Avignon, Montpellier, Nice, Nîmes, and Toulon), the mean value of the SPI in 2014 was quite similar to that of 2012 but lower than that in 2013 (see de Weger et al., 2013 for the SPI relevance to allergy).

Table 1. Set-up of the simulations for the participating models.

Model	Horizontal dispersion grid	Dispersion vertical	Meteo input	Meteo grid	Meteo vertical
EMEP	0.25° × 0.125°	20 levels up to 100 hPa	internal preprocessor	0.25° × 0.125°	IFS lvs 39–91 up to 100 hPa
EURAD-IM	15 km, Lambert conformal projection	23 layers up to 100 hPa	WRF based on ECMWF IFS	Same as CTM	Same as CTM
LOTOS-EUROS	0.25° × 0.125°	3 dyn. lyrs up to 3.5 km, sfc 25 m	ECMWF IFS 00 operational forecast, internal preprocessor	0.5° × 0.25°	IFS lvs 69–91 up to 3.5 km
MATCH	0.2° × 0.2°	52 layers up to 7 km	ECMWF IFS 00 from MARS, internal preprocessor	0.2° × 0.2°	IFS vertical: 91 lvs
MOCAGE	0.2° × 0.2°	47 layers up to 5 hPa (7 in ABL)	ECMWF IFS 00 operational forecast, internal preprocessor	0.125° × 0.125°	IFS vertical 91 lvs
SILAM	0.1° × 0.1°	9 layers up to 7.5 km	ECMWF IFS 00 operational forecast, internal preprocessor	0.125° × 0.125°	IFS lvs 62–137 up to ∼ 110 hPa

The start of the pollen season was earlier than in the previous 5 years. The duration of the season had been the longest one in Aix-en-Provence, Nice, and Nîmes since 2010. At the Ajaccio (Corsica) station, the SPI was higher in 2014 than at other stations, similarly to the situation in 2012.

In Andalusia, 2014 was the second warmest year of the last few decades but was more humid than usual, at 5 % above the typical relative humidity level (https://www.ncdc.noaa.gov/sotc/global/201413). However, after an intense olive flowering period in 2013, in 2014 the flowering intensity was lower and similar to 2012, in agreement with the biannual alterations of the season severity.

In northern Italy, the 2014 olive pollen season was less intense than the average of the previous 10 years (2004–2013). In contrast, in southern Italy, the 2014 season was more intense in the first part and less intense in the second part (after the beginning of June) than during previous seasons. No differences were noted with respect to the start and the end of the season in both cases.

In Thessaloniki, Greece, in 2014, the pollen season started at the same time as in the last few decades (first half of April) but ended about 1.5 months later (last half of October). The pollen season peak has been steady in May. The SPI was considerably higher in 2014 (418 pollen day m^{-3}) compared to the previous 2 years (approximately 300 pollen day m^{-3}). The overall shape of the pollen season in 2014 resembled that of the previous decade, but with a multimodal and smoother pattern.

3.2 Model results

The total seasonal olive pollen load (Figs. 3 and 4) expectedly correlates with the map of olive plantations (Fig. 1), which is also confirmed by the observations (Fig. 3). The highest load is predicted over Spain and Portugal, whereas the level in the eastern Mediterranean is not so high reflecting the smaller size of the areas covered by the olive trees and limited long-range transport over the Mediterranean. The model predictions differ up to a factor of 2–4 (Fig. 4), reflecting the diversity of modelling approaches, especially the deposition and vertical diffusion parameterisations (see Table 1 and Sect. 3.1).

Since the olive plantations are located within a comparatively narrow climatic range, flowering propagates through the whole region within a few weeks, starting from the coastal bands and progressing inland (not shown).

Hot weather during the flowering season leads to strong vertical mixing and a deep atmospheric boundary layer (ABL), which in turn promotes the pollen dispersion. As seen from Fig. 5, the pollen plumes can extend over the whole Mediterranean and episodically affect central Europe. Figs. 4 and 5 illustrate the differences between the models, e.g. substantially higher concentrations reported by EURAD-IM and MOCAGE compared to other models. The shortest transport with the fastest deposition is manifested by LOTOS-EUROS (also showing the lowest concentrations), while the longest one is suggested by MOCAGE.

The most important general parameters describing the season timing are its start and end (Fig. 6). Following Andersen (1991), these dates are computed as dates on which 5 and 95 % of the SPI are reached.

Computations of the model–measurement comparison statistics face the problem of non-stationarity and non-normal distribution of the daily pollen concentrations (Ritenberga et al., 2016). For such processes, the usual non-parametric statistics have to be treated with a great amount

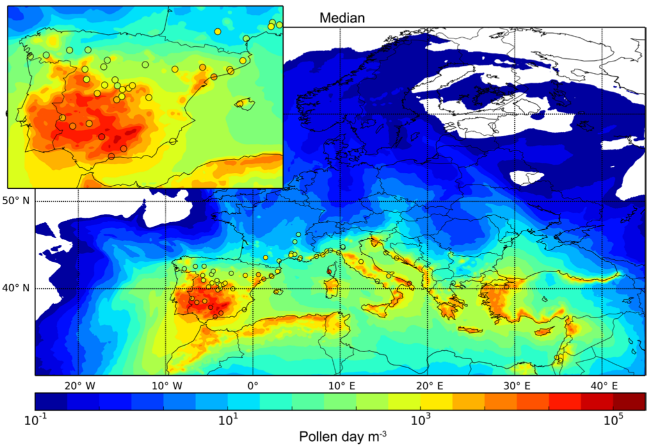


Figure 3. Observed (dots) and median-model-predicted (shades) seasonal pollen index (SPI, sum of daily concentrations), 2014 [pollen day m^{-3}].

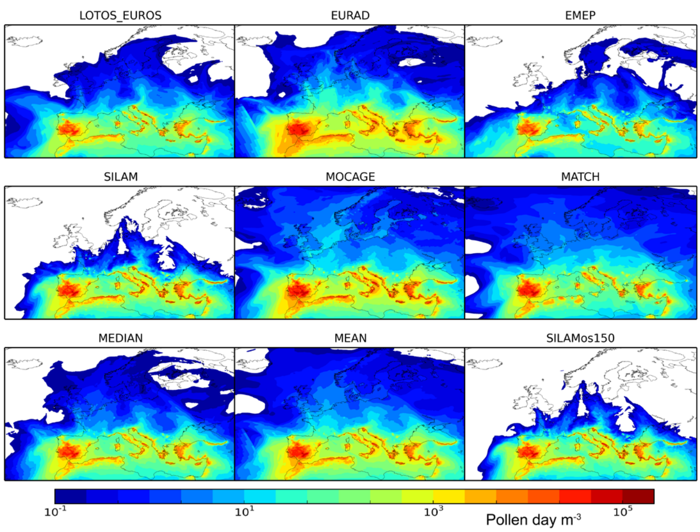


Figure 4. Modelled seasonal pollen index (SPI) by the individual ensemble members and mean models, 2014 [pollen day m^{-3}].

of care, since their basic assumptions are violated. Nevertheless, they can be formally calculated for both individual models and the ensemble (Figs. 7 and 8). The main characteristic of the ensemble, the discrete rank histogram, and the distribution of the modelled values for the below-detection-limit observations (Fig. 9) show that the spread of the obtained ensemble is somewhat too narrow in comparison with

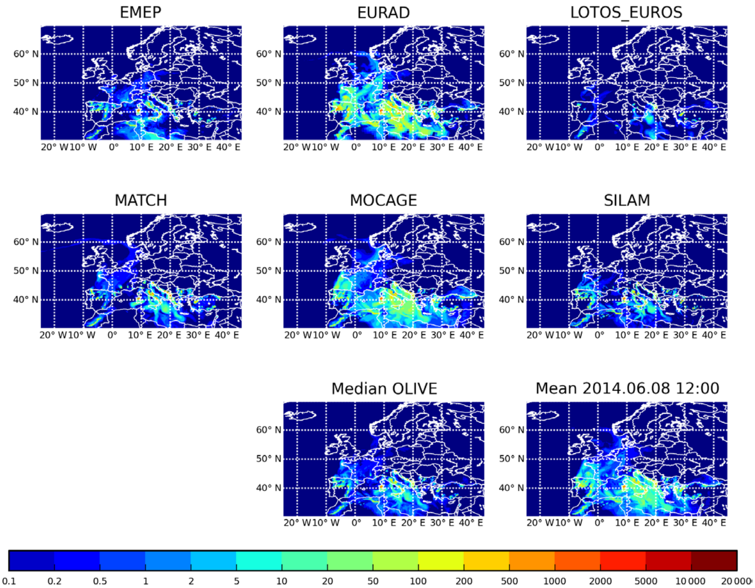


Figure 5. Example of hourly olive pollen concentrations: 12:00 UTC on 8 June 2014 [pollen m^{-3}].

the dynamic range of the observations. The same limitation was noticed for the birch ensemble.

The patterns in Figs. 6 and 7 reveal a systematic early bias of the predicted start and end of the season, which is easily seen from normalised cumulative concentration time series (Fig. 10). This bias is nearly identical for all models, except for EURAD-IM, which also shows a higher correlation coefficient than other models. The reasons for the problem and for the diversity of the model response are discussed in the next section.

4 Discussion

In this section, we consider the key season parameters and the ability of the presented ensemble to reproduce them (Sect. 4.1), the added value of the multi-model ensembles, including the optimised ensemble (Sect. 4.2), the main uncertainties that limit the model scores (Sect. 4.3), and the key challenges for future studies (Sect. 4.4).

4.1 Forecast quality: model predictions for the key season parameters

The key date of the pollen season is its start: this very date refers to adaptation measures that need to be taken by allergy sufferers. Predicting this date for olives is a significantly greater challenge than, for example, for birches: the heat sum has to be accumulated from 1 January with the season onset being in mid-April, whereas for birches the dates are 1 March and mid-March, respectively. As a result, the prediction of the start of the olive season strongly depends on the temperature predictions by the weather prediction model and the way this temperature is integrated into the heat sum. Inconsistency between these factors over the period of almost 4 months, even if small, can easily lead to a week of error. As one can see from Figs. 7 and 8, there is a systematic albeit spatially inhomogeneous bias of all models by up to 10 days (too early season). The exception is the SILAMos150 sensitivity run, which used the higher heat-sum threshold, by 150 degree days ($\sim 10\,\%$), than the standard level (Fig. 2). No other sensitivity runs, including the simulations driven by ERA-Interim fields, showed any significant improvement of this parameter. Importantly, EURAD-IM, which is driven by WRF meteorology fields, also showed a similar bias. Finally, the shift varies among the stations from near-zero (France, some sites in Italy, Croatia, Greece, and Israel) up to almost 3 weeks in north-western Spain. It means that no "easy" solution exists and an analysis of long-term time series is called for, aiming at a refinement of the heat-sum formulations and threshold values.

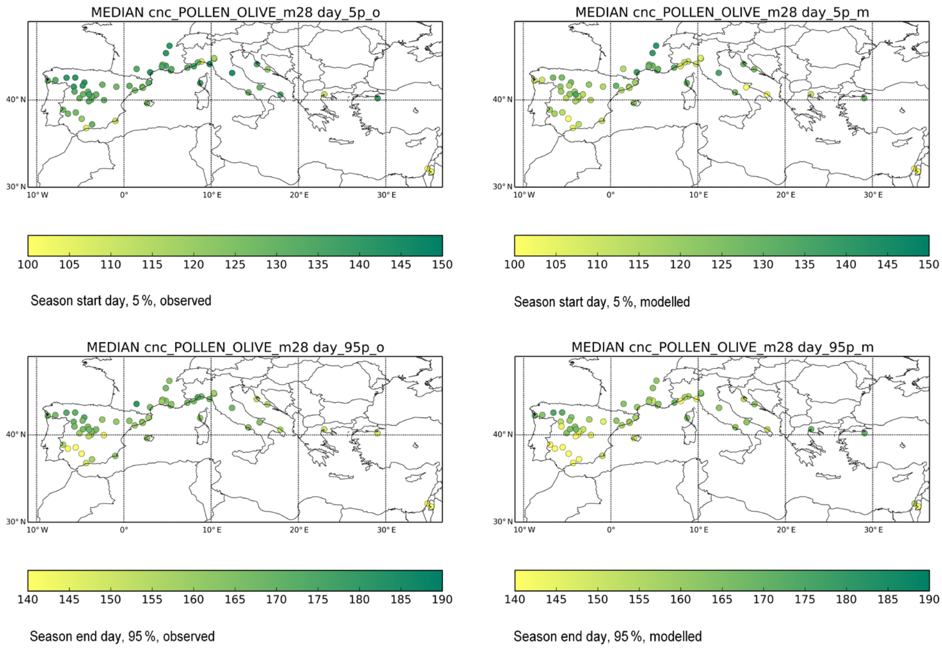


Figure 6. The start (5 % of the cumulative seasonal concentrations) and the end (95 % of the cumulative seasonal concentrations) of the olive season in 2014 as day of the year, predicted by the median of the ensemble and observed by the stations with a sufficient number of observations.

The end of the season showed an intriguing picture: EURAD-IM, despite starting the season as early as all other models, ends it 2 days too late instead of 5 days too early as all other models (see examples for two stations in Fig. 10). This indicates that WRF, in late spring, predicts a lower temperature than IFS, which leads to a longer-than-observed season in the EURAD-IM predictions. Other models showed the correct season length and, due to initial early bias, end it a few days too early. The de-biased run SILAMos150 shows an almost perfect shape and has within a 1 day accuracy for the start and end, which supports a 250 degree day as a season length parameter.

The most divergent model predictions are shown for the absolute concentrations (Fig. 8). With the mean observed April–June concentration of 35 pollen m^{-3} the range of predictions spans over a factor of four: EURAD-IM and MOCAGE being two times as high and EMEP and LOTOS-EUROS two time as low. Shifting the season by 5 days in the SILAMos150 run also changes the model bias, reflecting differences in the transport patterns and the impact of stronger vertical mixing in later spring. Spatially, the bias is quite homogeneous, except for southern Spain, where a heterogeneous pattern is controlled by local conditions at each specific site (Fig. 7).

Temporal correlation is generally high in coastal areas (Fig. 7) but at or below 0.5 at terrestrial stations on the Iberian Peninsula (the main olive plantations). This is primarily caused by the shifted season: the simulations with more accurate seasons showed the highest correlations among all models with $\sim 60\,\%$ of sites having a significant correlation ($p < 0.01$, Fig. 8).

Comparison with local statistical models made for single or a few closely located stations expectedly shows that local models are usually comparable to but somewhat more accurate (at their locations) than the European-scale dispersion models; also see discussion in Ritenberga et al. (2016). Thus, (Galan et al., 2001) analysed the performance of three popular local models for Córdoba, with the best one showing the mean error of 4.7 days at the start of the season but reaching up to 14 days in some years. A similar error was found for Andalusia (Galán et al., 2005) and two sites (Perugia and Ascoli Piceno) in Italy (Frenguelli et al., 1989) – 4.8 and

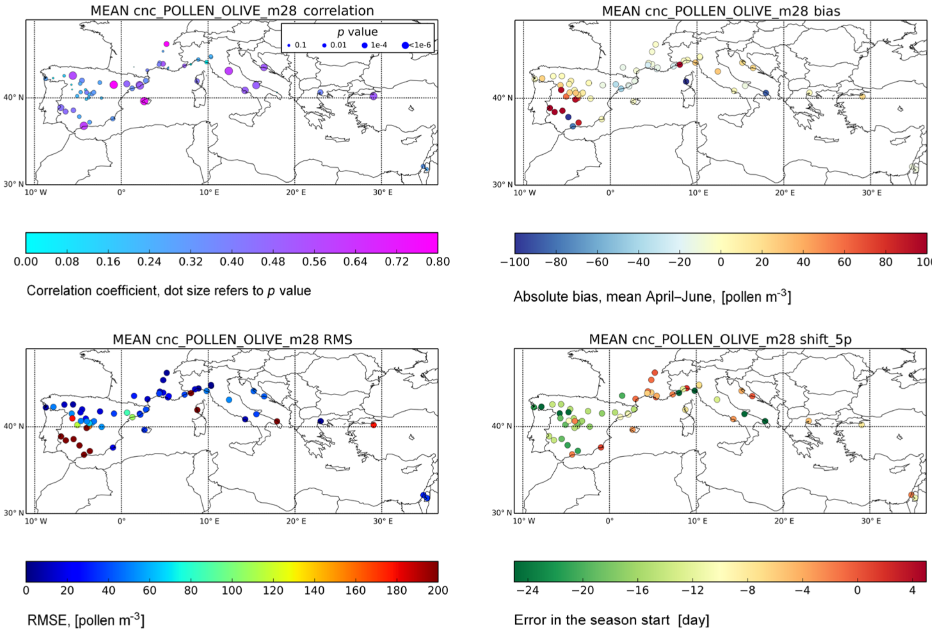


Figure 7. Results of model–measurement comparison for the ensemble mean: correlation coefficient for daily time series, mean bias April–June (pollen m^{-3}), RMSE (pollen m^{-3}), and error at the start of the season (days).

4.33 days of the standard error, respectively. A recent study (Aguilera et al., 2014) constructed three independent statistical models for Spain, Italy, and Tunisia and ended up with over 5 days of standard error for the Mediterranean. In another study, the authors admitted the scale of the challenges: "The specific moment for the onset of the olive heat accumulation period is difficult to determine and has essentially remained unknown" (Aguilera et al., 2013).

One of the strengths of continental-scale dispersion models is their ability to predict long-range transport events. However, a direct evaluation of this feature for olive pollen is difficult, since countries without olive plantations usually do not count its pollen. One can, however, refer to Fig. 3 (zoomed map of Spain), which shows that the ensemble successfully reproduces the drastic change of the SPI from nearly 10^5 pollen day m^{-3} in the south of Spain down to less than 100 pollen day m^{-3} in the north. Episode-wise, an example of a well-articulated case of olive pollen transport from Italy to Hungary in 2016 was brought up by Udvardy et al. (2017), who analysed it with adjoining SILAM simulations. The episode was also predicted well by the forward computations.

4.2 Ensemble added value

Arguably the main uncertainty of the model predictions was caused by the shift of the season start and end – the parameters were heavily controlled by temperature, i.e. least affected by transport features of the models. As a result, application of the "simple" ensemble technologies does not lead to a strong improvement. Some effect was still noticed, but it was less significant than in case of birch or traditional AQ forecasting. Therefore, in this section we also consider the possibility of ensemble-based fusion of the observational data with the model predictions. All ensembles were based on operational models; i.e. the SILAMos150 run was not included in either of them.

4.2.1 Mean ensembles: arithmetic average and median

Considering the mean-ensemble statistics, one should keep in mind that both the meteorological driver and the source term parameterisation were the same for all models (except for EURAD driven by WRF). This resulted in the underrepresentative ensemble (Fig. 9), for which several good and bad

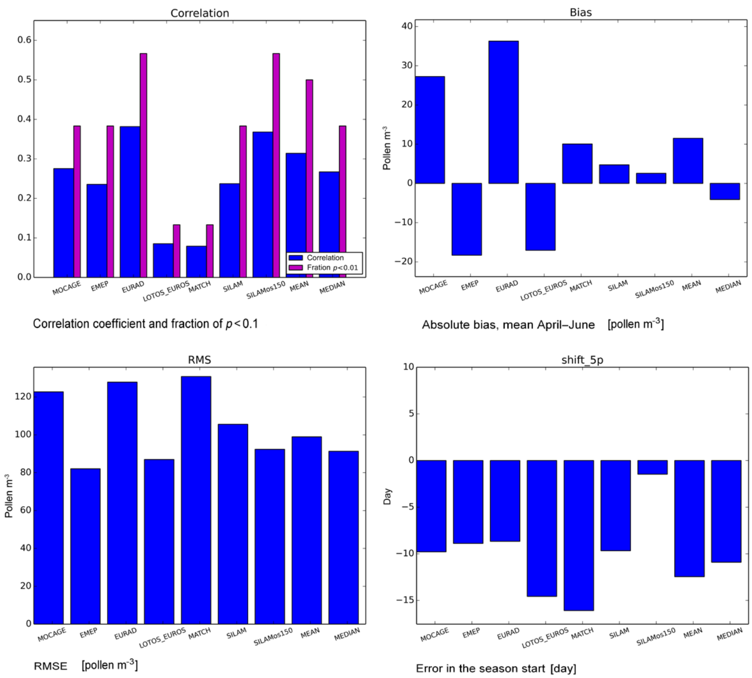


Figure 8. Scores of the individual models, mean over all stations. The same parameters as in Fig. 7. The sensitivity run SILAMos150 is explained in the discussion section.

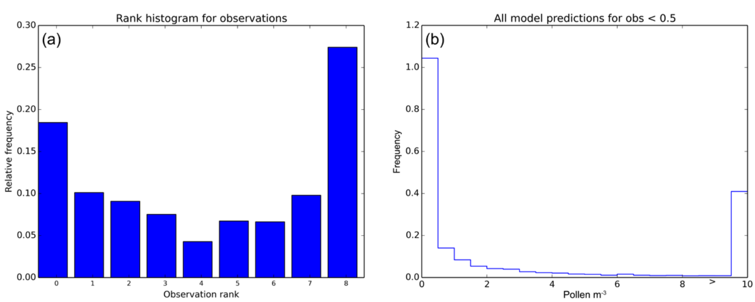


Figure 9. Ensemble characteristics. **(a)** Discrete rank histogram for the constructed ensemble (daily concentration statistics); **(b)** histogram of model predictions when observations were below the detection limit of 0.5 pollen m^{-3}.

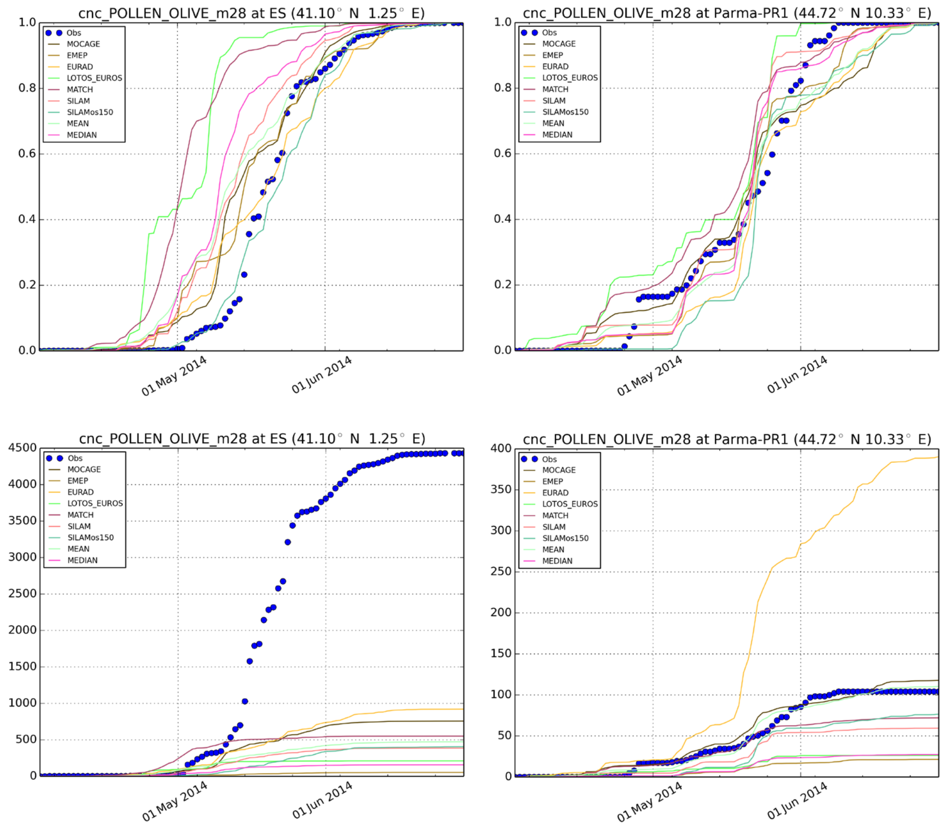


Figure 10. Cumulative time series of olive concentrations at Tarragona (Spain) and Parma (Italy). Upper row: normalised to the seasonal SPI [relative unit], lower row: absolute cumulative concentrations [pollen day m^{-3}].

features visible in all models propagate to the mean ensembles.

Among the simple means, the arithmetic average performed better than the median, largely owing to strong EURAD-IM impact. That model overestimated the concentrations and introduced a powerful push towards an extended season, thus offsetting the early bias of the other models. Since the median largely ignored this push, its performance was closer to that of other models. Nevertheless, both the mean and median demonstrated low RMSE, the median being marginally better.

4.2.2 Fusing the model predictions and observations into an optimised ensemble: gain in the analysis and predictive capacity

Developing the ensemble technology further, here we present the first attempt of fusion of the observational data with the multi-model ensemble for olive pollen.

In Sect. 3.1, Eq. (2) requires three parameters: the regularisation scaling parameters α and β, and length of the assimilation window T. For the purposes of the current feasibility study, several values for each of the parameters were tested and the robust performance of the ensemble was confirmed with very modest regularisation strength and for all considered lengths of the analysis window from 1 to 15 days.

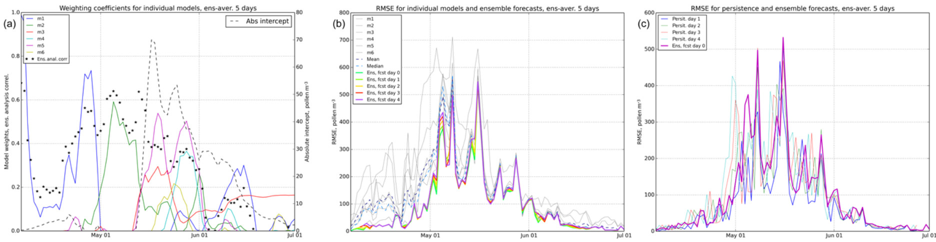


Figure 11. Optimal weights of the individual models and ensemble correlation score over the 5-day-long assimilation window (panel **a**); RMSE of the of individual models and the optimal ensemble forecasts against those of individual models and simple ensemble means **(b)**, and against persistence-based forecasts (panel **c**).

Finally, $\alpha = 0.1$, $\beta = 0.1$, $T = 5$ days were selected for the example below as a compromise between the smoothness of the coefficients, regularisation strength, and the optimisation efficiency over the assimilation window.

The optimised ensemble showed (Fig. 11a) that each of the six models had a substantial contribution over certain parts of the period. Over some times, e.g. during the first half of May, only one or two models were used, other coefficients being put to zero, whereas closer to the end of the month, all models were involved. Finally, prior to and after the main season, concentrations were very low and noisy, so the regularisation terms of Eq. (2) took over and pushed the weights to an a priori value of 1/6.

The bulk of the improvements came in the first half of the season (Fig. 11b). After the third peak in the middle of May, the effect of assimilation becomes small and the optimisation tends to use the intercept to meet the mean value, whereas the model predictions become small and essentially uncorrelated with the observations. This corroborates with the observed 8-day shift of the season, which fades out faster in the models than in the observed time series (Fig. 10).

There was little reduction in the predictive capacity of the optimised ensemble when going out of assimilation window and towards the forecasts. In essence, only the first peak of the concentrations (and RMSE) is better off with shorter forecasts. For the rest of the season (before and after the peak) the 5-day assimilation window led to a robust combination of the models that stayed nearly optimal over the next 5 days.

Comparison with other forecasts expectedly shows that the optimised ensemble not only has significantly better skills than any of the individual models, but is up to 25–30 % better than the mean and median of the ensemble (Fig. 11b). A stronger competitor was the "persistence forecast" when the next-day concentrations are predicted to be equal to the last observed daily value. The 1-day persistence appeared to be the best possible "forecast", which shows an RMSE at the beginning of May that is two times lower than in the 1-day forecast of the optimal ensemble (Fig. 11c). However, 2-day persistence forecast had about the same RMSE as the ensemble, and 3- and 4-day predictions were poor.

The strong performance of the 1-day persistence forecast is not surprising and, with the current standards of the pollen observations, has no practical value: the data are always late by more than 1 day (counting can start only next morning and become available about midday). The second problem of the persistence forecast is that it needs actual data; i.e. the scarcity of pollen network limits its coverage. Thirdly, persistence loses its skills very fast: the day +2 forecast has no superiority over the optimal ensemble, whereas day +3 and +4 persistence-based predictions are useless. Finally, at local scale, state-of-the-art statistical models can outperform it – see discussion in Ritenberga et al. (2016).

One should, however, point out that the 1-day predicting power of the persistence forecast (or more sophisticated statistical models based on it) can be a strong argument for the future real-time online pollen monitoring. It's delay can be as short as 1 h (Crouzy et al., 2016; Oteros et al., 2015b). These data have good potential to be used for next-day predictions for the vicinity of the monitor.

4.3 Sensitivity of the simulations to model and source term parameters

The above-presented results show that arguably the most significant uncertainty was due to shifting the start and the end of the season. It originated from the long heat-sum accumulation (since 1 January), where even a small systematic difference between the meteorology driving the multi-annual fitting simulations and that used for operational forecasts causes a significant season shift by late spring. In some areas, the resolution of the NWP model plays a role as well: the complex terrain in the north of Spain and in Italy requires dense grids with which to resolve the valleys. Other possible sources of uncertainties might need attention.

To understand the importance of some key parameters, a series of perturbed runs of SILAM was made:

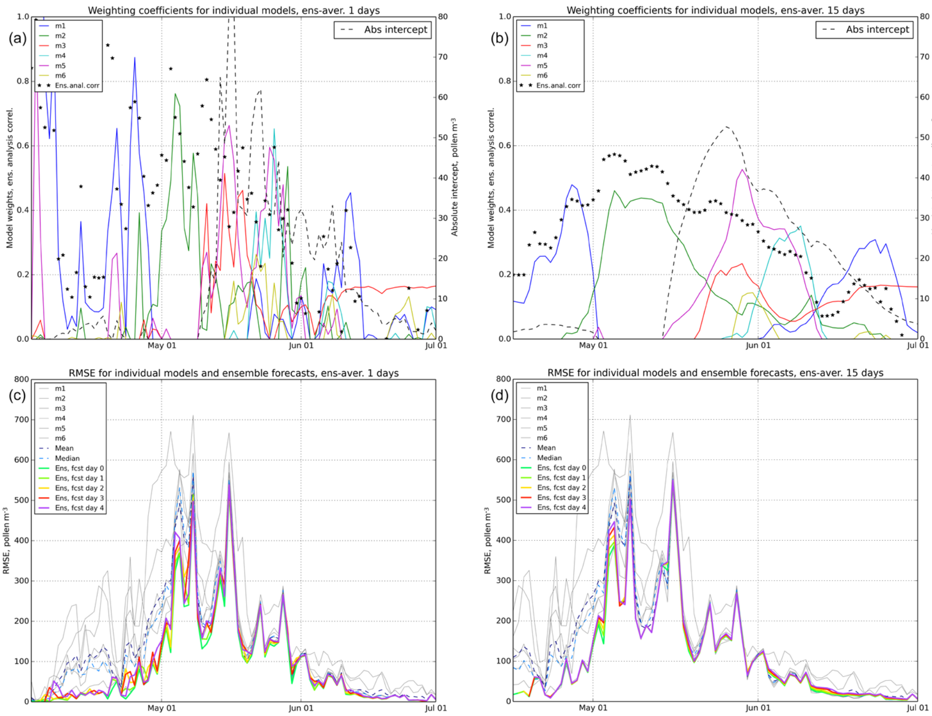


Figure 12. Sensitivity of optimised ensemble to the length of assimilation window. Upper row: optimal weights of the individual models and ensemble score over the 1- **(a)** and 15- **(b)** day-long assimilation windows; lower row: RMSE of the of individual models and the optimal ensemble forecasts against those of individual models. Note different time axes. Forecasts are available earlier for the 1-day analysis window.

- os100 and os150 runs with the season starting threshold increased by 100 and 150 degree days (the os150 run is referred in the above discussion as SILAMos150),
- ERA run with ERA-Interim meteorological fields, which were used for the source parameters fitting,
- series of three runs with reduced vertical mixing within the ABL and the free troposphere,
- smlpoll run with 20 µm size of the pollen grain,
- smlpoll_coarse run with 20 µm pollen size and coarse computational grid (0.2° × 0.2°).

The ERA simulations with ERA-Interim reduced the shift of the start of the season by 2 days but increased the shift of the end by 3 days, i.e. making the season shorter by 5 days. At the same time, the os150 run showed that a simple increase of the heat-sum threshold by ∼ 10 % (150 degree days) essentially eliminates the mean shift for 2014, but it remains unclear whether this adjustment is valid for other years.

Variations of the mixing parameterisation (perturbing the formula for the K_z eddy diffusivity) did not lead to significant changes: all scores stayed within 10 % of the reference SILAM simulations.

Evaluation of the impact of deposition parameterisations was more difficult since they are model specific. Higher deposition intensity causes both reduction of the transport distance and absolute concentrations. This issue might be behind the low values reported by LOTOS-EUROS and, conversely, the high concentrations of EURAD-IM and MOCAGE. Its importance was confirmed by the SILAM sensitivity simulations with smaller pollen size, smlpoll, and smlpoll_coarse. Both runs resulted in the mean concentrations more than doubling but with a marginal effect on temporal correlation. They also differed slightly from each other.

Variations of the fusion parameters showed a certain effect. For a short averaging window (5 days or less), the variations of weighting coefficients increased and the time series became noisier (Fig. 12). On return, the correlation increased almost up to 0.8–0.9 for some analysis intervals, but stayed the same for other periods. Also, the 1-day forecast RMSE decreased for some days but little difference was found for longer predictions.

4.4 Main challenges for the future studies

The current study is the first application of numerical models to olive pollen dispersion in Europe. One of its objectives was to identify the most pressing limitations of the current approach and the extent to which the ensemble and data fusion technologies can help in improving the forecasts.

The most evident issue highlighted by the exercise is the shift of the pollen season in some key regions, which is similar in all models, suggesting some unresolved inconsistencies between the heat-sum calculations of the source term and the features of the temperature predictions by the weather model. The issue suggests some factor(s) currently not included or misinterpreted in the source term. One of the candidate processes is the chilling-sum accumulation suggested by some studies, e.g. Aguilera et al. (2014). A switch to different types of phenological models with genetic differentiation of the populations following Chuine and Belmonte (2004) is another promising option.

The second issue refers to the underestimation of the pollen concentration in France, which probably originates from a comparatively large number of olive trees spread, for example, in private gardens but not accounted for in the agricultural maps of olive plantations.

The third set of questions refers to the pollen load prediction, i.e. the possibility of forecasting the overall season severity before it starts. Several statistical models have been presented in the literature, e.g. Ben Dhiab et al. (2017) for total annual load and Chuine and Belmonte (2004) for relative load. Their evaluation and implementation in the context of dispersion models is important.

An issue, mostly addressing the long-term horizon rather than the short-term forecasts, is the validity of the developed models in the conditions of changing climate. The models have to be robust to the trends in meteorological forcing. Purely statistical models are among the most vulnerable in this respect, because they just quantify the apparent correlations observed under certain conditions but do not explore the processes behind these relations.

Finally, the first steps towards the ensemble-based fusion of the model forecasts and pollen observations showed a strong positive effect. Further development of these techniques combined with progress towards near-real-time pollen data has very high potential for improving the forecasts.

5 Summary

An ensemble of six CAMS models was run through the olive flowering season of 2014 and compared with observational data of eight countries of European Aeroallergen Network (EAN).

The simulations showed a decent level of reproduction of the short-term phenomena but also demonstrated a shift of the whole season by about 8 days ($\sim 20\,\%$ of the overall pollination period). An ad-hoc adjustment of the starting heat-sum threshold by $\sim 10\,\%$ (150 degree days) on average resolves the issue and strongly improves the model skills, but its regional features and validity for other years and meteorological drivers remain unclear.

The ensemble members showed quite diverse pictures, demonstrating substantial variability, especially in areas remote from the main olive plantations. Nevertheless, the observation rank histogram still suggested a certain understatement of the ensemble variability in comparison with the observations. This partly originates from the synchronised source-term formulations and meteorological input used by all but one model.

Simple ensemble treatments, such as the arithmetic average and median, resulted in a more robust performance, but they did not outrun the best models over significant parts of the season. The arithmetic average turned out to be better than the median.

A data-fusion approach, which creates the optimal-ensemble model using the observations over preceding days for an optimal combination of the ensemble members, is suggested and evaluated. It was based on an optimal linear combination of the individual ensemble members and showed strong skills, routinely outperforming all individual models and simple ensemble approaches. It also showed strong forecasting skills, which allowed an application of the past-time model weighting coefficients over several days in the future. The only approach outperforming this fusion ensemble was the 1-day persistence-based forecast, which has no practical value due to the manual pollen observations and limited network density. It can, however, be used in the future when reliable online pollen observations become available.

A series of sensitivity simulations highlighted the importance of a meteorological driver, especially its temperature representation, and deposition mechanisms. The data fusion procedure was quite robust with regard to analysis window, still requiring 5–7 days to eliminate the noise in the model weighting coefficients.

Data availability. The model simulations presented in the paper are freely available on-request from the FMI team. The archive of 1.1 TB size is stored in the long-term FMI tape archive.

The Supplement related to this article is available online at https://doi.org/10.5194/acp-17-12341-2017-supplement.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The work was undertaken within the scope of Copernicus Atmospheric Monitoring Service (CAMS), funded by the European Union's Copernicus Programme. Support by performance-based funding of University of Latvia is also acknowledged. Observational data were provided by national pollen monitoring in Croatia, Greece, France, Italy (A.I.A.-R.I.M.A.®), Spain (REA: Aerocam, AeroUEx, RAA, UO, UC and Cantabria Health Council, REDAEROCAM, Health Castilla-Leon Council, RACyL, XAC, RIAG, PalinoCAM, UPCT, Basque Goverment/Public Health Directory), Hungary, Israel, and Turkey, members of the European Aeroallergen Network EAN. The olive source term is a joint development of the Finnish Meteorological Institute and EAN research teams, created within the scope of the Academy of Finland APTA project. This work contributes to the ICTA "Unit of Excellence" (MinECo, MDM2015-0552).

The material is published in the name of the European Commission; the commission is not responsible for any use that may be made of the information/material contained.

Edited by: Stefano Galmarini
Reviewed by: Slawomir Potempski and two anonymous referees

References

Aguilera, F., Ruiz, L., Fornaciari, M., Romano, B., Galán, C., Oteros, J., Ben Dhiab, A., Msallem, M., and Orlandi, F.: Heat accumulation period in the Mediterranean region: phenological response of the olive in different climate areas (Spain, Italy and Tunisia), Int. J. Biometeorol., 58, 867–876, 2013.

Aguilera, F., Fornaciari, M., Ruiz-Valenzuela, L., Galán, C., Msallem, M., Dhiab, A. Ben, la Guardia, C. D. de, del Mar Trigo, M., Bonofiglio, T., and Orlandi, F.: Phenological models to predict the main flowering phases of olive (Olea europaea L.) along a latitudinal and longitudinal gradient across the Mediterranean region, Int. J. Biometeorol., 59, 629–641, https://doi.org/10.1007/s00484-014-0876-7 2014.

Aguilera, F., Ben Dhiab, A., Msallem, M., Orlandi, F., Bonofiglio, T., Ruiz-Valenzuela, L., Galán, C., Díaz-de la Guardia, C., Giannelli, A., Trigo, M. M., García-Mozo, H., Pérez-Badia, R., and Fornaciari, M.: Airborne-pollen maps for olive-growing areas throughout the Mediterranean region: spatio-temporal interpretation, Aerobiologia, 31, 421–434, 2015.

Alba, F. and Diaz De La Guardia, C.: The effect of air temperature on the starting dates of the Ulmus, Platanus and Olea pollen seasons in the SE Iberian Peninsula, Aerobiologia, 14, 191–194, 1998.

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

Baklanov, A. and Sorensen, J. H.: Parameterisation of radionuclide deposition in atmospheric long-range transport modelling, Phys. Chem. Earth Pt. B, 26, 787–799, 2001.

Barranco, D., Fernández-Escobar, R., and Rallo, L.: El Cultivo del Olivo, 8th Edn., Mundi-Prensa, Madrid, 2008.

Bechtold, P., Bazile, E., Guichard, F., Mascart, P., and Richard, E.: A mass-flux convection scheme for regional and global models, Q. J. Roy. Meteor. Soc., 127, 869–886, 2001.

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

Bott, A.: A positive definite advection scheme obtained by nonlinear renormalization of the advective fluxes, Mon. Weather Rev., 117, 1006–1016, 1989.

Cebrino, J., Portero de la Cruz, S., Barasona, M. J., Alcázar, P., Moreno, C., Domínguez-Vilches, E., and Galán, C.: Airborne pollen in Córdoba City (Spain) and its implications for pollen allergy, Aerobiologia, 33, 281–291, https://doi.org/10.1007/s10453-016-9469-8, 2017.

CEC: CORINE Land Cover Technical Guide, European Environment Agency, Luxembourg, 1993.

Champeaux, J. L., Masson, V., and Chauvin, F.: ECOCLIMAP: a global database of land surface parameters at 1 km resolution, Meteorol. Appl., 12, 29–32, https://doi.org/10.1017/S1350482705001519, 2005.

Chuine, I. and Belmonte, J.: Improving prophylaxis for pollen allergies?: Predicting the time course of the pollen load of the atmosphere of major allergenic plants in France and Spain, Grana, 43, 65–80, https://doi.org/10.1080/00173130410019163, 2004.

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

D'Amato, G., Cecchi, L., Bonini, S., Nunes, C., Annesi-Maesano, I., Behrendt, H., Liccardi, G., Popov, T., and van Cauwenberge, P.: Allergenic pollen and pollen allergy in Europe, Allergy, 62, 976–990, https://doi.org/10.1111/j.1398-9995.2007.01393.x, 2007.

de Weger, L. A., Bergmann, K.-C., Rantio-Lehtimäki, A., Dahl, Å., Buters, J. T. M., Dechamp, C., Belmonte, J., Thibaudon, M., Cecchi, L., Besancenot, J.-P., Gala, C., and Waisel, Y.: Impact of pollen, in: Allergenic Pollen, A Review of the Production, Release, Distribution and Health Impacts, edited by: Sofiev, M., Bergmann, K.-C., Springer Netherlands, Dordrecht, p. x + 247, https://doi.org/10.1007/978-94007-4881-1_6, 2013.

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmos. Chem. Phys., 7, 3749–3769, https://doi.org/10.5194/acp-7-3749-2007, 2007.

Fernandez-Escobar, R., Benlloch, M., Navarro, C., and Martin, G. C.: The time of floral induction in the olive, J. Am. Soc. Hortic. Sci., 117, 304–307, 1992.

Fornaciari, M., Pieroni, L., Ciuchi, P., and Romano, B.: A regression model for the start of the pollen season in Olea europaea, Grana, 37, 110–113, 1998.

Frenguelli, G., Bricchi, E., Romano, B., Mincigrucci, G., and Spieksma, F. T. M.: A predictive study on the beginning of the pollen season for Gramineae and Olea europaea L., Aerobiologia, 5, 64–70, 1989.

Galan, C., Garcia-Mozo, H., Carinanos, P., Alcazar, P., and Dominguez-Vilches, E.: The role of temperature in the onset of the Olea europaea L. pollen season in southwestern Spain, Int. J. Biometeorol., 45, 8–12, 2001.

Galan, C., Antunes, C., Brandao, R., Torres, C., Garcia-Mozo, H., Caeiro, E., Ferro, R., Prank, M., Sofiev, M., Albertini, R., Berger, U., Cecchi, L., Celenk, S., Grewling, L., Jackowiak, B., Jfiger, S., Kennedy, R., Rantio-Lehtimfiki, A., Reese, G., Sauliene, I., Smith, M., Thibaudon, M., Weber, B., Weichenmeier, I., Pusch, G., and Buters, J. T. M.: Airborne olive pollen counts are not representative of exposure to the major olive allergen Ole e. 1. Allergy, 68, 809–812, https://doi.org/10.1111/all.12144, 2013.

Galan, C., Alcazar, P., Oteros, J., Garcia-Mozo, H., Aira, M. J., Belmonte, J., Diaz de la Guardia, C., Fernandez-Gonzalez, D., Gutierrez-Bustillo, M., Moreno-Grau, S., Perez-Badia, R., Rodriguez-Rajo, J., Ruiz-Valenzuela, L., Tormo, R., Trigo, M. M., and Dominguez-Vilches, E.: Airborne pollen trends in the Iberian Peninsula, Sci. Total Environ., 550, 53–59, https://doi.org/10.1016/j.scitotenv.2016.01.069, 2016.

Galán, C., García-Mozo, H., Vázquez, L., Ruiz, L., De La Guardia, C. D., and Trigo, M.: 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, 2005.

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

García-Mozo, H., Galán, C., Jato, V., Belmonte, J., Díaz, C., Guardia, D., Fernández, D., Aira, M. J., Roure, J. M., Ruiz, L., Trigo, M. M., and Domínguez-vilches, E.: Quercus pollen season dynamics in the Iderian Peninsula: response to meteorological parameters and posible consequences of climate change, Ann. Agr. Env. Med., 13, 209–224, 2006.

Garcia-Mozo, H., Yaezel, L., Oteros, J., and Galan, C.: Statistical approach to the analysis of olive long-term pollen season trends in southern Spain, Sci. Total Environ., 473–474, 103–109, https://doi.org/10.1016/j.scitotenv.2013.11.142, 2014.

Genikhovich, E., Pavlova, T. V., and Kattsov, V. M.: On complexing the ensemble of climate models (in Russian: O komplexirovanii ansamblya klimaticheskih modelej), Proc. Voeikov Main Geoiphysical Observatory, 7, 28–46, 2010.

Giorgi, F. and Chameides, W. L.: Rainout lifetimes of highly soluble aerosols and gases as inferred from simulations with a general circulation model, J. Geophys. Res.-Atmos., 91, 14367–14376, https://doi.org/10.1029/JD091iD13p14367, 1986.

Gioulekas, D., Papakosta, D., Damialis, A., Spieksma, F. T. M., Giouleka, P., and Patakas, D.: Allergenic pollen records (15 years) and sensitization in patients with respiratory allergy in Thessaloniki, Greece, Allergy, 59, 174–184, 2004.

Gómez, J. A., Infante-Amate, J., González de Molina, M., Vanwalleghem, T., Taguas, E. V., and Lorite, I.: Olive cultivation, its impact on soil erosion and its progression into yield impacts in Southern Spain in the past as a key to a future of increasing climate uncertainty, Agriculture, 4, 170–198, https://doi.org/10.3390/agriculture4020170, 2014.

Hass, H., Jakobs, H. J., and Memmesheimer, M.: Analysis of a regional model (EURAD) near surface gas concentration predictions using observations from networks, Meteorol. Atmos. Phys., 57, 173–200, 1995.

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

Holtslag, A. A. and Nieuwstadt, F. T. M.: Scaling the atmospheric boundary layer, Bound.-Lay. Meteorol., 36, 201–209, 1986.

Johansson, L., Epitropou, V., Karatzas, K., Karppinen, A., Wanner, L., Vrochidis, S., Bassoukos, A., Kukkonen, J., and Kompatsiaris, I.: Fusion of meteorological and air quality data extracted from the web for personalized environmental information services, Environ. Modell. Softw., 64, 143–155, https://doi.org/10.1016/j.envsoft.2014.11.021, 2015.

Josse, B., Simon, P., and Peuch, V.: Radon global simulations with the multiscale chemistry and transport model MOCAGE, Tellus B, 56, 339–356, 2004.

Jäger, S., Mandroli, P., Spieksma, F., Emberlin, J., Hjelmroos, M., Rantio-Lehtimaki, A., and Al, E.: News, Aerobiologia, 11, 69–70, 1995.

Kalyoncu, A., Qoplii, L., Selguk, Z., Emri, A., Kolagan, B., and Kocabas, A.: Survey of the allergic status of patients with bronchial asthma in Turkey: a multicenter study, Allergy, 50, 451–456, 1995.

Kouznetsov, R. and Sofiev, M.: A methodology for evaluation of vertical dispersion and dry deposition of atmospheric aerosols. J. Geophys. Res.-Atmos., 117, D01202, https://doi.org/10.1029/2011JD016366, 2012.

Kukkonen, J., Olsson, T., Schultz, D. M., Baklanov, A., Klein, T., Miranda, A. I., Monteiro, A., Hirtl, M., Tarvainen, V., Boy, M., Peuch, V.-H., Poupkou, A., Kioutsioukis, I., Finardi, S., Sofiev, M., Sokhi, R., Lehtinen, K. E. J., Karatzas, K., San José, R., Astitha, M., Kallos, G., Schaap, M., Reimer, E., Jakobs, H., and Eben, K.: A review of operational, regional-scale, chemical weather forecasting models in Europe, Atmos. Chem. Phys., 12, 1–87, https://doi.org/10.5194/acp-12-1-2012, 2012.

Langner, J., Bergström, R., and Pleijel, K.: European Scale Modeling of Sulphur, Oxidized Nitrogen and Photochemical Oxidants, Model Dependent Development av Evaluation for the 1994 Growing Season, SMHI, Norkoping, 1998.

Linkosalo, T., Ranta, H., Oksanen, A., Siljamo, P., Luomajoki, A., Kukkonen, J., and Sofiev, M.: A double-threshold temperature sum model for predicting the flowering duration and relative intensity of Betula pendula and B. pubescens, Agr. Forest Meteorol., 150, 1579–1584, https://doi.org/10.1016/j.agrformet.2010.08.007, 2010.

Louis, J.-F.: A parametric model of vertical eddy fluxes in the atmosphere, Bound.-Lay. Meteorol., 17, 187–202, 1979.

Loureiro, G., Rabaca, M. A., Blanco, B., Andrade, S., Chieira, C., and Pereira, C.: Aeroallergens' sensitization in an allergic paediatric population of Cova da Beira, Portugal, Allergol. Immunopath., 33, 192–198, 2005.

Mandrioli, P., Comtois, P., and Levizzani, V. (Eds.): Methods in Aerobiology, Pitagora Editrice, Bologna, 1998.

Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R. L., Denier van der Gon, H. A. C., Drouin, A., Elbern, H., Emili, E., Engelen, R. J., Eskes, H. J., Foret, G., Friese, E., Gauss, M., Giannaros, C., Guth, J., Joly, M., Jaumouillé, E., Josse, B., Kadygrov, N., Kaiser, J. W., Krajsek, K., Kuenen, J., Kumar, U., Liora, N., Lopez, E., Malherbe, L., Martinez, I., Melas, D., Meleux, F., Menut, L., Moinat, P., Morales, T., Parmentier, J., Piacentini, A., Plu, M., Poupkou, A., Queguiner, S., Robertson, L., Rouïl, L., Schaap, M., Segers, A., Sofiev, M., Tarasson, L., Thomas, M., Timmermans, R., Valdebenito, Á., van Velthoven, P., van Versendaal, R., Vira, J., and Ung, A.: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, Geosci. Model Dev., 8, 2777–2813, https://doi.org/10.5194/gmd-8-2777-2015, 2015.

Martet, M., Peuch, V.-H., Laurent, B., B., M., and Bergametti, G.: Evaluation of long-range transport and deposition of desert dust with the CTM Mocage, Tellus B, 61, 449–463, 2009.

Masson, V., Champelaux, J.-L., Chauvin, F., Meriguet, C., and Lacaze, R.: A global database of land surface parameters at 1-km resolution in meteorological and climate models, J. Climate, 16, 1261–1282, 2003.

Memmesheimer, M., Friese, E., Ebel, A., Jakobs, H. J., Feldmann, H., Kessler, C., and Piekorz, G.: Long-term simulations of particulate matter in Europe on different scales using sequential nesting of a regional model, Int. J. Environ. Pollut., 22, 108–132, 2004.

Moriondo, M., Orlandini, S. De, P., N., and Mandrioli, P.: Effect of agrometeorological parameters on the phenology of pollen emission and production of olive trees (Olea europea L.), Aerobiologia, 17, 225–232, https://doi.org/10.1023/A:1011893411266, 2001.

Negrini, A. C., Ariano, R., Delbono, G., Ebbli, A., Quaglia, A., Arobba, D., Allergologia, A., Paolo, O. S., Ligure, P., and Sv, I.-P. L.: Incidence of sensitisation to the pollens of Urticaceae (Parietaria), Poaceae and Oleaceae (Olea europaea) and pollen rain in Liguria (Italy), Aerobiologia, 8, 355–358, 1992.

Orlandi, F., Romano, B., and Fornaciari, M.: Effective pollination period estimation in olive (Olea europaea L.): A pollen monitoring application, Sci. Hortic.-Amsterdam, 105, 313–318, https://doi.org/10.1016/j.scienta.2005.01.012, 2005a.

Orlandi, F., Vazquez, L. M., Ruga, L., Bonofiglio, T., Fornaciari, M., Garcia-Mozo, H., Domínguez, E., Romano, B., and Galan, C.: Bioclimatic requirements for olive flowering in two mediterranean regions located at the same latitude (Andalucia, Spain, and Sicily, Italy), Annales Agricultural and Environmental Medicine Volume 12, 47–52, 2005b.

Orlandi, F., Lanari, D., Romano, B., and Fornaciari, M.: New model to predict the timing of olive (Olea europaea) flowering: a case study in central Italy, New Zeal. J. Crop Hort., 34, 93–99, https://doi.org/10.1080/01140671.2006.9514392, 2006.

Oteros, J., Garcia-Mozo, H., Vazquez, L., Mestre, A., Dominguez-Vilches, E., and Galan, C.: Modelling olive phenological response to weather and topography, Agr. Ecosyst. Environ., 179, 62–68, 2013.

Oteros, J., Orlandi, F., Aguilera, F., Ben, A., and Bonofiglio, T.: Better prediction of Mediterranean olive production using pollen-based models, Agron. Sustain. Dev., 34, 685–694, https://doi.org/10.1007/s13593-013-0198-x, 2014.

Oteros, J., García-Mozo, H., Alcázar, P., Belmonte, J., Bermejo, D., Boi, M., Cariñanos, P., Díaz de la Guardia, C., Fernández-González, D., González-Minero, F., Gutiérrez-Bustillo, A. M., Moreno-Grau, S., Pérez-Badia, R., Rodríguez-Rajo, F. J., Ruíz-Valenzuela, L., Suárez-Pérez, J., Trigo, M. M., Domínguez-Vilches, E., and Galán, C.: A new method for determining the sources of airborne particles, J. Environ. Manage., 155, 212–218, 2015a.

Oteros, J., Pusch, G., Weichenmeier, I., Heimann, U., Möller, R., Röseler, S., Traidl-Hoffmann, C., Schmidt-Weber, C., and Buters, J. T. M.: Automatic and online pollen monitoring, Int. Arch. Allergy Imm., 167, 158–166, https://doi.org/10.1159/000436968, 2015b.

Pereira, C., Valero, A., Loureiro, C., and Dávila, I.: Iberian study of aeroallergens sensitisation in allergic rhinitis, European Annals of Allergy and Clinical Immunology, 38, 186–194, 2006.

Petroff, A. and Zhang, L.: Development and validation of a size-resolved particle dry deposition scheme for application in aerosol transport models, Geosci. Model Dev., 3, 753–769, https://doi.org/10.5194/gmd-3-753-2010, 2010.

Potempski, S. and Galmarini, S.: *Est modus in rebus*: analytical properties of multi-model ensembles, Atmos. Chem. Phys., 9, 9471–9489, https://doi.org/10.5194/acp-9-9471-2009, 2009.

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., and Sofiev, M.: An operational model for forecasting ragweed pollen release and dispersion in Europe, Agr. Forest Meteorol., 182–183, 43–53, https://doi.org/10.1016/j.agrformet.2013.08.003, 2013.

Ritenberga, O., Sofiev, M., Kirillova, V., Kalnina, L., and Genikhovich, E.: Statistical modelling of non-stationary processes of atmospheric pollution from natural sources: example of birch pollen, Agr. Forest Meteorol., 226–227, 96–107, https://doi.org/10.1016/j.agrformet.2016.05.016, 2016.

Robertson, L. and Langner, J.: An Eulerian limited-area atmospheric transport model, J. Appl. Meteorol., 38, 190–210, 1999.

Rojo, J. and Pérez-Badia, R.: Spatiotemporal analysis of olive flowering using geostatistical techniques, Sci. Total Environ., 505, 860–869, 2015.

Rojo, J., Orlandi, F., Perez-Badia, R., Aguilera, F., Ben Dhiab, A., Bouziane, H., Diaz de la Guardia, C., Galan, C., Gutiarrez-Bustillo, A. M., Moreno-Grau, S., Msallem, M., Trigo, M. M., and Fornaciari, M.: Modeling olive pollen intensity in the Mediterranean region through analysis of emission sources, Sci. Total Environ., 551–552, 73–82, https://doi.org/10.1016/j.scitotenv.2016.01.193, 2016.

Sánchez-Mesa, J. A., Serrano, P., Cariñanos, P., Prieto-Baena, J, Moreno, C., Guerra, F., and Galan, C.: Pollen allergy in Cordoba city: frequency of sensitization and relation with antihistamine sales, J. Invest. Allerg. Clin., 15, 50–56, 2005.

Schaap, M., Timmermans, R. M. A., Roemer, M., Boersen, G. A. C., Builtjes, P. J. H., Sauter, F. J., Velders, G. J. M., and Beck, J. P.: The LOTOS-EUROS model: description, validation and latest developments, Int. J. Environ. Pollut., 32, 270–290, 2008.

Scott, B. C.: Parameterization of sulphate removal by precipitation, J. Appl. Meteorol., 17, 11275–11389, 1979.

Seinfeld, J. H. and Pandis, S. N.: Atmospheric Chemistry and Physics, 1st Edn., Wiley, New York, 1998.

Siljamo, P., Sofiev, M., Ranta, H., Linkosalo, T., Kubin, E., Ahas, R., Genikhovich, E., Jatczak, K., Jato, V., Nekovar, J., Minin, A., Severova, E., and Shalabova, V.: Representativeness of point-wise phenological Betula data collected in different parts of Europe, Global Ecol. Biogeogr., 17, 489–502, https://doi.org/10.1111/j.1466-8238.2008.00383.x, 2008.

Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Flechard, C. R., Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Nyíri, A., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J.-P., Valdebenito, Á., and Wind, P.: The EMEP MSC-W chemical transport model – technical description, Atmos. Chem. Phys., 12, 7825–7865, https://doi.org/10.5194/acp-12-7825-2012, 2012.

Simpson, D., Fagerli, H., Jonson, J. E., Tsyro, S., Wind, P., and Tuovinen, J.-P.: Transboundary Acidification, Eutrophication and Ground Level Ozone in Europe, Part 1: Unified EMEP Model Description, EMEP Report 1/2003, Met Norway, Oslo, 2003.

Sofiev, M.: Extended resistance analogy for construction of the vertical diffusion scheme for dispersion models, J. Geophys. Res., 107, ACH 10-1–ACH 10-8, https://doi.org/10.1029/2001JD001233, 2002.

Sofiev, M.: 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, 2017.

Sofiev, M., Siljamo, P., Valkama, I., Ilvonen, M., and Kukkonen, J.: 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, 2006.

Sofiev, M., Siljamo, P., Ranta, H., Linkosalo, T., Jaeger, S., Rasmussen, A., Rantio-Lehtimaki, A., Severova, E., and Kukkonen, J.: 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, 2012.

Sofiev, M., Berger, U., Prank, M., 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., and Peuch, V.-H.: 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, 2015a.

Sofiev, M., Vira, J., Kouznetsov, R., Prank, M., Soares, J., and Genikhovich, E.: Construction of the SILAM Eulerian atmospheric dispersion model based on the advection algorithm of Michael Galperin, Geosci. Model Dev., 8, 3497–3522, https://doi.org/10.5194/gmd-8-3497-2015, 2015b.

Spano, D., Cesaraccio, C., Duce, P., and Snyder, R. L.: Phenological stages of natural species and their use as climate indicators, Int. J. Biometeorol., 42, 124–133, https://doi.org/10.1007/s004840050095, 1999.

Spieksma, F.T: Pollinosis in Europe: new observations and developments, Rev. Palaeobot. Palyno., 64, 35–40, 1990.

Udvardy, O., Tedeschini, E., Sofiev, M., Palamarchuk, J., Makra, L., Kajtor-Apatini, D., and Magyar, D.: Utazo Allergenek Mediterran eredetu viragpor Magyarorszagon, Allergologia, June, 25–27, 2017 (in Hungarian).

Venkatram, A.: Estimating the convective velocity scale for diffusion applications, Bound.-Lay. Meteorol., 15, 447–452, 1978.

Walcek, C. J. and Aleksic, N. M.: A simple but accurate mass conservative, peak-preserving, mixing ratio bounded advection algorithm with FORTRAN code, Atmos. Environ., 32, 3863–3880, https://doi.org/10.1016/S1352-2310(98)00099-5, 1998.

Williamson, D. L. and Rasch, P.: Two-dimensional semi-Lagrangian transport with shape-preserving interpolation, Am. Meteorol. Soc., 117, 102–129, 1989.

Zhang, L., Gong, S., Padro, J., and Barrie, L.: A size-segregated particle dry deposition scheme for an atmospheric aerosol module, Atmos. Environ., 35, 549–560, 2001.