**ADAM MICKIEWICZ
UNIVERSITY
IN POZNAŃ**

Jakub Nowosad

Forecasting of

Corylus, *Alnus*, and *Betula* pollen concentration

in the air in Poland

Prognozowanie

stężeń pyłku leszczyny, olszy i brzozy

w powietrzu atmosferycznym na obszarze Polski

Doctoral dissertation

Praca doktorska

Supervisor

Promotor

dr hab. Alfred Stach, prof. UAM

Poznań 2016

Acknowledgements

List of publications

1. **Nowosad J.**, Stach A., Kasprzyk I., Latałowa M., Puc M., Myszkowska D., Weryszko-Chmielewska E., Piotrowska-Weryszko K., Chłopek K., Majkowska-Wojciechowska B., Uruska A. (2015) Temporal and spatiotemporal autocorrelation of daily concentrations of *Alnus*, *Betula*, and *Corylus* pollen in Poland. *Aerobiologia*, 31(2):159-177, doi: 10.1007/s10453-014-9354-2 (IF = 1.38)

2. **Nowosad J.**, Stach A., Kasprzyk I., Weryszko-Chmielewska E., Piotrowska-Weryszko K., Puc M., Grewling Ł., Pędziszewska A., Uruska A., Myszkowska D., Chłopek K., Majkowska-Wojciechowska B. (2015) Forecasting model of *Corylus*, *Alnus*, and *Betula* pollen concentration levels using spatiotemporal correlation properties of pollen count. *Aerobiologia*, doi: 10.1007/s10453-015-9418-y (IF = 1.38)

3. **Nowosad J.** (2015) Spatiotemporal models for predicting high pollen concentration level of *Corylus*, *Alnus*, and *Betula*. *International Journal of Biometeorology*, doi: 10.1007/s00484-015-1077-8 (IF = 3.25)

Abstract

Background: *Corylus*, *Alnus*, and *Betula* trees are among the most important sources of allergic pollen in the temperate zone of the Northern Hemisphere. The pollen concentrations of *Corylus*, *Alnus*, and *Betula* in the air have a large impact on the quality of life and productivity of allergy sufferers. The production, release, dispersal, transport, and deposition of pollen grains are controlled by interacting environmental factors. Understanding the behavior of atmospheric pollen concentration, as well as developing predictive models, can greatly help allergic sufferers. Such models could also be useful in agriculture, forestry, and many fields of science.

Aims:

1. To determine mean multi-year characteristics of temporal and space–time autocorrelation of the pollen counts of *Corylus*, *Alnus*, and *Betula* in Poland.
2. To create and evaluate *Corylus*, *Alnus*, and *Betula* pollen concentration levels predictions based on previous pollen count values from given sites.
3. To develop spatiotemporal predictive models of *Corylus*, *Alnus*, and *Betula* pollen concentration levels, using preprocessed gridded meteorological data.

Materials: The monitoring of the concentrations of *Corylus*, *Alnus*, and *Betula* pollen in the air was conducted in Bydgoszcz, Gdańsk, Kraków, Lublin, Łódź, Olsztyn, Poznań, Rzeszów, Siedlce, Sosnowiec, and Szczecin. The studies covered between 2 and 16 years of measurements. In addition, AGRI4CAST Interpolated Meteorological Data were used as predictor variables.

Methods:

1. Autocorrelation and cross-correlation functions were used to investigate temporal (one taxon, one location) and spatial (one taxon, two locations) patterns. Calculation of 61 correlograms and 84 cross-correlograms was performed for up to the previous 25 days. Mathematical models of correlograms were developed to smooth random fluctuation of the data and to assess the share and range of individual structures (factors). In addition, the cross-correlograms were divided into four classes using cluster analysis.

2. For each taxon, two levels of pollen concentration were distinguished according to the first allergy symptoms during exposure. Pollen level data were divided into a training set and a test set, using a stratified random split. The upsampling technique was used to reduce the class imbalance. The random forest method was used to predict the pollen concentration levels of *Corylus*, *Alnus*, and *Betula*. The previous 4 days' pollen counts from all of the studied sites were used as independent variables. The effect of predictors (independent variables) was calculated. The results of the final 24 models were evaluated using the test sets.

3. For each taxon, two levels of pollen concentration were distinguished according to the first allergy symptoms during exposure. The data was split into three sets: (i) a training set used for building the models; (ii) a first test set used to evaluate the performance of the models over time; (iii) a second test set used to evaluate the performance of the models in the unmeasured locations. The random forest method was used to predict the high pollen concentration levels of *Corylus*, *Alnus*, and *Betula*. An optimizing probability threshold technique was applied to the pollen level data to decrease the effect of class imbalance. Preprocessed meteorological variables based on gridded data were used as predictor variables. The importance of independent variables was determined using a mean decrease in accuracy. *Corylus*, *Alnus*, and *Betula* models were evaluated on two test sets.

Results:

1. Three main groups of factors were associated with temporal variations of *Corylus*, *Alnus*, and *Betula* pollen concentrations: (i) measurement errors and diurnal variability; (ii) exchanges of air masses after the passage of a single weather front; (iii) longer-lasting factors. An increase in the correlation between pollen concentrations in pairs of monitoring sites was found to be delayed by 1–3 days. This reflects the spatial configuration of monitoring stations and the prevailing circulation conditions in the pollen seasons of the taxa studied. Based on the character and range of autocorrelation, the monitoring sites were divided into three groups. These groups were found to have similar local topography and climate. In addition, it was found that the pollen characteristics in Gdańsk differ from those in the rest of the study sites.

2. The importance of independent variables showed temporal and spatial dependency. In most of the models, variables from a day before had the largest impact. Furthermore, independent variables from the same site as a dependent variable were the most important factors in 66% of the models. *Corylus* models showed the lowest values of performance statistics. This could be a mixed result of (i) an insufficient number of events with high *Corylus* pollen concentration levels, and (ii) a relatively short time series. On the other hand, it is possible to predict *Alnus* and *Betula* high pollen concentration levels using past pollen count data from monitoring sites. In addition, false predictions of high levels in *Corylus*, *Alnus*, and *Betula* models were connected with specific weather conditions.

3. Random forest models using preprocessed gridded meteorological data gave moderately reliable predictions of *Corylus* and highly reliable predictions of *Alnus* and *Betula* high pollen concentration levels. Cumulated growing degree days and 16-day averages of potential evaporation were the most important predictors.

Conclusions:

1. The study provided an understanding of the temporal autocorrelation and revealed

the main groups of factors connected with temporal variations of *Corylus*, *Alnus*, and *Betula* pollen concentrations. Space–time autocorrelation properties were determined. In addition, three groups of monitoring sites were distinguished. These results will be helpful in improving the quality of forecasting models.

2. *Corylus* models performed poorly. On the other hand, the study showed the possibility of predicting high *Alnus* and *Betula* pollen concentration levels with substantial accuracy using past pollen concentration data from monitoring sites. With development of semi-automatic and automatic systems for counting airborne pollen, these results can be used for building simpler models which require data only from aerobiological monitoring sites. Furthermore, it would be worthwhile to combine meteorological and aerobiological data for predicting high levels of pollen concentration.

3. *Corylus*, *Alnus*, and *Betula* high pollen concentration levels were predicted using preprocessed gridded meteorological data. The final models are capable of predicting pollen levels in continuous areas rather than in a single location. Furthermore, the modeling framework can be applied to predict pollen concentrations of the different pollen taxa in the study sites as well as in other areas.

6

Abstract (in Polish)

Tło badań: Leszczyna (*Corylus* L.), olsza (*Alnus* Mill.) i brzoza (*Betula* L.) są uznawane za jedne z najważniejszych źródeł pyłku wywołującego alergie w strefie klimatów umiarkowanych na półkuli północnej. Stężenia pyłku leszczyny, olszy i brzozy w powietrzu atmosferycznym mają znaczny wpływ na jakość życia oraz pracy alergików. Produkcja, uwalnianie, transport i rozprzestrzenianie się oraz depozycja ziaren pyłku jest wynikiem działania wielu współoddziałujących czynników środowiskowych. Zrozumienie dynamiki przebiegu pylenia i występowania pyłku w powietrzu, a także stworzenie modeli predykcyjnych, może istotnie pomóc alergikom. Takie modele mogą być także użyteczne w rolnictwie, leśnictwie oraz innych dziedzinach nauki.

Cele pracy:

1. Określenie średnich wieloletnich charakterystyk czasowej i czasoprzestrzennej autokorelacji dobowych stężeń pyłku leszczyny, olszy i brzozy w Polsce.

2. Stworzenie i ocena prognoz poziomów stężeń pyłku leszczyny, olszy i brzozy zbudowanych w oparciu o przeszłe dane pyłkowe z badanych stacji.

3. Zbudowanie czasoprzestrzennych modeli prognozujących poziomy stężeń pyłku leszczyny, olszy i brzozy z użyciem danych meteorologicznych.

Materiały: Monitoring średnich dobowych stężeń pyłku leszczyny, olszy i brzozy w powietrzu atmosferycznym był przeprowadzony w Bydgoszczy, Gdańsku, Krakowie, Lublinie, Łódzi, Olsztynie, Poznaniu, Rzeszowie, Siedlcach, Sosnowcu oraz Szczecinie i obejmował od dwóch do szesnastu lat pomiarów. Dodatkowo, jako zmienne niezależne,

wykorzystano pochodzące z reanaliz meteorologicznych dane w regularnej siatce (AGRI4CAST Interpolated Meteorological Data).

Metody:

1. Do zbadania wzorców zmienności czasowej stężeń pyłku leszczyny, olszy i brzozy w powietrzu atmosferycznym (jeden takson - jedno stanowisko) zostały użyte funkcje autokorelacji, natomiast do analizy zróżnicowania przestrzennego (jeden takson - para stanowisk) - funkcje kroskorelacji. Obliczono dla okresu 25 dni wstecz 61 korelogramów oraz 84 krosskorelogramy oraz dopasowano do nich modele (funkcje) matematyczne. Modelowanie miało na celu wygładzenie losowych fluktuacji krzywych empirycznych oraz umożliwienie interpolacji i ekstrapolacji wartości zmiennych, a w efekcie uzyskanie zasięgów oraz udziału poszczególnych struktur (czynników). Dodatkowo, wydzielono cztery grupy kroskorelogramów za pomocą analizy skupień.

2. Dla każdego z badanych taksonów, na podstawie wartości progowych wywołujących pierwsze symptomy alergiczne, wydzielono dwa poziomy stężeń pyłku. Dane poziomów stężeń pyłku zostały za pomocą stratyfikowanego próbkowania losowego podzielone na zbiór treningowy i testowy. Do zmiejszenia wpływu niezbilansowania liczby przypadków w badanych poziomach posłużyła technika nadpróbkowania (ang. *upsampling*). Poziomy stężeń pyłku leszczyny, olszy i brzozy były prognozowane za pomocą metodyki losowych lasów (ang. *random forest*). Jako zmienne niezależne użyto stężenia pyłku z czterech poprzednich dni ze wszystkich badanych stacji. Obliczono względne znaczenie poszczególnych, uwzględnionych w modelu, predyktorów (zmiennych niezależnych). Do oceny wynikowych 24 modeli posłużył niezależny zbiór testowy.

3. Na podstawie wartości progowych wywołujących pierwsze symptomy alergiczne, dla każdego z badanych taksonów wydzielono dwa poziomy stężeń pyłku. Zbiór danych został podzielony na trzy podzbiory: (i) podzbiór treningowy użyty do budowania modeli, (ii) pierwszy podzbiór testowy służący do oceny jakości modeli

w czasie, (iii) drugi podzbiór testowy użyty do oceny jakości modeli w przestrzeni (w lokalizacjach nieuwzględnionych przy konstrukcji modeli). Wysokie poziomy stężeń pyłku leszczyny, olszy i brzozy prognozowano wykorzystując metodykę losowych lasów. Wpływ efektu nierówności liczby przypadków w klasach został zmniejszony za pomocą techniki optymalizacji progu prawdopodobieństwa (ang. *optimizing probability threshold technique*). Jako zmienne niezależne użyto przetworzone dane pochodzące z reanaliz meteorologicznych w regularnej siatce. Istotność predyktorów określono stosując technikę średniej redukcji dokładności (ang. *mean decrease in accuracy*). Modele uzyskane dla poszczególnych taksonów poddano walidacji w oparciu o dwa zbiory testowe.

Wyniki:

1. Zmienność czasową stężeń pyłku leszczyny, olszy i brzozy powiązano z trzema grupami czynników: (i) błędami pomiarowymi oraz regularną i losową zmiennością stężeń w cyklu dobowym, (ii) wymianą mas powietrza po przejściu pojedynczego frontu atmosferycznego, (iii) czynnikami o dłuższym okresie powtarzalności. Pomiędzy stacjami pomiarowymi zaobserwowano opóźnienia lub wyprzedzenia wartości korelacji stężeń pyłku w atmosferze sięgające od jednego do trzech dni. Odzwierciedla to przestrzenny układ stacji pomiarowych oraz dominujące w okresie sezonów pyłkowych badanych taksonów warunki cyrkulacyjne. Na podstawie charakteru i zasięgu autokorelacji wyróżniono trzy grupy stanowisk. Grupy te łączą podobne cechy lokalnej topografii i klimatu. Dodatkowo zauważono, że przebieg i intensywność pylenia w Gdańsku różni się zdecydowanie od reszty badanych lokalizacji.

2. Istotność wpływu zmiennych niezależnych wykazała zależność czasową i przestrzenną. W większości modeli stężenia z dnia poprzedniego wykazywały największy wpływ na wyniki. Dodatkowo, w 66% modeli najbardziej istotne były zmienne niezależne z tej samej stacji. Najmniej dokładne prognozy dały modele dla leszczyny. Jest to efektem: (i) niewystarczającej liczby przypadków dni z wysokim poziomem

stężeń pyłku leszczyny, (ii) relatywnie krótkiej serii pomiarowej. Modele poziomów stężeń pyłku olszy i brzozy charakteryzowały się wysokimi wartościami miar jakości. Występowanie dni z wysokimi stężeniami pyłku olszy i brzozy można zatem prognozować z dużą trafnością. Wykazano także związek przypadków niepoprawnych prognoz wysokiego poziomu stężeń pyłku leszczyny, olszy i brzozy ze specyficznymi warunkami meteorologicznymi.

3. Modele lasów losowych wykorzystujące przetworzone dane meteorologiczne dały umiarkowanie pewne predykcje w przypadku leszczyny i wysoce sprawdzalne predykcje wysokich poziomów stężeń pyłku olszy i brzozy. Najważniejszymi predyktorami była suma efektywna temperatury (ang. *growing degree days*) oraz średnie szesnastodniowe wartości parowania potencjalnego.

Podsumowanie:

1. Badanie pozwoliło na określenie charakterystyk autokorelacji czasowej, jak i określenie głównych grup czynników wpływających na zmienność czasową stężeń pyłku leszczyny, olszy i brzozy. Scharakteryzowano także własności czasoprzestrzennej autokorelacji stężeń pyłku tych taksonów. Dodatkowo, wydzielono trzy grupy stacji pomiarowych. Uzyskane wyniki będą pomocne w ulepszaniu jakości modeli predykcyjnych.

2. Modele dla stężeń pyłku leszczyny miały niską sprawdzalność. Z drugiej strony, badanie wykazało, że przeszłe dane pyłkowe z wielu stacji pozwalają na prognozowanie poziomów stężeń pyłku olszy i brzozy ze znaczną dokładnością. Wraz z rozwojem półautomatycznych i automatycznych systemów mikroskopowych do zliczania pyłku, uzyskane wyniki powinny pozwolić na tworzenie prostszych modeli prognostycznych, wymagających jedynie danych pomiarowych ze stacji monitoringu aerobiologicznego. Dodatkowo wykazano, że wartościowym kierunkiem prac nad prognozowaniem wysokich poziomów stężeń pyłku byłaby próba integracji danych aerobiologicznych i meteorologicznych.

3. Wysokie poziomy stężeń pyłku leszczyny, olszy i brzozy były prognozowane

z użyciem przetworzonych danych meteorologicznych w regularnej siatce. Wynikowe, bardzo zadowalające, modele dają możliwość prognozowania poziomów stężeń na całym obszarze Polski, a nie tylko punktowo dla poszczególnych stanowisk. Dodatkowo, stworzony schemat postępowania może być zastosowany przy modelowaniu stężeń pyłku innych taksonów, jak i w innych obszarach.

Słowa kluczowe: pyłek alergizujący, stężenie pyłku w atmosferze, Betulaceae, autokorelacja czasoprzestrzenna, modele czasoprzestrzenne, modelowanie predykcyjne, uczenie maszynowe

Table of Contents

Chapter 1

Introduction

1.1 Preface

Corylus, *Alnus*, and *Betula* pollen are among the main causes of allergy in the temperate zone of the Northern Hemisphere. The prevalence of allergic diseases has increased over the last decades (World Allergy Organization 2013). In Europe, sensitization to tree pollen from the Betulaceae family is present in about 21–24% of the population (Heinzerling et al. 2009).

The production, release, dispersal, and deposition of pollen grains are affected by many environmental and genetic factors. The most important of these factors are meteorological variables such as temperature, wind, humidity, cloud cover, and precipitation. The relationship between these processes is highly complex and nonlinear. Understanding the behavior of atmospheric pollen concentrations, as well as developing predictive models, can greatly help allergic sufferers.

Most *Corylus*, *Alnus*, and *Betula* statistical models are based on the relationship between pollen concentration properties and meteorological conditions (Bringfelt et al. 1982; Cotos-Yáñez et al. 2004; Castellano-Méndez et al. 2005; Rodriguez-Rajo et al. 2006; Hilaire et al. 2012). In addition, these models are usually focused on temporal predictions.

There are also pollen forecast models that rely on physical transport and dispersion modeling (Vogel et al. 2008; Sofiev et al. 2013). These models are capable of spatial prediction of pollen concentrations. However, pollen dispersion models have common shortcomings. Firstly, they are dependent on information about distribution of pollen sources. Secondly, the quality of these models requires an understanding of the physical processes involved in pollen emission.

1.2   Pollen sources

Corylus L. (hazel), *Alnus* Mill. (alder), and *Betula* L. (birch) are members of the Betulaceae Grey family (Bremer et al. 2009). In Poland, *Alnus glutinosa*, *Alnus incana*, and *Betula pendula* are the predominant species from this family. The less common species are *Betula pubsescens*, *Corylus avellana*, and their cultivars.

Alnus and *Betula* are present in almost all of the European continent. However, their density varies among the regions. Five species of *Alnus* are present in Europe, with *Alnus glutinosa* and *Alnus incana* as the most common. Typical habitats of *Alnus* are forests and woodlands. *Alnus glutinosa* also often occurs in wet areas, such as bogs and river valleys (Tutin 1993). *Alnus* is not a predominant species on a large scale, with its percentages rarely exceeding 20%. The highest densities of this taxon are found in the Boreal region, including Poland, the Baltic countries (Lithuania, Latvia, Estonia), and Belarus (Skjøth et al. 2008). In Europe, *Betula* pollen-type sources are four native species (*Betula pubescens*, *Betula pendula*, *Betula humilis*, and *Betula nana*) and two non-native species (*Betula papyrifera* and *Betula utilis*) (Walters 1993). *Betula* are the predominant broadleaved species in Russia, Scandinavia, and Baltic countries but also have significant prevalence in Poland and Scotland (Skjøth et al. 2008). *Corylus* are mostly planted as ornamentals or in nut production fields (Tutin 1993). This taxon presents on a small scale in the main European forests (Skjøth et al. 2012). Although its density is low, *Corylus* covers most of Europe from Scandinavia to Italy. Three species (*Corylus avellana*, *Corylus*

colurna, and *Corylus maxima*) are *Corylus* pollen-type sources in Europe (Skjøth et al. 2012).

The location of allergic pollen sources could extend knowledge about levels of pollen concentration as well as be used in building pollen concentration models. However, there is a lack of information about the location of tree pollen sources. Skjøth et al. (2012) provided an overview of general spatial distributions of *Corylus*, *Alnus*, and *Betula* trees. There are also a few gridded inventories of pollen sources. Köble & Seufert (2001) calculated the distribution of 115 different tree species on the European scale. Skjøth et al. (2008) created a tree species inventory containing 39 groups of species covering Europe, parts of Africa, and parts of Asia. However, coarse spatial resolution of these inventories could not reveal all valuable information. The inventories also do not include city sources of tree pollen. In addition, statistical information about *Corylus* distribution is not available on a regional scale.

1.3 Production and transport of *Corylus*, *Alnus*, and *Betula* pollen

The number of pollen grains in the air varies in space and time. The production, emission, dispersion, transport, and deposition of pollen grains are controlled by interacting environmental factors (Dahl et al. 2013) (Fig. 1). The relationships between these factors are complex and nonlinear, and any change in these factors influences the pollen concentration in the air.

1.3.1 Pollen production

Pollen is a powder containing a plant's male sexual component. *Corylus*, *Alnus*, and *Betula* pollens are carried on the wind from the anther of a male flower to the stigma (female part) of another plant. This process is known as pollination. To increase the

chances of sexual reproduction, trees from the Betulaceae family produce large quantities of pollen.

This pollen production depends on the age and health of the trees. The pollen production is also determined by several site factors, such as soil type, topography, climate, and location. For example, pollen production decreases with altitude owing to the shorter growing season. Weather conditions during the previous growing season are also important (Frenguelli et al. 1991; Dahl & Strandhede 1996; Latałowa et al. 2002; Emberlin et al. 2007; Stach et al. 2008).
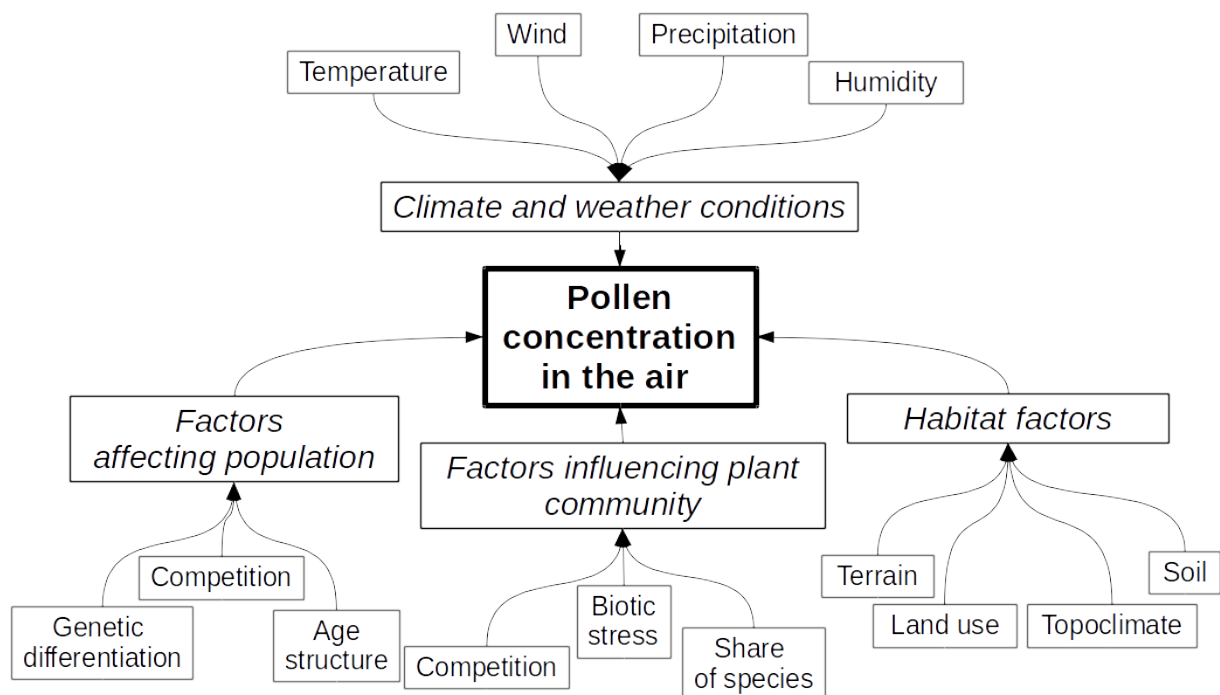
Figure 1: The factors affecting pollen concentration in the atmosphere

Each taxon produces different quantities of pollen. The European Aeroallergen Network calculation of annual pollen indexes (the sum of average daily pollen concentrations over a year) showed that for the period 2000–2009, the *Corylus* mean value was 398 (max=3,239); the mean value for *Alnus* was 1,048 (max=8,055); and the mean value for *Betula* was 3,782 (max=32,708) (Skjøth et al. 2012).

Furthermore, *Corylus*, *Alnus*, and *Betula* pollen production varies significantly from year to year (Dahl & Strandhede 1996; Weryszko-Chmielewska et al. 2001; Rodriguez-Rajo et al. 2006; Ranta & Satri 2007). It is possible that in one year many flowers and fruits are

produced in a population, and that in the following year production is low or moderate (Rodriguez-Rajo et al. 2006; Mahura et al. 2007). This process is known as masting, and a year of high abundance of pollen is called a mast year.

1.3.2 Pollen release

In most anemophilous trees, bursting of the anthers is controlled primarily by high temperature, low humidity, and moderate wind speed (Pacini & Hesse 2004). Thus *Corylus*, *Alnus*, and *Betula* pollination periods are dependent on weather patterns. The severity of *Corylus* and *Alnus* pollen seasons is influenced by rainfall and temperatures in the previous year, as well as by conditions at the time of pollen release and dispersal (Emberlin et al. 2007). Moreover, the start and duration of *Corylus* and *Alnus* pollen seasons vary considerably from year to year. The main pollination period of *Betula* is more stable, although the impact of temperature on pollen season properties is still evident (Méndez et al. 2005). In addition, a diurnal variation of these taxa has been reported (Galán et al. 1991; Norris-Hill & Emberlin 1991; Latałowa et al. 2005). Usually, *Corylus*, *Alnus*, and *Betula* pollen concentrations are marked by relatively low values at night and have their highest proportion of daily totals in the afternoon (Kasprzyk et al. 2001; Latałowa et al. 2005; Toth et al. 2011). The patterns of diurnal variation of *Alnus* and *Betula* also differ among sites and years (Berggren et al. 1995; Latałowa et al. 2005).

1.3.3 Pollen dispersal and deposition

The number of pollen grains in the air is a consequence of pollen-related processes in the atmosphere, on a wide range of scales (Sofiev et al. 2013):

1. The micro-scale: connected with pollen release. This is within a few meters of the plants.
2. The local scale: processes within the nearest kilometer of the source.
3. The regional scale: dispersion over distances up to 100 km.

4. The synoptic scale: dispersion up to 1,000–2,000 km.

5. The continental scale: processes over 5,000 km.

6. The global scale: processes over 5,000 km. The synoptic, continental, and global scales are often described as long-distance transport.

The distance pollen grains travel from a tree depends on the pollen grain size. The pollen grain sizes of *Corylus*, *Alnus*, and *Betula* are comparable. Accorsi et al. (1991) identified the average size of *Corylus* pollen grain as $18\,\mu m \times 23\,\mu m$, *Alnus* pollen grain as $22\,\mu m \times 34\,\mu m$, and *Betula* pollen grain as $18\,\mu m \times 28\,\mu m$. Sofiev et al. (2006) stated that owing to their low density, *Betula* pollen grains resemble the behavior of anthropogenic aerosols with a diameter smaller than $10\,\mu m$ (PM 10).

After the release of pollen grains into the atmosphere, the main factors influencing dispersal are gravity, wind speed and direction, and turbulence (Di-Giovanni & Kevan 1991). Most of the observed pollen grains come from local sources (Damialis et al. 2005). Only a small portion of pollen grains, called the "regional component", reaches beyond the local scale. The fraction of grains that will be a part of larger-scale dispersion depends on uplift turbulent motions (Gregory 1961), while the wind becomes the main transporting force at the meso-scale (Sofiev et al. 2013).

The number of pollen grains counted in a certain location can be significantly affected by episodes of long-distance transport from remote sources (Mahura et al. 2007). Oikonen et al. (2005) suggested that in some years, non-local *Betula* pollen contributes considerably to the total *Betula* pollen sum in Finnish Lapland. Ranta et al. (2006) found that long-distance transport of *Betula* pollen can greatly affect the timing of a local *Betula* pollen season. In addition, episodes of *Betula* pollen transport have been identified as coming from Poland and Germany to Denmark (Skjøth et al. 2007); from Russia to Finland (Siljamo et al. 2008); and from Latvia, Sweden, Denmark, Belarus, Ukraine, Moldova, Germany, and Poland to Lithuania (Veriankaite et al. 2009).

At all scales, there are several processes by which pollen grains are transferred from the air to solid surfaces: dry deposition, impact or gravitational sedimentation, and wet deposition as a consequence of precipitation scavenging (Sofiev et al. 2013). Occasionally, pollen grains after deposition may be resuspended back into the air.

1.4   Impact of pollen on human health

Corylus, *Alnus*, and *Betula* are considered to be among the most important sources of allergic pollen in the temperate climatic zone of the Northern Hemisphere (D'Amato et al. 2007). Proteins inside the pollen grains of these taxa have been identified as allergens (Weger et al. 2013). The human immune system can react to their presence and treat them as a threat; thus, they can cause allergenic reactions.

According to Heinzerling et al. (2009), average sensitization rates to *Corylus*, *Alnus*, and *Betula* in Europe are 22.8%, 21.2%, and 24.2%, respectively. In Poland, sensitization rates are comparable: 22.3% for *Corylus*, 22.8% for *Alnus*, and 27.7% for *Betula* (Heinzerling et al. 2009). Furthermore, the prevalence of pollen allergy is increasing worldwide (Heinrich et al. 2002).

Individual responses can vary considerably based on hereditary factors, exposure over time, and concentration of inhaled allergens. First symptoms in patients allergic to *Corylus*, *Alnus*, and *Betula* pollen were visible during exposure to concentrations greater than 35, 45, and 20 pollen/m^3 of air, respectively. Symptoms of allergic disease in all subjects were connected with concentrations greater than 80 pollen/m^3 of air for *Corylus*, 85 pollen/m^3 of air for *Alnus*, and 75 pollen/m^3 of air for *Betula* (Rapiejko et al. 2007). In addition, the major allergens in *Corylus*, *Alnus*, and *Betula* pollen are immunochemically similar (Ipsen et al. 1985); therefore, members of the Betulaceae Grey family can cause cross-reactions (Valenta et al. 1991).

The most commonly recognized allergic problem is a seasonal allergic rhinitis (hay fever). The symptoms of this condition include sneezing, nasal itching, rhinorrhoea (runny nose),

and nasal congestion (blocked nose). Itchy eyes, ears, mouth, swelling around the eyes, headache, and impaired smell can also occur. In addition, asthma is closely linked with pollen allergy.

Major socio-economic effects of allergy include absence from work (absenteeism) and reduced productivity at work (presenteeism). The annual cost of allergic diseases has been estimated to be between approximately 55 and 151 billion Euros in the European Union (Zuberbier et al. 2014).

1.5 Aerobiological monitoring

The origin of aerobiological measurements can be found in the nineteenth century. Around 1860, French chemist and microbiologist Louis Pasteur (1822–1895) built a series of glass bottles with long curved necks to examine the biological contents of dust in the air of Paris. Later, the French physician Pierre Miquel (1850–1922) designed various types of samplers to monitor pollen and fungal spores in the atmosphere. From 1866, English physician Charles H. Blackley, now known as the father of aerobiology, counted pollen grains using a sampler he built. He created pollen calendars for Manchester, England. In 1873, Blackley demonstrated a direct relationship between the quantity of pollen in air and the severity of allergic symptoms. In the second half of the nineteenth century, many scientists built their own devices: Maddox (1870), Cunningham (1873), Miquel (1878), France (1884), and Roster (1885) (Scheifinger et al. 2013).

In 1946, Oren C. Durham (1889–1967) introduced his gravimetric sampler. This device was used first in the monitoring of *Ambrosia* pollen. However, in a short time the measurements were extended to all types of pollen, and the sampler was applied in pollen monitoring in the USA, Canada, Mexico, and Cuba (Scheifinger et al. 2013). Nowadays, this device is still in use in developing countries (Singh & Mathur 2012). In the 1950s, Jim Hirst designed a volumetric spore trap. Modifications of the Hirst trap (Burkhard and Lanzoni traps) are still in operation today (Berger et al. 2014).

In Poland, the first aerobiological measurements were performed by Mieczysław Obtułowicz, Anna Weiss et al. in Kraków and Rabka in the 1960s (Śpiewak 2003). In the next two decades, aerobiological monitoring was initiated in Warszawa by Mieczysław Dąbrowski and Edward Zawisza, in Bydgoszcz by Ryszard Gniazdowski and Franciszek Klimas, and in Łódź by Małgorzata Wnuk and Krzysztof Buczyłko (Rapiejko 1995). All of these measurements were performed using gravimetric methods. In the 1980s, volumetric methods were gradually introduced (Śpiewak 2003). In 1982, Krystyna Obtułowicz and Kazimierz Szczepanek started continuous measurements of pollen concentration in the air in Kraków. They were the first in Poland to begin collaboration with the European Aeroallergen Network (Obtułowicz et al. 1990). At the end of the 1980s, Krzysztof Buczyłko, Edward Zawisza, and Piotr Rapiejko formed the Allergen Research Center in Łódź (Ośrodek Badania Alergenów Środowiskowych). The aim of the Allergen Research Center was to expand the network of measurement of pollen and fungal spore concentration (Śpiewak 2003; Rapiejko 1995). In the early 1990s, Małgorzata Latałowa began aerobiological studies in Gdańsk, while Alicja Stach began studies in Poznań. In the following years, the aerobiological monitoring network expanded with additional sites in Lublin (1995), Rzeszów (1996), Sosnowiec (1996), and Wrocław (2002) (http://www.aero.cm-uj.krakow.pl/historia.html).

All of the monitoring stations involved in the European Aeroallergen Network use the Hirst-type volumetric spore trap. However, a recent study of Galán et al. (2014) showed notable variations in sample preparation and analysis. The main differences included adhesives, mounting media, counting methods, number of horizontal sweeps, pollen types identified per year, and approach to internal quality control testing. A set of minimum requirements was produced to improve the quality of aerobiological monitoring (Galán et al. 2014). However, the value obtained for gathered pollen concentration in the air is an estimation of the whole biological population (Comtois et al. 1999); therefore, the pollen count obtained is prone to error.

Comtois et al. (1999) distinguished two possible sources of error in aerobiology: systematic errors (bias) and random errors. Bias is not determined by chance and it is usually unidirectional (systematic) (Comtois et al. 1999; Oteros et al. 2013). Bias is often con-

nected with the counting method used or other methodological issues, such as the adhesive used for the trapping surface (Comtois & Mandrioli 1997; Galán & Dominguez-Vilches 1997; Alcazar et al. 1999; Gottardini et al. 2009). The pollen traps are situated on the roofs of buildings to decrease bias from local sources (Mullins & Emberlin 1997). However, systematic error is still possible due to differences in the relative position of samplers and in the heights of buildings. Random errors are the representation of uncertainties. These can be divided into human errors and statistical errors. Human errors, often described as technician errors, are made by analysts in counting or identifying pollen grains (Oteros et al. 2013). Statistical errors account for the largest proportion of non-biased errors in aerobiological monitoring (Pedersen & Moseholm 1993; Comtois et al. 1999). These arise from the sampling methods used to estimate real airborne particles. Only a small proportion of a slide is read, and therefore the value obtained is an approximation of the true value. Comtois et al. (1999) discovered a mean error of 23% between the count using standard protocols and the total population count. Their results also showed a significant correlation between error and the abundance of pollen taxa on a sampled slide. Thus, the error increases when the number of pollen grains decreases.

Currently, pollen analysis requires a labor-intensive approach. It requires a qualified observer to identify and count pollen grains using a light microscope. Afterwards, this data has to be entered into a computer for processing and further analysis. These tasks are highly demanding in terms of time and cost. As a consequence, there is a need for the development of automatic systems for identification and counting airborne pollen (Boucher et al. 2002). Recently, Holt & Bennett (2014) reviewed methods for automatic pollen counting and recognition. They showed that simple, time-consuming tasks can already be automatized, while more complex problems will probably be addressed within the next few decades. The possible benefits of automatic systems include (i) an increased number of counted pollen grains, (ii) faster and more efficient data generation, and (iii) improved objectivity and consistency of data (Stillman & Flenley 1996; Holt & Bennett 2014).

1.6 Modeling and forecasting of *Corylus*, *Alnus*, and *Betula* pollen concentrations in the air: State of the art

1.6.1 SPATIAL ANALYSIS

Mandrioli & Ariatti (2001) identified aerobiology as a discipline as well as a tool for other disciplines. Aerobiology studies are interesting for scientists from numerous disciplines, ranging from those of medicine, cultural heritage, indoor air quality, biology, and agriculture to those of meteorology, climatology, and geography.

Spatial analyses of *Corylus*, *Alnus*, and *Betula* pollen are primarily connected with three types of studies:

1. **The comparison of pollen concentration or pollen season properties from two or more locations.** Stach et al. (2008) analyzed factors that influence the severity of *Betula* pollen seasons in Poland and the United Kingdom. They suggested that a severity of *Betula* pollen seasons at Poznań, Worcester, and London is connected to different phases of the North Atlantic Oscillation. Puc & Kasprzyk (2013) studied the properties of *Corylus* and *Alnus* pollen seasons in Szczecin and Rzeszów. Their study showed the differences in the characteristics of pollination in the two cities. In addition, the effect of long-distance transport of *Corylus* and *Alnus* pollen was observed in Rzeszów. Sauliene et al. (2014) compared pollen seasons of *Corylus*, *Alnus*, and *Betula* in Riga, Moscow, and Vilnius. Their study showed significant differences between *Corylus* and *Alnus* pollen seasons in Riga and Moscow. The average pollen concentration of *Betula* was twice higher in Moscow than in Riga and Vilnius. Moreover, Moscow had the shortest period of pollen season for all of the analyzed taxa.

2. **The characterization of the temporal and spatial variability of pollen concentrations.** Emberlin et al. (2002) investigated patterns in the start of *Betula* pollen seasons across Europe. Their analyses showed differences between regions.

Most of the sites exhibited a trend towards earlier start dates of *Betula* pollen seasons. However, Turku showed cyclic patterns in start dates, and the trend in Kevo was towards later starts of *Betula* pollen seasons. Rieux et al. (2008) characterized spatial variation of *Betula* pollen concentration properties over southeast France using regression and cluster analysis. They identified distance between sites, latitude, and altitude as significant predictors of airborne pollen variations. They also determined that the clustering of sites with similar pollen concentration properties can be connected with biogeographical or climatic factors. Myszkowska et al. (2010) analyzed the dynamics of spatial variations of *Corylus* and *Alnus* pollen seasons in Poland. This study showed that geographical position influences the properties of pollen seasons and the annual totals of pollen grains of *Corylus* and *Alnus*. In addition, different spatial patterns between annual totals of *Corylus* and *Alnus* pollen grains were observed.

3. **The application of backwards trajectory analysis to establish the source regions affecting the specific location (monitoring site).** Skjøth et al. (2008) determined the sources of *Betula* pollen in Copenhagen and Roskilde. Their study showed that not only forests but also urban areas are a significant source of *Betula* pollen. They also found one long-range transport episode from a remote source in Poland and Germany. Veriankaite et al. (2009) analyzed regional and long-range transport of *Betula* pollen in Lithuania using two variants of the dispersion model SILAM (System for Integrated modeLling of Atmospheric coMposition) and the trajectory model HYSPLIT. In years 2004–2007, they found 24 cases of high pollen concentration before the local pollen season. Their results identified the long-range sources of pollen grains. However, the spatiotemporal patterns of long-range transport were highly variable. Skjøth et al. (2009) determined possible sources of *Betula* pollen for north London. Back-trajectory analysis showed that *Betula* pollen grains in north London, aside from coming from trees within London, could originate from Continental Europe and southern England.

1.6.2 MODELS

Two main groups of models, statistical and numerical, are used in studies of *Corylus*, *Alnus*, and *Betula* pollen concentration in the air. Statistical (observation-oriented) models focus on the relationship between a dependent variable (pollen data) and one or more independent variables. These models do not require knowledge of emission or diffusion of pollen. Statistical modeling of pollen data has two main purposes: (i) to identify and describe the numerical relationship between pollen concentration and independent variables such as meteorological conditions, and (ii) to correctly predict the pollen concentration levels or values, or the characteristics of pollen seasons. On the other hand, numerical models use the mathematical equations and algorithms of atmospheric dispersion to estimate pollen concentration, using data from known sources. The main goals of numerical models are (i) estimation of pollen concentration, and (ii) determination of pollen source regions affecting a certain region (Sofiev et al. 2013).

Statistical models have varied in terms of modeling techniques, predictors, and output variables. The following statistical techniques have been used: multiple regression (Emberlin et al. 1993; Laaidi 2001); stepwise multiple regression (Bringfelt et al. 1982; Myszkowska 2013); additive logistic models and partially linear models (Cotos-Yáñez et al. 2004); artificial neural networks (Castellano-Méndez et al. 2005; Puc 2012); ARIMA models (Rodriguez-Rajo et al. 2006); and stochastic gradient boosting (Hilaire et al. 2012).

In recent years two numerical models of *Betula* pollen concentration have been created (Vogel et al. 2008; Sofiev et al. 2013). Vogel et al. (2008) parameterized *Betula* pollen dispersion into the model system COSMO-ART. Detailed spatial inventory data of the *Betula* stands in Switzerland was derived and used to provide information about the distribution of pollen sources. As a result, the diurnal variation of the pollen concentration for four days was simulated and compared with pollen concentrations measured by six pollen monitoring stations. Sofiev et al. (2013) developed a *Betula* pollen emission model as part of SILAM (System for Integrated modeLling of Atmospheric coMposition). Dispersion and flowering models, as well as phenological and aerobiological data, were used

Table 1: Summary of studies on modeling and forecasting of *Corylus*, *Alnus*, and *Betula* pollen concentrations in the air

Authors	Taxon	Output	Inputs	The most important variables	Methods	Validation	Location
Bringfelt et al. (1982)	*Betula*	Daily pollen concentration	Meteorological variables (temperature, global radiation, duration of sunshine, relative humidity, air pressure change, precipitation, wind direction and speed)	Temperature, relative humidity	Empirical (multiple regression)	No	Stockholm (Sweden)
Emberlin et al. (1993)	*Betula*	Pollen season properties (start, duration, annual totals)	Meteorological variables (relative humidity, sunshine hours, wind direction and velocity, air temperature, precipitation, and synoptic situation)	Temperature, precipitation	Empirical (multiple regression)	Partialy (two years)[a]	London (England)
Laaidi (2001)	*Betula*	Pollen season properties (start and duration)	Meteorological variables (precipitation, temperature)	Temperature	Empirical (multiple regression)	Yes (one year)	Burgundy (France)
Cotos-Yáñez et al. (2004)	*Betula*	Daily pollen concentration levels	Meteorological variables (precipitation, humidity, temperature, hours of sunshine, wind speed and direction) and the number of pollen grains	No information	Empirical (generalized additive model, partially linear model)	Yes (one year)	Vigo (Spain)
Castellano-Méndez et al. (2005)	*Betula*	Daily pollen concentration levels	Meteorological variables (precipitation, temperature) and the number of pollen grains	No information	Empirical (artificial neural networks)	Yes (two years)	Santiago de Compostela (Spain)
Rodriguez-Rajo et al. (2006)	*Alnus*	Daily pollen concentration	Meteorological variables (precipitation, relative humidity, temperature, hours of sunshine, wind speed and direction) and the number of pollen grains	Temperature, sun hours, precipitation	Empirical (ARIMA)	Yes (one year)	NW Spain
Vogel et al. (2008)	*Betula*	Pollen concentration	Meteorological variables (temperature, humidity, wind speed) and the distribution of *Betula* stands	-	Numerical (COSMO-ART)	Yes (four days)	Switzerland
Hilaire et al. (2012)	*Alnus* and *Betula*	Daily pollen concentration	Meteorological variables (temperature, humidity, evaporation, precipitation, radiation, wind speed and direction) and the number of pollen grains	Temperature, humidity	Empirical (stochastic gradient boosting)	Yes (last 25% of the data, at least 3 years)	Switzerland
Puc (2012)	*Betula*	Daily pollen concentration	Meteorological variables (wind speed, precipitation, relative humidity, temperature) and the number of pollen grains	Temperature and humidity	Empirical (artificial neural networks)	Yes (15 % of the cases, randomly)	Szczecin (Poland)
Myszkowska (2013)	*Betula*	Pollen season properties (start, end, peak value, peak day)	Meteorological variables (temperature, precipitation, relative humidity, cloudiness, relative sunshine)	Temperature	Empirical (multiple regression)	Yes (two years)	Kraków (Poland)
Sofiev, Siljamo, et al. (2013), Siljamo et al. (2013)	*Betula*	Pollen concentration	Meteorological variables (wind speed, precipitation, relative humidity, temperature), dates of phenological phases, and the distribution of *Betula* stands	Temperature	Numerical (SILAM)	Yes (one year)	Europe

[a]Only validation of the pollen season start model was performed

27

as input information. The performance of the *Betula* pollen emission model using SILAM was evaluated by Siljamo et al. (2013).

Overall, models have tested the possibility of predicting pollen season characteristics (Emberlin et al. 1993; Laaidi 2001; Myszkowska 2013); pollen concentration values (Bringfelt et al. 1982; Rodriguez-Rajo et al. 2006; Vogel et al. 2008; Hilaire et al. 2012; Puc 2012; Sofiev et al. 2013); and pollen concentration levels (Cotos-Yáñez et al. 2004; Castellano-Méndez et al. 2005) (Table 1). Different variables have been used in pollen models; however, the majority of the studies used meteorological variables as predictors. The number of pollen grains at an analyzed monitoring site was also often used. Temperature was determined to be the most important independent variable. In addition, relative humidity and precipitation were often described as relevant (Table 1).

Different approaches to the validation of the models were taken in past studies. The majority of studies used the aerobiological data from one or two years as a validation set (Table 1). Hilaire et al. (2012) used the most recent 25% of the data; Puc (2012) validated model results on 15% of random days; and Vogel et al. (2008) simulated the diurnal variation of *Betula* pollen concentrations for four consecutive days. Although Emberlin et al. (1993) analyzed pollen season start, duration, and severity, they validated only the model of the pollen season start. No validation was performed in the work of Bringfelt et al. (1982).

The results of Bringfelt et al. (1982) gave multiple correlation coefficient values between *Betula* pollen concentration and meteorological values up to 0.81. Emberlin et al. (1993) found multiple correlation coefficient values to be up to 0.62, 0.31, and 0.42 for *Betula* pollen season start, duration, and severity, respectively. Laaidi (2001) predicted *Betula* pollen season start seven to nine days before the actual dates of the validation set. He also forecasted a pollen season duration of 15 days, while it was 14 days. The study of Cotos-Yáñez et al. (2004) in Vigo gave almost 90% probability of detecting high *Betula* pollen concentration levels. The models of Castellano-Méndez et al. (2005) predicted between 83 and 100% of *Betula* pollen concentrations exceeding 20, 30, 70, and 80 grains/m^3

thresholds. Rodriguez-Rajo et al. (2006) note that "the estimated curve[s] accurately describe the *Alnus* pollen grains' behaviour", without giving any information on the accuracy. Vogel et al. (2008) parameterized the pollen grain emission into the model system COSMO-ART and compared the results with four consecutive days. They surmised that "the level of daily concentration is well modelled". Hilaire et al. (2012) tested *Alnus* and *Betula* models using the data from 14 pollen measurement stations. Their *Betula* model gave an accuracy up to approximately 0.87 (Locarno), and their *Alnus* model accuracy was up to 0.78 (Geneva). However, information about the lowest values of the performance statistic was missing. The prediction accuracy of Puc (2012) model was 0.86 on a test set. Myszkowska (2013) used multiple regression to predict pollen season properties. The results gave multiple correlation coefficients of 0.41–0.62 for pollen season start, 0.46–0.65 for pollen season end, 0.28–0.49 for peak concentration, and 0.33–0.52 for peak day. Siljamo et al. (2013) evaluated the model developed by Sofiev et al. (2013). The SILAM pollen predictions were correct in 65% of cases of low pollen concentration and in 56% of cases of high pollen concentration. The verifiability in cases of moderate values was worse, with only 33% of correctly predicted cases for moderate pollen concentration level.

Lastly, prediction models vary in terms of spatial extent. Studies have focused on modeling pollen concentration values in one city (Bringfelt et al. 1982; Emberlin et al. 1993; Cotos-Yáñez et al. 2004; Castellano-Méndez et al. 2005; Puc 2012; Myszkowska 2013); in a region (Laaidi 2001; Rodriguez-Rajo et al. 2006); in a country (Vogel et al. 2008; Hilaire et al. 2012); and even on a continental scale (Sofiev et al. 2013).

Nevertheless, there are still several problems with modeling and forecasting of *Corylus*, *Alnus*, and *Betula* pollen concentrations in the air. Firstly, the majority of previous studies have focused on a temporal prediction. Although some authors used the data from several monitoring sites, spatial aspects were omitted or marginalized. Thus, there is a need to analyze not only temporal but also spatiotemporal characteristics of pollen concentration. Numerical models require several pieces of information, such as detailed spatial distribution of pollen sources, reliable phenological data (for example, spatial distribution of the

beginning of flowering), and the parameterization of dispersion modules. Gridded inventories of pollen sources are rare, and their spatial resolution is coarse. There is also insufficient information about pollen sources in cities. Phenological data are often incomparable. The observed phases vary from country to country, and the quality of data is inconsistent due to different principles and procedures of observation. On the other hand, statistical models are based on a relationship between pollen data and independent variables. Many statistical methods make assumptions about the data, and therefore variables need to be transformed. Different transformation techniques were tested by Toro et al. (1998) for multiple regression models. However, these techniques should also be compared in machine learning models. The majority of studies using statistical models showed the large influence of air temperature, relative humidity, and precipitation on pollen concentration in the air. However, there are differences in the temporal range of the applied predictors. Hourly mean temperature, daily mean temperature, or even monthly mean temperature were used in previous studies. Both numerical and statistical models of *Corylus*, *Alnus*, and *Betula* pollen concentrations in the air have common problems with the validation of results. *Betula* and *Corylus* have a biannual cycle (Jäger et al. 1991; Latałowa et al. 2002; Rodriguez-Rajo et al. 2006); nevertheless, in some studies only one year of data was used for validation. In many other studies, two years of observation were used as a testing set. This could still give very unstable results in terms of accuracy, because of the occurrence of years with abnormal weather conditions (precipitation, temperature, etc.) influencing the pollen season properties and pollen concentration values. A longer, continuous time series of aerobiological observations should mitigate the difficulties of temporal validation. Pollen monitoring stations are not randomly located; they are usually in large, academic cities. Therefore, the reliable spatial validation of the models is difficult to achieve. Another problem with validation is the choice of a model's performance statistics. In cases of classification problems, such as prediction of low or high pollen concentration level, accuracy cannot be the best indicator of a model's performance. For example, if the testing set consists of 90 days with low pollen concentration levels and 10 days with high pollen concentration levels, a prediction that all days have a low level will give an accuracy of

90%. Moreover, the use of different performance statistics in various studies makes the models difficult to compare. Thus, there is a need to determine the optimal set of validation statistics for forecasting *Corylus*, *Alnus*, and *Betula* pollen concentrations. Finally, there is a lack of studies of *Corylus* and *Alnus* pollen concentration models. Although the number of *Corylus* and *Alnus* pollen grains in the air is usually lower than that of *Betula*, it is still sufficient to provoke allergic symptoms.

1.7   Objectives

The general objective of this study is to develop spatiotemporal predictive models of *Corylus*, *Alnus*, and *Betula* pollen concentration levels in the air in Poland.

In accordance with the general objective, the specific objectives are the following:

1. To determine the characteristics of temporal and space–time autocorrelation of *Corylus*, *Alnus*, and *Betula* pollen concentration in the air in Poland.
2. To create and evaluate *Corylus*, *Alnus*, and *Betula* pollen concentration levels predictions using past pollen count data from monitoring sites.
3. To develop a framework for the predictive mapping of pollen concentration levels.
4. To identify the optimal set of meteorological variables for predicting *Corylus*, *Alnus*, and *Betula* pollen concentration levels.
5. To demonstrate the use of preprocessed, gridded meteorological data for spatiotemporal predictive models of *Corylus*, *Alnus*, and *Betula* pollen concentration levels.

Chapter 2

Materials

2.1 Study area

Poland is a lowland country located in Central Europe. Its area is 322,575 sq km and the average height above sea level is 173 m (Dmochowska 2013). Approximately 46% of the country is higher than 150 m above sea level, and only about 3% of the country is higher than 500 m above sea level. Mount Rysy (2,499 m), located in the Tatra Range of the Carpathians, is the highest peak in Poland. Poland is divided into five topographic zones from north to south: (i) the Baltic coastal plains; (ii) the lake region consisting of the Pomeranian and Mazurian Lakelands, with the elevation up to 309 m above sea level; (iii) the Land of Great Valleys with the Wielkopolska, Mazovian, and Silesian lowlands, occupied by vast plains; (iv) the uplands; and (v) the mountains formed by the Sudeten and the Carpathian ranges. Agricultural land covers approximately 60% of the surface area of the country. Forests cover about 30% of the country's land area. Forest coverage exceeds 50% of the area in the western and northwestern parts of Poland. The forest area is also large in the Carpathians and the Masurian Lake District. The forest coverage is lowest in the central part of Poland. The *Alnus* share of forested areas is estimated to be about 5.3%, while the *Betula* share is estimated to be about 7.4% (Domaszewicz 2013). There is a lack of information about the share of *Corylus*. Built-up and urbanized areas

occupy approximately 5% of the total area of Poland (Dmochowska 2013).

The climate of Poland presents transient conditions influenced by the vast land areas to the east and by the Atlantic Ocean to the west. Thus, Poland's climate is often described as a hybrid oceanic/continental temperate climate. The main pressure systems that affect the weather in Poland are the Icelandic low and the Azores high. There are also seasonal influences of the Siberian High and Arctic High. In addition, the absence of meridional mountain ranges in Poland facilitates the penetration of oceanic air masses from the west and continental air masses from the east. Owing to its relatively small area, the Baltic Sea influences mainly the narrow coastal zone. In the rest of the country, the impact of the Baltic Sea on Poland's climate is limited (Woś 2010).

The spatial variations in the air temperature are primarily connected with the large-scale atmospheric circulation. In most of the country, the average annual temperature decreases from southwest to northeast. However, this pattern does not apply to mountainous areas. Elevation above sea level, temperature drop with altitude, and slope exposure are the main factors that influence the annual temperatures in these areas (Woś 2010). The lowest average annual temperatures are recorded in the Tatra and Karkonosze Mountains. Lower Silesia, the Wielkopolska Lowland, and the eastern part of the Pomeranian Lakeland are the warmest regions, with an average annual temperature up to approximately 8.5°C (Lorenc 2005).

The daily and annual average vapor pressure show a pattern similar to that of the air temperature. The highest values of vapor pressure are recorded in summer, while the lowest values are in winter. The annual average vapor pressure is between 8.4 and 9.9 hPa in most parts of the country. Only in the mountains are these values much lower, sometimes even lower than 6 hPa. Although the value of annual average vapor pressure is somewhat variable, the average vapor pressure can change substantially throughout the year (Woś 2010).

Poland has a predominantly western circulation and the majority of winds blow from the western sector (west, southwest, and northwest) (about 40–55%). Winds from the eastern

sector are reported in about 15–30% of observations. Southern and northern winds are the least common. In most of the country, the average wind speed ranges between 2 and 4 m/s. The highest values of average wind speed are reported in the mountains, with the annual average wind speed exceeding 11 m/s. The wind speed displays distinct seasonal variations. It is lowest in summer (usually in July) and highest in winter (usually in January) (Woś 2010). Moreover, a diurnal variation of wind speed is clearly seen in spring and summer season owing to the impact of atmospheric convection (Czernecki 2013).

Precipitation is mainly connected with the passage of meteorological fronts. In addition, the spatial distribution of precipitation is influenced by orographic features and exposure to the dominant westerlies. The multi-year annual precipitation totals range between 500 and 700 mm. Precipitation totals also vary substantially from year to year. The Wielkopolska and Mazovian Lowlands are the driest areas, with annual totals lower than 550 mm. The highest values of annual precipitation totals, exceeding 1250 mm, are reported in the Sudeten and the Carpathians (Woś 2010).

Total solar radiation in Poland varies temporally and spatially. It depends on the atmospheric circulation, and therefore it is controlled by a cloudiness and atmospheric turbidity factor. Average total solar radiation in Poland is 9.98 MJ/m²/day (Woś 2010). The lowest values of total radiation are measured in winter and the highest in summer. In addition, higher values of total radiation are reported in central and east Poland, while lower values are reported in the southeast, east, and northeast (Woś 2010).

The growing season in Poland is approximately 220 days long, on average. Except for the mountainous areas, the growing season lasts longest in southwestern Poland, with more than 225 days in a year, while the shortest is in the northeastern part of the country, with less than 205 days in a year. Moreover, in the mountains, the growing season is shorter than 110 days a year (Woś 2010).

Corylus, *Alnus*, and *Betula* pollen concentrations depend on local variations in the vegetative flora and the weather conditions in the year preceding the pollination period and

during the pollen season (Latałowa et al. 2002; Puc 2007). Air temperature is often described as the main factor that influences the start of the pollen season (Piotrowska & Kubik-Komar 2012). In addition, light intensity, evaporation, low humidity, and moderate wind speed have an impact on pollen release (Helbig et al. 2004; Pacini & Hesse 2004; Dahl et al. 2013). The dispersal of pollen grains is influenced by gravity, wind speed and direction, and turbulence (Di-Giovanni & Kevan 1991). Afterwards, pollen grains are transferred to solid surfaces either by dry deposition or wet deposition (caused mainly by precipitation) (Sofiev et al. 2013). Sporadically, the grains of pollen can be resuspended into the air by gusty winds over relatively dry terrain (Sofiev et al. 2013).

The start and length of *Corylus* and *Alnus* pollen seasons are very variable from year to year. In Poland, the *Corylus* season usually starts between early February and late March and lasts on average for seven weeks. However, the *Corylus* pollen season can last between about three and nine weeks. The time range of the *Alnus* pollen season is similar, with a length between one and ten weeks and an average of about seven weeks (Puc & Kasprzyk 2013). In Kraków, the *Betula* pollen season occurs between the first half of April and mid-May (Myszkowska & Piotrowicz 2009). Its pollination period is relatively short and less variable than that of *Corylus* or *Alnus*. *Betula* pollination usually lasts three weeks on average (Myszkowska & Piotrowicz 2009).

2.2 Aerobiological data

The main data source was the daily pollen count of *Corylus*, *Alnus*, and *Betula* from 11 cities in Poland: Bydgoszcz, Gdańsk, Kraków, Lublin, Łódź, Olsztyn, Poznań, Rzeszów, Siedlce, Sosnowiec, and Szczecin. The aerobiological studies covered between two and 16 years of measurements (Figs. 2, 3, 4).

Daily average pollen concentrations of *Corylus*, *Alnus*, and *Betula* were measured using the recommendations of the European Aerobiology Society's Quality Control Working Group (Galán et al. 2014). Pollen counts were determined by a volumetric spore trap of the

Figure 2: The location of the aerobiological monitoring stations used for the study

Hirst design (Hirst 1952), located 12 m above ground level or higher. The trap sucks air through the narrow inlet opening at a rate of 10 l/min. Airborne particles are deposited on an adhesive tape wrapped around a moving drum (Leventin et al. 2000). The drum is slowly turned by clockwork at a constant speed of 2 mm/h, performing a complete turn during one week. Two different pollen counting methods were used. Pollen grains were counted along 12 vertical transects (Stach 2000), or along 4 horizontal transects (Galán et al. 2007). Both methods follow similar trends and provide close approximations to the pollen count from the entire slide (Cariñanos & Emberlin 2000). Their sum was multiplied by a factor dependent on the surface of the slide. The results were expressed as a daily average number of pollen grains in m^3 of air sampled per 24 h (Comtois 1998).
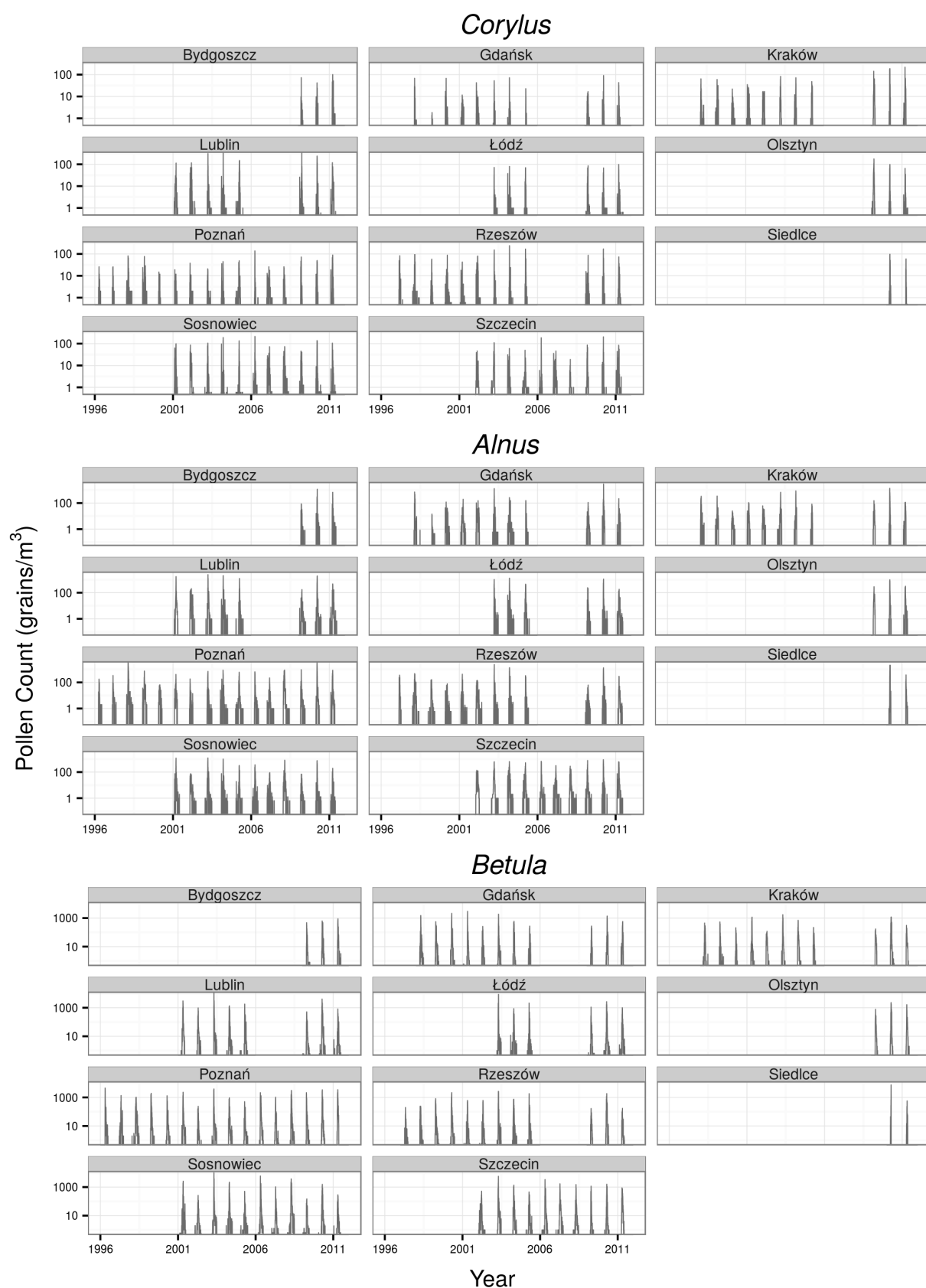
Figure 3: Pollen counts of *Corylus*, *Alnus*, and *Betula* for all of the analyzed years in each location on a logarithmic scale
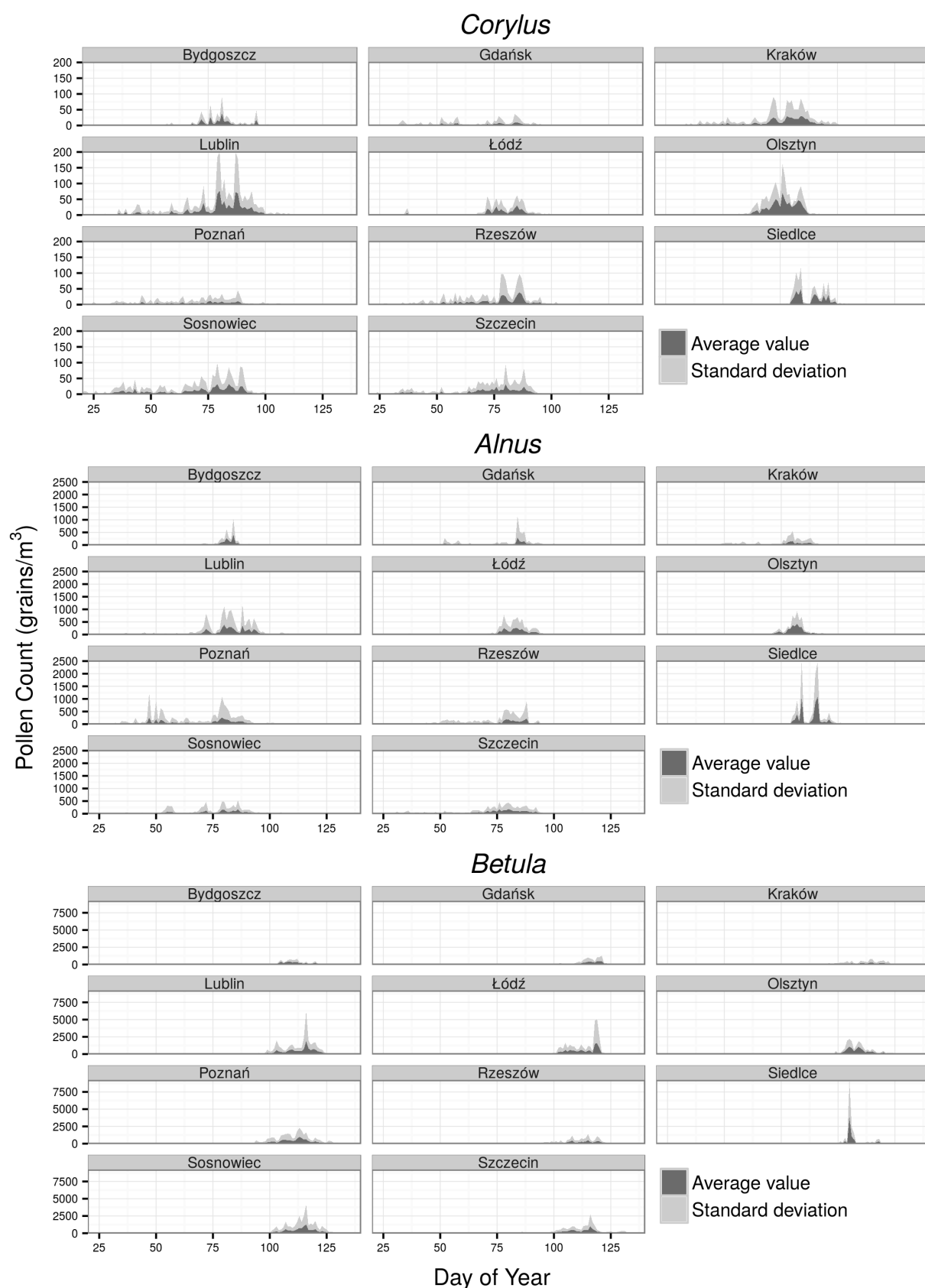
Figure 4: Average value of pollen counts of *Corylus*, *Alnus*, and *Betula* for all of the analyzed years in each location. One standard deviation of pollen count values is represented by the light grey area

2.3 Grid data

AGRI4CAST Interpolated Meteorological Data, created by the European Commission's Joint Research Centre (http://ec.europa.eu/jrc/), were used as an additional data source (Baruth et al. 2007). This database contains meteorological parameters such as average daily temperature, maximum daily temperature, minimum daily temperature, mean daily vapor pressure, mean daily wind speed at 10 m, mean daily rainfall, Penman potential evaporation, and daily total global radiation (Fig. 5). Most of the variables were interpolated from weather stations on a 25x25 km grid. Only Penman potential evaporation and daily total global radiation were calculated using available data. AGRI4CAST Interpolated Meteorological Data contain indicators from 1975 to 2014. This covers the European Union member states, neighboring European countries, and all countries bordering on the Mediterranean.
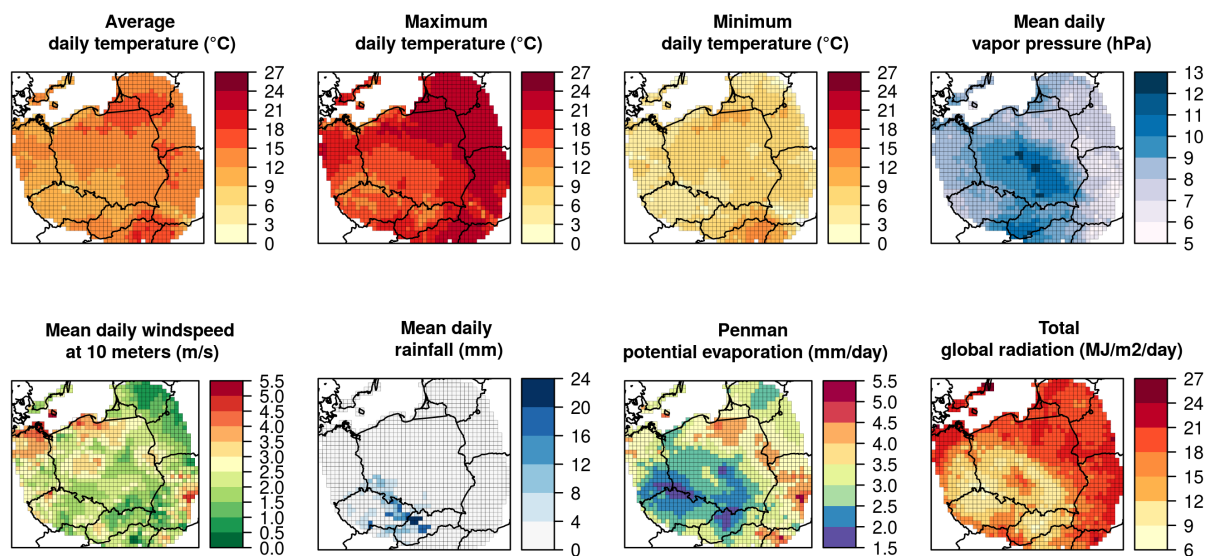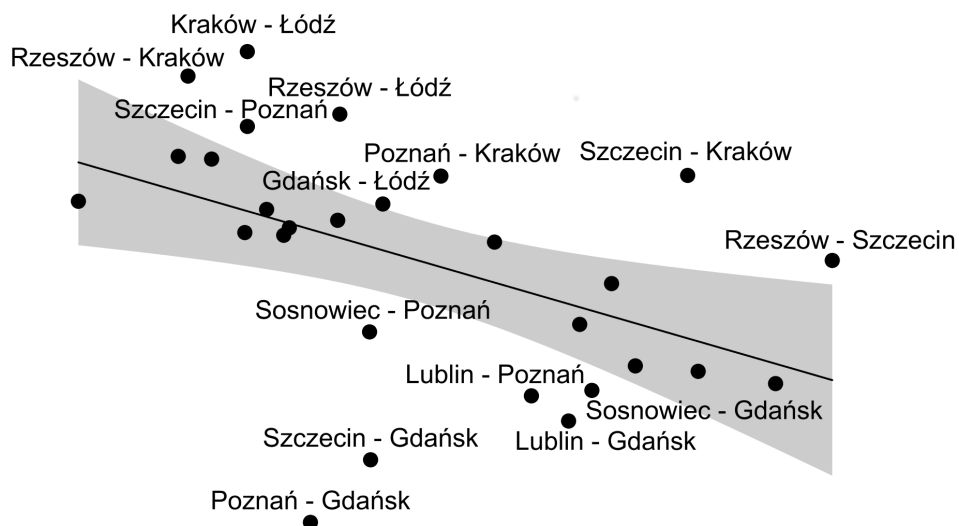
Figure 5: Examples of AGRI4CAST Interpolated Meteorological Data on 25 April 2011, based on the data for Poland and the area within 200 km of the Polish border

Chapter 3

Temporal and spatiotemporal autocorrelation of daily concentrations of *Alnus*, *Betula*, and *Corylus* pollen in Poland

CrossMark

ORIGINAL PAPER

Temporal and spatiotemporal autocorrelation of daily concentrations of *Alnus, Betula*, and *Corylus* pollen in Poland

J. Nowosad · A. Stach · I. Kasprzyk · Ł. Grewling · M. Latałowa ·
M. Puc · D. Myszkowska · E. Weryszko- Chmielewska · K. Piotrowska-Weryszko ·
K. Chłopek · B. Majkowska-Wojciechowska · A. Uruska

Abstract The aim of the study was to determine the characteristics of temporal and space–time autocorrelation of pollen counts of *Alnus*, *Betula*, and *Corylus* in the air of eight cities in Poland. Daily average pollen concentrations were monitored over 8 years (2001–2005 and 2009–2011) using Hirst-designed volumetric spore traps. The spatial and temporal coherence of data was investigated using the autocorrelation and cross-correlation functions. The calculation and mathematical modelling of 61 correlograms were performed for up to 25 days back. The study revealed an association between temporal variations in *Alnus*, *Betula*, and *Corylus* pollen counts in Poland and three main groups of factors such as: (1) air mass exchange after the passage of a single weather front (30–40 % of pollen count variation); (2) long-lasting factors (50–60 %); and (3) random factors, including diurnal variations and measurements errors (10 %). These results can help to improve the quality of forecasting models.

J. Nowosad (✉) · A. Stach
Institute of Geoecology and Geoinformation, Adam
Mickiewicz University, Dzięgielowa 27, 61-680 Poznań,
Poland
e-mail: nowosad@amu.edu.pl

I. Kasprzyk
Department of Environmental Biology, University of
Rzeszów, Zelwerowicza 4, 35-601 Rzeszów, Poland

Ł. Grewling
Laboratory of Aeropalynology, Faculty of Biology, Adam
Mickiewicz University, Umultowska 89, 61-614 Poznań,
Poland

M. Latałowa · A. Uruska
Department of Plant Ecology, University of Gdańsk, Wita
Stwosza 59, 80-308 Gdańsk, Poland

M. Puc
Department of Botany and Nature Conservation,
University of Szczecin, Felczaka 3c, 71-412 Szczecin,
Poland

D. Myszkowska
Department of Clinical and Environmental Allergology,
Jagiellonian University Medical College, Śniadeckich 10,
31-531 Kraków, Poland

E. Weryszko- Chmielewska
Department of Botany, University of Life Sciences in
Lublin, Akademicka 15, 20-950 Lublin, Poland

K. Piotrowska-Weryszko
Department of General Ecology, University of Life
Sciences in Lublin, Akademicka 15, 20-950 Lublin,
Poland

K. Chłopek
Faculty of Earth Sciences, University of Silesia,
Będzińska 60, 41-200 Sosnowiec, Poland

B. Majkowska-Wojciechowska
Department of Immunology, Rheumatology and Allergy,
Faculty of Medicine, Medical University, Pomorska 251,
92-215 Łódź, Poland

Springer

1 Introduction

Measured at a point, an air pollen count is the result of many factors affecting the production and dispersion of pollen in the atmosphere. Hence, changing one or a few factors usually does not cause an immediate and abrupt change in pollen concentrations. Rather, the change is gradual and somewhat delayed. The same holds for its spatial dimension. Understanding the character of atmospheric pollen concentrations can greatly help to improve the quality of forecasting models. Therefore, it is important in the prophylaxis of pollen allergies.

Alder (*Alnus* Mill.), birch (*Betula* L.), and hazel (*Corylus* L.) belong to the order Fagales Engl. and the family Betulaceae Gray (Bremer et al. 2009). Those trees are important sources of allergenic pollen in the temperate climatic zone of the Northern Hemisphere (Stach et al. 2008). The major pollen allergens from members of the family Betulaceae are structurally and immunochemically similar. Therefore, alder, birch, and hazel allergens have a high degree of cross-reactivity (Mothes and Valenta 2004). Seasons with abundant hazel or alder pollen can lead to stronger reactions to birch pollen and extend the birch pollen season (Emberlin et al. 1997; Puc 2003; Rodriguez-Rajo et al. 2004). Approximately 15 % of the European population suffers from allergies, and Poland is one of the countries with the highest allergy incidence rates, up to 45 %. It mostly affects children and young people (Heinzerling et al. 2009; Piotrowska and Kubik-Komar 2012; Malkiewicz et al. 2013). Sensitisation rates to trees belonging to the family Betulaceae are high in Central/Western Europe, with Poland showing high sensitisation rates for alder (22.8 %), birch (27.7 %), and hazel (22.3 %) (Heinzerling et al. 2009). The number of pollen grains needed to provoke an allergic reaction in people depends on individual reactivity and differs according to species and region. In Poland, the first allergy symptoms are observed at 20 pollen/m^3 of air for birch, 35 pollen/m^3 of air for hazel, and 45 pollen/m^3 of air for alder (Rapiejko et al. 2007).

The pollen grain identification is conducted at the genus level. Each genus has a dominant species in Poland: *Alnus glutinosa*, *Betula pendula*, and less common *Betula pubescens*, *Corylus avellana*, and their cultivars. All species within a particular genus bloom at very similar times. Alder and hazel pollen is the first to appear in the air. The pollination period and the start of the pollen season are highly variable from year to year and depend on the rather unstable weather conditions in late winter and spring (Smith et al. 2007; Kaszewski et al. 2008; Hájková et al. 2009; Rodriguez-Rajo et al. 2009; Myszkowska et al. 2011). Alder and hazel shed their pollen from the end of January to mid-April, while the birch pollination period is relatively short and occurs in the second half of April and early May (Weryszko-Chmielewska et al. 2001; Kluza-Wieloch and Szewczak 2006; Puc 2007; Grewling et al. 2012a).

Pollination is affected by the geographical location and local climate. The phases and duration of pollen seasons as well as the skewness, kurtosis, and annual totals of pollen grains heavily depend on the geographical position (Myszkowska et al. 2010). Nevertheless, annual totals of alder and hazel pollen show different spatial patterns. For *Alnus*, the annual pollen total increases from east to west, and for *Corylus*, there is an increase in the northerly direction (Myszkowska et al. 2010). Furthermore, weather conditions such as mean air temperature, total precipitation, relative humidity, and wind direction and speed have an important effect on the duration and intensity of pollen release and concentration in the air. As a result, these variables can be used for constructing forecast models (Rodriguez-Rajo et al. 2009; Puc 2012).

Spatial and space–time autocorrelation is functionally important in many ecosystems. Pollen production varies in both space and time; therefore, documenting its patterns and understanding the causes of variations is important to aerobiology. Moreover, describing temporal autocorrelation (one taxon—one location) and space–time autocorrelation (one taxon—two/many locations) should be the first step in developing models of pollen concentrations. Nowadays, autocorrelation analysis is rarely used in aerobiological research (Rodriguez-Rajo et al. 2006), except for mast seeding studies. Liebhold et al. (2003) quantified within-population spatial synchrony in mast dynamics of North American oaks, Garrison et al. (2006) studied spatial synchrony and temporal patterns of acorn production in California black oaks, and Ranta and Satri (2007) analysed the

43

synchronisation of high and low pollen years of *Betula*, *Alnus*, *Corylus*, *Salix*, and *Populus*. There has been no comprehensive analysis of alder, birch, and hazel temporal and space–time synchrony so far.

The main objective of the present study was to determine mean multi-year characteristics of temporal and space–time autocorrelation of the pollen counts of *Alnus*, *Betula*, and *Corylus* in the air of eight cities in Poland.

2 Materials and methods

2.1 Study area

The monitoring of the concentrations of *Alnus*, *Betula*, and *Corylus* pollen in the air was conducted in Gdańsk,

Fig. 1 Sites used for the study of temporal and spatiotemporal autocorrelation of daily pollen concentrations in Poland

Kraków, Lublin, Łódź, Poznań, Rzeszów, Sosnowiec, and Szczecin (Fig. 1; Table 1). The studies covered 8 years of measurement (2001–2005 and 2009–2011) in Gdańsk, Kraków, Lublin, Poznań, Rzeszów, and Sosnowiec, 7 years in Szczecin (2002–2005 and 2009–2011), and 6 years in Łódź (2003–2005 and 2009–2011). The minimum distance between the locations was 75 km (Kraków–Sosnowiec), and the maximum one 640 km (Rzeszów–Szczecin) (Fig. 1).

Poland is a country in Central Europe with an average elevation of 173 m and an area of 322,575 square kilometres (Dmochowska 2013). Only 3 % of the Polish territory is higher than 500 m. Despite its lowland character and a relatively small and compact area, Poland displays marked climatic differences. There is a NW–SE oceanic–continental gradient connected with the predominantly westerly circulation from the north Atlantic Ocean. In the analysed data set, this gradient is best represented by the locations of Szczecin and Rzeszów. This pattern is locally modified by the impact of the Baltic Sea and relief. At the locations concerned, the mean annual temperature ranges from 7.4 to 8.8 °C, annual precipitation totals, from 500 to 750 mm, and the duration of the growing season, from 205 to 225 days (Lorenc 2005). The Kolmogorov–Smirnov two-sample test shows no significant difference ($D = 0.12$, p value $= 0.14$) between the daily temperature for 8 years of study (2001–2005 and 2009–2011) and a 30-year time series of measurement (1983–2012).

Poland has an average urbanisation level, with an urban population of about 62 % and artificial surfaces covering 4 % of the land (Dmochowska 2013). Various kinds of agricultural land cover 62.5 % of

Table 1 Characteristics of the study sites: latitude, longitude, and altitude of the aerobiological monitoring sites, area and population of the cities, and mean temperatures recorded at meteorological stations located in them

City	λ (DD)	ϕ (DD)	Altitude (a.s.l.)	Area (km²)	Population (in thousands)	Mean temperature (°C) (1983–2012)				
						Annual	January	February	March	April
Gdańsk	18.6131	54.3856	10	262	460	7.42	−1.40	−1.40	1.78	6.66
Kraków	19.9559	50.0637	212	327	758	8.47	−1.90	−1.11	3.12	8.94
Lublin	22.5402	51.2437	198	147	348	7.81	−2.66	−2.22	1.93	8.31
Łódź	19.4748	51.7715	216	293	719	8.43	−1.62	−1.05	2.80	8.76
Poznań	16.9243	52.4671	91	262	551	8.96	−0.62	−0.18	3.59	9.08
Rzeszów	22.0160	50.0293	209	117	182	8.46	−2.10	−1.28	2.87	8.79
Sosnowiec	19.1389	50.2972	252	91	214	8.47	−1.50	−0.81	3.08	8.72
Szczecin	14.5478	53.4395	30	301	409	8.96	0.18	0.57	3.74	8.60

the country's area, including arable land, permanent crops, pastures, and heterogeneous farmland. Predominant crops include potatoes, sugar beets, wheat, rye, and barley (Dmochowska 2013). Forests, mostly coniferous, occupy about 30 % of the country's area. Their share is the largest in northern and western Poland (Milewski 2013).

2.2 Aerobiological data

Daily average pollen concentrations of three taxa such as *Alnus*, *Betula,* and *Corylus* were analysed over 8 years (2001–2005 and 2009–2011), depending on the availability of data. The investigations were carried out using Hirst-designed volumetric spore traps (Hirst 1952). Ten litres per minute was the intake of air through the narrow inlet opening. Airborne particles attached themselves to an adhesive tape wrapped around a moving drum. The drum moved at a constant speed of two millimetres per hour, performing a complete turn during 1 week. Every week, the tape was cut into seven segments and microscopic slides were made. Pollen grains were counted from twelve vertical or four horizontal strips, and their sum was multiplied by a factor dependent on the surface of a slide (Frenguelli 2003). The results were expressed as a daily average number of pollen grains in m^3 of air sampled per 24 h (Comtois 1998).

2.3 Statistical analysis

All calculations were carried out using the Variowin, ISATIS, and R software packages (Pannatier 1996; Bleines et al. 2011; R Core Team 2013). The limits of pollen seasons were calculated using the 90 % method (Nilsson and Persson 1981) whereby a season starts when 5 % of the total catch has been achieved and ends when 95 % has been reached.

Data of all years and particular site have been combined into a single table, in which X was the cumulative number of a day from 1 January 2001, and Y—the value of daily concentrations of taxon A in site B. In order to speed up calculations and eliminate the large number of lags with zero concentration values, the period from 1 to 180 days was taken into account each year.

Second-order properties, such as the mean, variance, and serial correlation, play a key role in the study of time-series data (Cowpertwait and Metcalfe 2010).

The number of time steps between the variables is known as the lag. If the correlation between variables depends only on the lag, the variables may be correlated and the model is second-order stationary. Autocorrelation, also known as serial correlation, is a correlation of a variable with itself at different times. Consider a time-series model that is second-order stationary. The autocovariance function (*acvf*), γ_k, can be defined as a function of the lag k:

$$\gamma_k = E[(x_t - \mu)(x_{t+k} - \mu)] \tag{1}$$

The function γ_k does not depend on t because the expectation is the same at all times t. The lag k autocorrelation function (*acf*), ρ_k, is

$$\rho_k = \frac{\gamma_k}{\sigma^2} \tag{2}$$

The estimate of autocovariance c_k is defined as:

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \tag{3}$$

where $\bar{x}$ is the sample mean of the time series.

The most satisfactory estimate of the kth lag autocorrelation ρ_k is

$$r_k = \frac{c_k}{c_0} \tag{4}$$

The r_k values can be called the sample autocorrelation function. All sample autocorrelations lie between -1 and 1. A plot of the correlation function against the lag is referred to as a correlogram (Fig. 2).

Consider a time-series model for variables x and y that are stationary in the mean and the variance. If the serial correlations of those variables and their correlations with each other at different time lags depend only on a lag, the combined model is second-order stationary. Therefore, a cross-covariance function (*ccvf*), $\gamma_k(x, y)$, can be defined as a function of the lag, k:
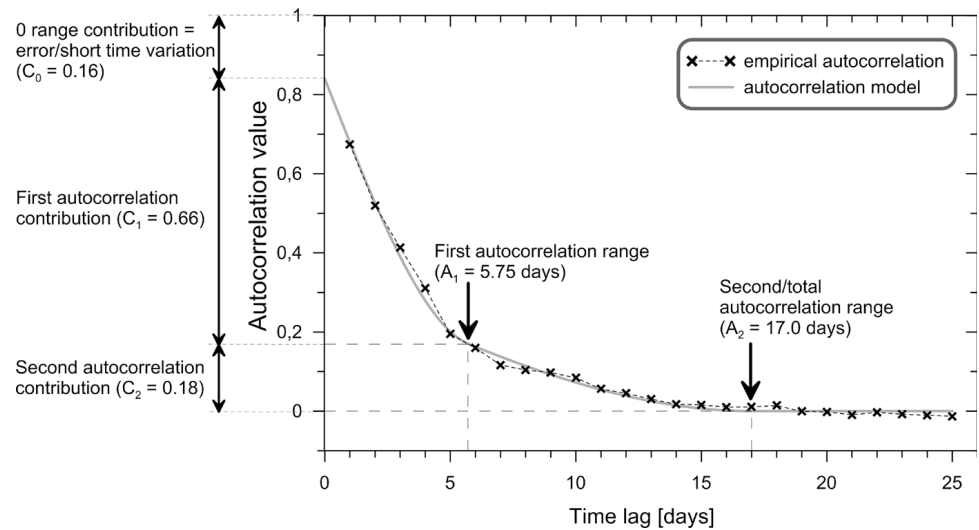
$$\gamma_k(x, y) = E[(x_{t+k} - \mu_x)(y_t - \mu_y)] \tag{5}$$

The lag k cross-correlation function (*ccf*), $\rho_k(x, y)$ is

$$\rho_k(x, y) = \frac{\gamma_k(x, y)}{\sigma_x \sigma_y} \tag{6}$$

The cross-correlation function, in contrast to the autocorrelation function, is not symmetric (Fig. 3). The estimate *ccvf*, $c_k(x, y)$ is defined as:

Fig. 2 Explanation of the terms of a correlogram and its mathematical model

The sample *ccf* is calculated as:

$$c_k(x,y) = \frac{1}{n} \sum_{t=1}^{n-k} (x_{t+k} - \bar{x})(y_t - \bar{y}) \qquad (7)$$

$$r_k(x,y) = \frac{c_k(x,y)}{\sqrt{c_0(x,x)c_0(y,y)}} \qquad (8)$$

The *ccf* function can be shown as a plot of $r_k(x,y)$ against k, which is called the cross-correlogram.

The zero structure (C_0) on the correlogram plot (Fig. 2) is represented by an extrapolated point of intersection between the correlogram model and the *y*-axis. It gives the estimated value of autocorrelation at the 'zero' interval. The results of two measurements made at the same place and at the same time should be identical, which means that the curve of the model should touch the *y*-axis where the correlation coefficient is equal to 1. However, usually this is not so: only 6 out of the 61 estimated *r* values for the 0-day interval equal 1 ($C_0 = 0.0$); see Table 2. The source of the differences is measurement errors and the temporal variability of pollen concentrations during the sampling period, i.e. within 1 day. Therefore, the C_0 structure in Fig. 2 and Tables 2 and 3 is referred to as 'error/short time variation'.

Typically, an empirical correlogram does not steadily decrease from the maximum to zero. Sometimes, it is composed of several segments with a different inclination. The connections of the segments show distinct changes in the shape of the curve. This characteristic of a temporal correlogram is interpreted as a result of the impact of factors/processes operating

at different scales (intervals), for example a single event, a diurnal cycle, and a seasonal cycle. Mathematical correlogram modelling allows an estimation of the length (time range) of its individual segments, and this provides a basis for reasoning about their origins. Theoretical basis of this methodology is presented in geostatistical textbooks, inter alia, Goovaerts (1997) and Chiles and Delfiner (2012).

The calculation of 61 correlograms (8 locations × 3 taxa × 2 or 3 variants with excluding extremes) and 84 cross-correlograms (28 pairs of locations × 3 taxa × 1 variant) as well as their mathematical modelling were performed for up to 25 days back. This process consisted of three stages. Sample correlograms and cross-correlograms were calculated at different lags. Next, parametric functions were identified using one of the conditionally negative definite functions, such as the Gaussian, the exponential and the spherical model, or by nesting these models. Finally, sample correlograms were fitted to the parametric form by adjusting model parameters: the sill, nugget, and range. The best fit was obtained by minimising the indicative goodness-of-fit (IGF) criterion (Pannatier 1996).

The mathematical modelling of the correlograms was intended to smooth random fluctuations of the empirical data, to interpolate and extrapolate the data, and to objectively assess the range and share of individual structures.

Since all the measurement series exhibited highly skewed distributions with a predominance of low and average values and single, far outlying extreme values,
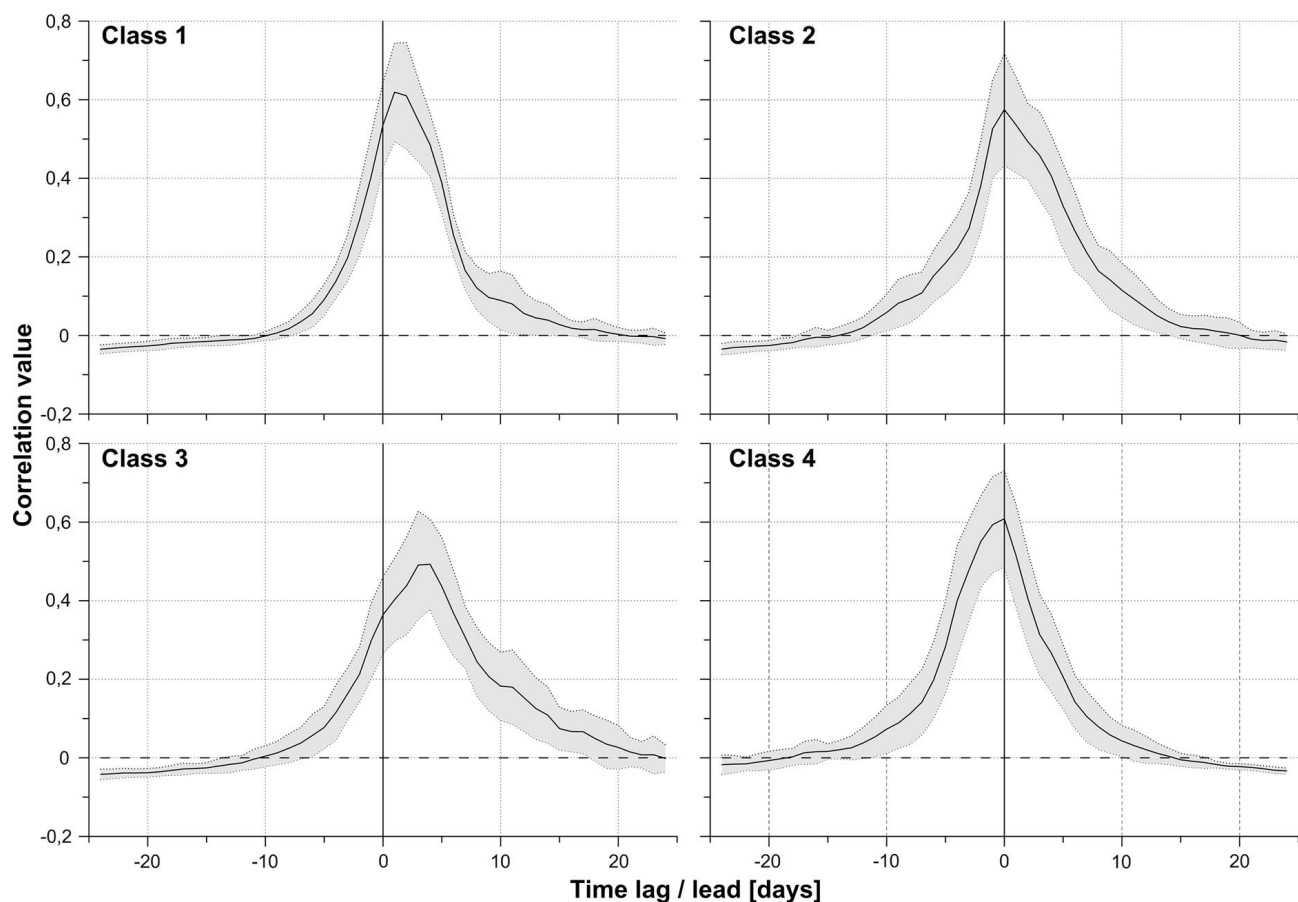
Fig. 3 Average cross-correlograms and their standard deviations for individual classes

the autocorrelation analysis was conducted twice: for the entire set and with the extremes eliminated. For the determination of extreme values, two threshold values were used. The first was put at the place of disruption of the frequency histogram (selection 1), the other cut off the data outliers (selection 2) (Fig. 5).

The Pearson's coefficient of product–moment correlation between all the cross-correlograms was calculated. Next, based on the r value, a hierarchical cluster analysis using Ward's method was carried out (Ward Jr 1963). Finally, on the basis of a plot of linkage distances across steps, the cross-correlograms were divided into four classes (Fig. 3).

3 Results

3.1 Data characteristics

The mean annual alder, birch, and hazel pollen concentrations were 2,952.67, 7,657.00, and 435.65

grains, respectively. The deposition of pollen of those taxa differed widely between particular years and measurement sites. The annual alder pollen totals ranged from 429.2 grains (Sosnowiec in 2009) to 13,582.2 (Poznań in 2010). The annual birch pollen counts varied from 1,182.6 (Rzeszów in 2009) to 34,095 (Lublin in 2003), and hazel pollen from 68.4 (Gdańsk in 2005) to 1,267.2 (Kraków in 2010).

The start of an average *Alnus* pollen season was on the 67th day of the year (March 8) (SD = 16.08 days), and its average duration was 26 days. The earliest start of the *Alnus* season occurred on 2 February 2004 in Kraków, and the latest, on 1 April 2005 in Gdańsk. The *Betula* season started on the 105th day of the year (April 15) (SD = 5.58 days) and lasted 17.5 days on the average. The Gdańsk results differed from the other sites, with an average birch pollen season starting here on April 22 (112th day of the year). The average duration of the *Corylus* pollen season was about 30 days, with the season starting on the 60th day of the year (March 1) (SD = 17.73 days). Variations in the
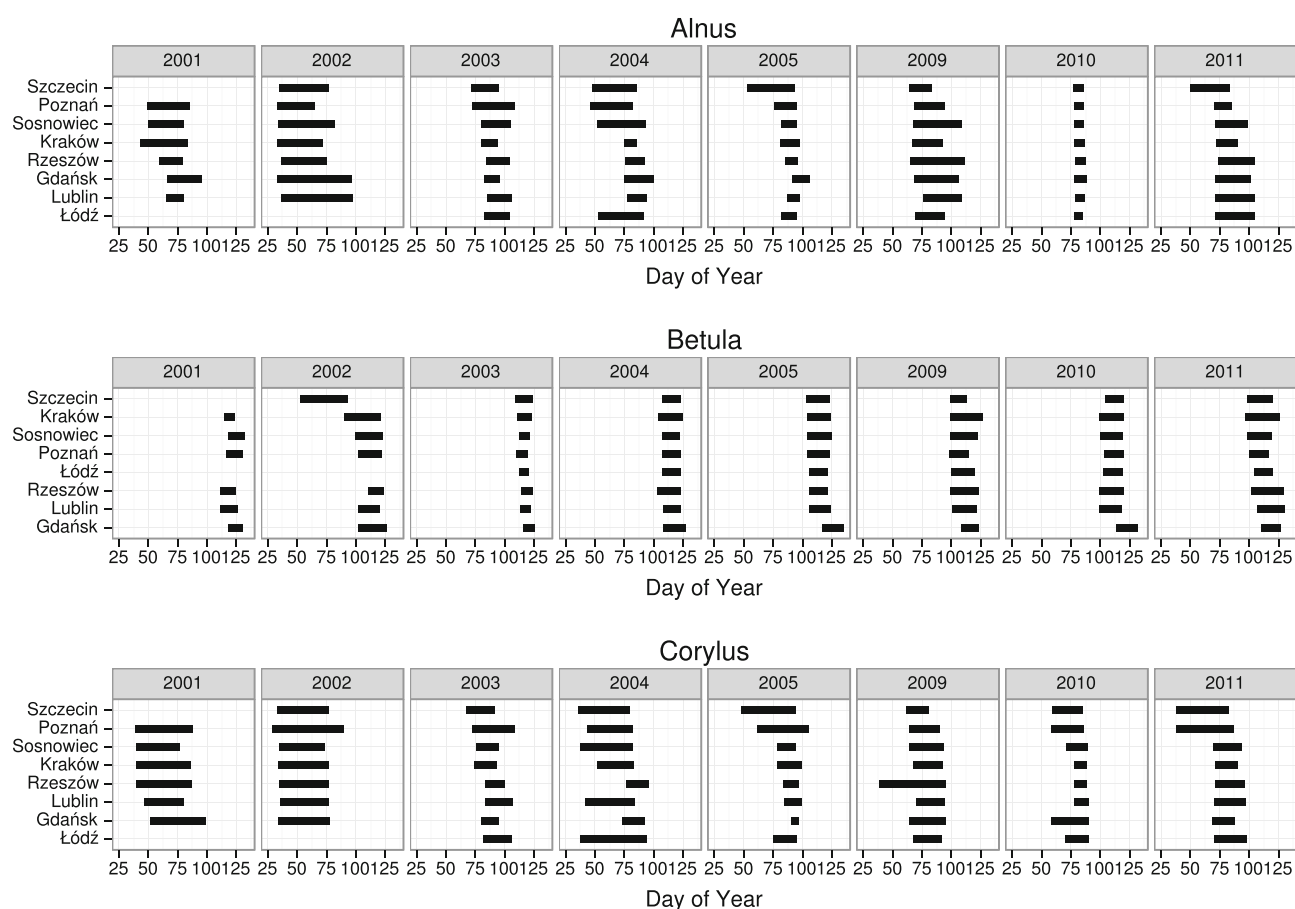
Fig. 4 Length of the pollen season for the same taxa in particular years at various locations. The sites are ordered by the average starting date of the pollen season

duration of the pollen season depended not only on the taxon, but also on the location. The standard deviation of the duration value for alder, birch, and hazel was 13.33, 5.77, and 13.31 days, respectively (Fig. 4).

The descriptive statistics of daily pollen concentrations were calculated for the pollen seasons. Pollen concentration data were characterised by a strongly right-skewed frequency distribution. The median values of 19.0 (SD = 267.62), 133.0 (SD = 845.05), and 4.7 (SD = 30.94) grains per cubic metre of air represented *Alnus*, *Betula*, and *Corylus*, respectively. Pollen counts differed widely among years and sites. The lowest pollen count median for all data was in 2004, at 9.0, and the highest in 2010, at 124.0. The highest medians of daily pollen concentrations were recorded in Szczecin (*Alnus*—53.0) and Lublin (*Betula*—279.2, *Corylus*—13.0), and the lowest ones in Sosnowiec (*Alnus* - 10.0), Kraków (*Betula*—66.0), and Poznań (*Corylus*—1.0).

3.2 Temporal autocorrelation

Autocorrelation coefficients of the analysed taxa with a 1-day lag were calculated (Table 2). The average autocorrelation of daily counts was 0.68 with a range of 0.47–0.84. The autocorrelation figures of daily pollen counts with a 1-day lag were similar in the individual taxa. The average autocorrelation of daily *Alnus*, *Betula*, and *Corylus* pollen counts with a 1-day lag was 0.71 (range 0.47–0.84), 0.71 (range 0.58–0.75), and 0.65 (range 0.54–0.75), respectively.

On average, correlogram shows a decline for the 3.5 days lag (*Alnus*—5.2, *Betula*—2.9, *Corylus*—3.5). Afterwards, the value kept decreasing steadily. Thus, autocorrelation coefficients reduced to zero after an average of 15.0 days (range 8.2–22.5) (Figs. 5, 6; Table 2). The removal of extreme values increased autocorrelation (0.70 on average) and expanded its range (19.5 and 18.2 days on average) (Tables 2, 3).

Table 2 Parameters of correlogram models of pollen counts

Taxa	City	Data subset	C_0	T1	C_1	A_1	T2	C_2	A_2	T3	C_3	A_3	$r-1$ day
Alnus	Gdańsk	All data	0.20	Sph	0.59	2.50	Sph	0.21	11.00				0.471
Alnus	Gdańsk	Selection 1	0.15	Exp	0.58	5.00	Sph	0.24	24.00				0.565
Alnus	Gdańsk	Selection 2	0.13	Exp	0.44	3.25	Sph	0.18	10.50	Sph	0.20	20.25	0.565
Alnus	Kraków	All data	0.05	Sph	0.48	6.50	Sph	0.48	11.75				0.772
Alnus	Kraków	Selection 1	0.10	Sph	0.30	3.00	Sph	0.61	22.50				0.729
Alnus	Kraków	Selection 2	0.07	Sph	0.26	2.75	Sph	0.21	11.50	Sph	0.48	19.25	0.741
Alnus	Lublin	All data	0.06	Sph	0.56	4.75	Sph	0.38	10.25				0.696
Alnus	Lublin	Selection 1	0.08	Sph	0.25	2.50	Sph	0.16	7.50	Sph	0.49	15.75	0.722
Alnus	Lublin	Selection 2	0.01	Sph	0.35	2.80	Sph	0.22	8.50	Sph	0.42	14.00	0.751
Alnus	Łódź	All data	0.10	Sph	0.91	8.50							0.761
Alnus	Łódź	Selection 1	0.00	Sph	0.26	3.50	Sph	0.74	10.50				0.776
Alnus	Poznań	All data	0.00	Sph	0.33	4.75	Sph	0.48	9.75	Sph	0.20	15.00	0.835
Alnus	Poznań	Selection 1	0.15	Sph	0.27	4.75	Sph	0.58	15.25				0.709
Alnus	Rzeszów	All data	0.07	Sph	0.28	4.00	Sph	0.65	8.20				0.709
Alnus	Rzeszów	Selection 1	0.04	Sph	0.55	3.25	Sph	0.39	15.25				0.685
Alnus	Rzeszów	Selection 2	0.11	Sph	0.31	3.00	Sph	0.56	10.25				0.659
Alnus	Sosnowiec	All data	0.04	Sph	0.46	5.25	Sph	0.51	11.00				0.750
Alnus	Sosnowiec	Selection 1	0.02	Sph	0.18	4.25	Exp	0.80	13.50				0.725
Alnus	Sosnowiec	Selection 2	0.00	Exp	0.37	6.50	Sph	0.62	10.25				0.789
Alnus	Szczecin	All data	0.09	Sph	0.30	5.00	Sph	0.62	17.75				0.722
Alnus	Szczecin	Selection 1	0.17	Sph	0.16	3.00	Sph	0.66	22.00				0.666
Alnus	Szczecin	Selection 2	0.09	Sph	0.24	4.75	Sph	0.67	17.50				0.740
Betula	Gdańsk	All data	0.09	Sph	0.26	2.75	Sph	0.41	6.25	Sph	0.26	16.25	0.652
Betula	Gdańsk	Selection 1	0.02	Sph	0.32	2.50	Sph	0.32	6.75	Sph	0.36	16.50	0.706
Betula	Kraków	All data	0.10	Sph	0.15	2.25	Sph	0.16	10.75	Sph	0.61	20.50	0.740
Betula	Kraków	Selection 1	0.13	Sph	0.14	2.50	Sph	0.12	13.50	Sph	0.66	31.00	0.725
Betula	Kraków	Selection 2	0.04	Sph	0.21	3.25	Sph	0.22	10.75	Sph	0.58	27.00	0.800
Betula	Lublin	All data	0.18	Sph	0.25	2.25	Sph	0.27	10.00	Sph	0.32	17.00	0.600
Betula	Lublin	Selection 1	0.05	Sph	0.15	2.50	Sph	0.20	9.75	Sph	0.65	22.25	0.794
Betula	Łódź	All data	0.00	Sph	0.49	2.75	Sph	0.52	11.75				0.686
Betula	Łódź	Selection 1	0.07	Sph	0.16	2.25	Sph	0.33	10.75	Sph	0.47	20.25	0.748
Betula	Łódź	Selection 2	0.08	Sph	0.21	2.75	Sph	0.29	9.25	Sph	0.45	19.50	0.664
Betula	Poznań	All data	0.03	Sph	0.23	2.25	Sph	0.24	9.75	Sph	0.52	17.50	0.748
Betula	Poznań	Selection 1	0.06	Sph	0.22	2.25	Sph	0.18	9.75	Sph	0.57	20.00	0.744
Betula	Rzeszów	All data	0.10	Sph	0.25	2.75	Sph	0.20	6.75	Sph	0.45	18.00	0.647
Betula	Rzeszów	Selection 1	0.10	Sph	0.12	3.00	Sph	0.22	8.75	Sph	0.57	21.80	0.756
Betula	Rzeszów	Selection 2	0.14	Sph	0.12	3.00	Sph	0.17	8.75	Sph	0.61	24.00	0.681
Betula	Sosnowiec	All data	0.02	Exp	0.40	2.75	Sph	0.29	8.50	Sph	0.30	13.75	0.578
Betula	Sosnowiec	Selection 1	0.03	Sph	0.22	3.75	Sph	0.26	6.75	Sph	0.50	17.50	0.767
Betula	Szczecin	All data	0.19	Sph	0.24	4.50	Sph	0.58	16.50				0.678
Betula	Szczecin	Selection 1	0.00	Sph	0.20	2.50	Sph	0.19	9.50	Sph	0.63	19.25	0.822
Corylus	Gdańsk	All data	0.05	Exp	0.65	4.75	Sph	0.29	12.50				0.657
Corylus	Gdańsk	Selection 1	0.36	Sph	0.21	3.75	Sph	0.41	20.75				0.492
Corylus	Kraków	All data	0.05	Sph	0.34	3.00	Sph	0.27	13.75	Sph	0.36	22.50	0.752

Table 2 continued

Taxa	City	Data subset	C_0	T1	C_1	A_1	T2	C_2	A_2	T3	C_3	A_3	$r - 1$ day
Corylus	Kraków	Selection 1	0.05	Sph	0.21	4.25	Sph	0.29	15.25	Sph	0.48	24.50	0.813
Corylus	Lublin	All data	0.15	Sph	0.31	2.25	Sph	0.29	9.00	Sph	0.25	16.75	0.597
Corylus	Lublin	Selection 1	0.18	Exp	0.21	3.50	Sph	0.57	21.50				0.653
Corylus	Lublin	Selection 2	0.10	Sph	0.26	2.50	Sph	0.22	15.00	Sph	0.42	22.25	0.713
Corylus	Łódź	All data	0.22	Sph	0.31	2.50	Sph	0.29	9.50	Sph	0.18	18.25	0.540
Corylus	Łódź	Selection 1	0.14	Sph	0.36	2.25	Sph	0.22	11.00	Sph	0.27	20.00	0.604
Corylus	Poznań	All data	0.16	Sph	0.21	3.00	Sph	0.64	17.75				0.682
Corylus	Poznań	Selection 1	0.07	Sph	0.37	3.50	Sph	0.56	17.75				0.685
Corylus	Rzeszów	All data	0.03	Sph	0.66	3.75	Sph	0.31	14.25				0.697
Corylus	Rzeszów	Selection 1	0.13	Sph	0.45	2.50	Sph	0.39	16.00				0.590
Corylus	Rzeszów	Selection 2	0.00	Sph	0.68	3.50	Sph	0.32	15.00				0.704
Corylus	Sosnowiec	All data	0.09	Sph	0.46	4.25	Sph	0.44	12.75				0.655
Corylus	Sosnowiec	Selection 1	0.09	Sph	0.25	2.25	Sph	0.18	8.00	Sph	0.40	17.25	0.643
Corylus	Sosnowiec	Selection 2	0.05	Exp	0.50	4.25	Sph	0.44	14.25				0.635
Corylus	Szczecin	All data	0.20	Sph	0.28	4.25	Sph	0.53	20.75				0.648
Corylus	Szczecin	Selection 1	0.18	Sph	0.22	3.75	Sph	0.60	22.25				0.671
Corylus	Szczecin	Selection 2	0.16	Sph	0.24	3.25	Sph	0.61	23.00				0.674

C_0—nugget component (error/short time variation), function type—T1 (type 1), T2 (type 2), T3 (type 3): Sph—Spherical, Exp—Exponential, time span of individual structures in days (A_1, A_2 and A_3), their share in total variability (C_1, C_2 and C_3), and the correlation coefficient with a 1-day delay ($r - 1$ day) for distinguished taxons (*Alnus*, *Betula*, *Corylus*), cities (Gdańsk, Kraków, Lublin, Łódź, Poznań, Rzeszów, Sosnowiec, Szczecin), and subsets of data (all data and data with extremes eliminated—selection 1 and selection 2)

Table 3 Descriptive statistics of parameters of correlogram models

Data subset	C_0		C_1		C_2		C_3^a		A_1		A_2		A_3^a	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
All data	0.09	0.07	0.39	0.18	0.39	0.15	0.35	0.14	3.80	1.55	11.75	3.65	17.55	2.53
Selection 1	0.10	0.08	0.27	0.12	0.38	0.20	0.50	0.12	3.18	0.83	14.10	5.63	20.50	4.17
Selection 2	0.08	0.05	0.32	0.15	0.36	0.19	0.45	0.13	3.50	1.10	12.65	4.18	20.89	4.11

C_0—nugget component (error/short time variation), share in total variability (C_1, C_2 and C_3), and time span of individual structures in days (A_1, A_2 and A_3) for distinguished subsets of data (all data and data with extremes eliminated—selection 1 and selection 2)

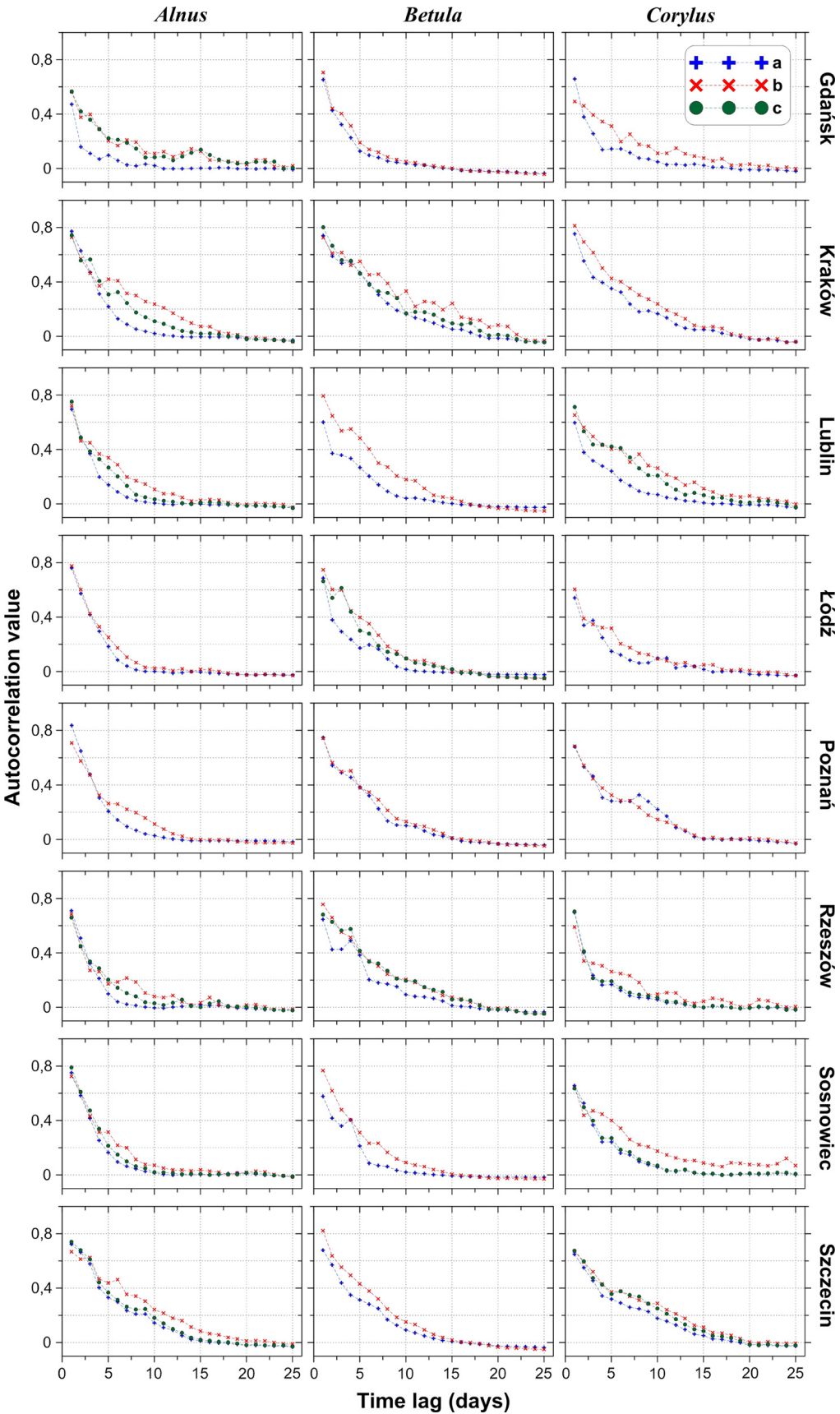[a] Third structure was present in 29 of 61 correlogram models (47.5 %)

At least two individual structures were found in 60 out of the 61 correlogram models. A third structure was identified for 29 correlogram models (47.5 %): 5 out of 22 for *Alnus* (22.7 %), 17 out of 19 for *Betula* (89.5 %), and 7 out of 20 for *Corylus* (35.0 %) (Tables 2, 3).

3.3 Space–time autocorrelation

The correlation of individual pollen counts between the monitoring sites as a function of distance was shown in Figs. 7 (the same day) and 8 (with a 1-day offset). Few sites with regimes of pollen counts different to the remaining were identified. Pollen concentrations in Gdańsk were notably distinctive, especially to those noted in Szczecin, Poznań, and Lublin. Also, the correlation between some pairs of the cities was higher than others, especially in Kraków–Łódź (*Alnus, Betula,* and *Corylus* pollen), Lublin–Sosnowiec, Szczecin–Sosnowiec (*Betula* pollen), Gdańsk–Łódź, and Szczecin–Kraków (*Corylus* pollen).

Correlation coefficients of 28 individual pairs of locations were calculated for each taxon. The average

Alnus   *Betula*   *Corylus*

Gdańsk

Kraków

Lublin

Łódź

Poznań

Rzeszów

Sosnowiec

Szczecin

Autocorrelation value

Time lag (days)

a
b
c

0,8   0,4   0   0   5   10   15   20   25

51

◄**Fig. 5** Matrix of sample (experimental) correlogram plots for the entire data set (*a*) and after the elimination of extreme figures (*b*-first threshold, *c*-second threshold) for the individual locations (*rows*) and taxa (*columns*)

value was 0.58 (range 0.20–0.89). Correlation coefficients varied from 0.566 to 0.583. Furthermore, they were clearly, though not very strongly, dependent on the distance between pairs of monitoring sites (Fig. 7). With a 1-day lag, the correlation coefficients dropped to 0.49 (range 0.15–0.70) but the impact of distance still existed. The Kolmogorov–Smirnov two-sample test showed no significant differences between correlation coefficients of pairs of the observed taxa (*Alnus–Betula*, *Alnus–Corylus*, and *Betula–Corylus*).

Figures 9, 10, 11 show cross-correlograms of *Alnus*, *Betula*, and *Corylus* pollen counts in Poland. Cross-correlograms of alder were the most homogeneous, with symmetric and simple shapes. Moreover, the temporal range of pollen counts was the shortest, in most cases it did not exceed 10 days. Most of birch and hazel cross-correlograms had an asymmetrical shape, often with a few oscillations. Its range was generally longer, sometimes even more than 20 days. Furthermore, correlation rose along with a 1- to 5-day lag/lead in numerous pairs of monitoring sites. The strongest variation was seen in the case of Gdańsk. Most of its cross-correlograms were asymmetric and showed considerable variation.

The average cross-correlograms and their standard deviations for individual classes was shown in Fig. 3. In classes 2 and 4, the maximum correlation occurs on the same day (zero lag). They differ in a rate of correlation decrease in the direction of leads and delays. Their shapes are almost a reflection symmetry. Class 1 has a symmetric shape with a maximum delayed by 1 day. Class 3 has the lowest maximum correlation (0.5), the largest shift (lag of 4–5 days), and the most asymmetric shape.

The Chi-square test was used to examine whether the cross-correlogram classes (Fig. 3) depended on the taxon or the location. It showed a statistically significant relationship between the classes and the taxa (p value $\simeq 0.0002$), and between the classes and the location (p value $\simeq 0.01$). The first class was mostly *Alnus* (80 % of the class), the second class—*Corylus* (55 %), and the third class—*Betula* (57 %). Only in the fourth class, no taxa were dominating. However, the fourth class differentiates between the locations.

This class occurred mostly in Kraków (10 out of 52) and Rzeszów (11 out of 52), and very rarely in Gdańsk (2 out of 52) and Szczecin (1 out of 52). In the third class, differentiation between the locations also occurred, with zero or a few cases in Sosnowiec (0 out of 28), Kraków (1 out of 28), and Rzeszów (1 out of 28) (Table 4).

The locations were classified by the frequency of the occurrence of the cross-correlogram class (Figs. 3, 9, 10, 11). There were major similarities between Kraków and Rzeszów, and Lublin and Łódź. Three groups were distinguished, the first consisting of Kraków, Rzeszów, and Sosnowiec; the second of Lublin, Łódź, and Poznań; and the third of Gdańsk and Szczecin.

4 Discussion

Local and regional flora and land-use substantially affect the pollen spectrum. However, pollen concentrations in the air depend on more factors, including the vegetation structure, ability of particular taxa to spread their pollen, and the topoclimate and weather conditions in the current and preceding years. Especially, the effect of temperature on the intensity of the pollen season is often stressed (Corden et al. 2002; Rasmussen 2002).

Most pollen studies have focused on changes through time; hence, patterns across space remain largely unknown (Grewling et al. 2012b; León Ruiz et al. 2012; Ziello et al. 2012). Little research has been devoted to a spatial analysis of tree pollen in Poland. Myszkowska et al. (2010) investigated the relationship between the geographical location and the dynamics of *Alnus* and *Corylus* pollen seasons in Poland. However, temporal and spatiotemporal autocorrelation has not been taken into account before.

The form of the correlograms provides an understanding of the temporal variations in *Alnus*, *Betula*, and *Corylus* pollen counts in Poland. Notable changes in the shapes of the correlograms suggest that variations may be associated with three main groups of factors. 10 % of pollen count variations can be due to random factors, including diurnal fluctuations and measurement errors (Tables 2, 3). The autocorrelation value dropped substantially after 3.5 days. Thus, approximately 30–40 % of the variation could be caused by an exchange of air masses after the passage

Fig. 6 Matrix of
correlogram models (**a**—all
data, **b**—first threshold) for
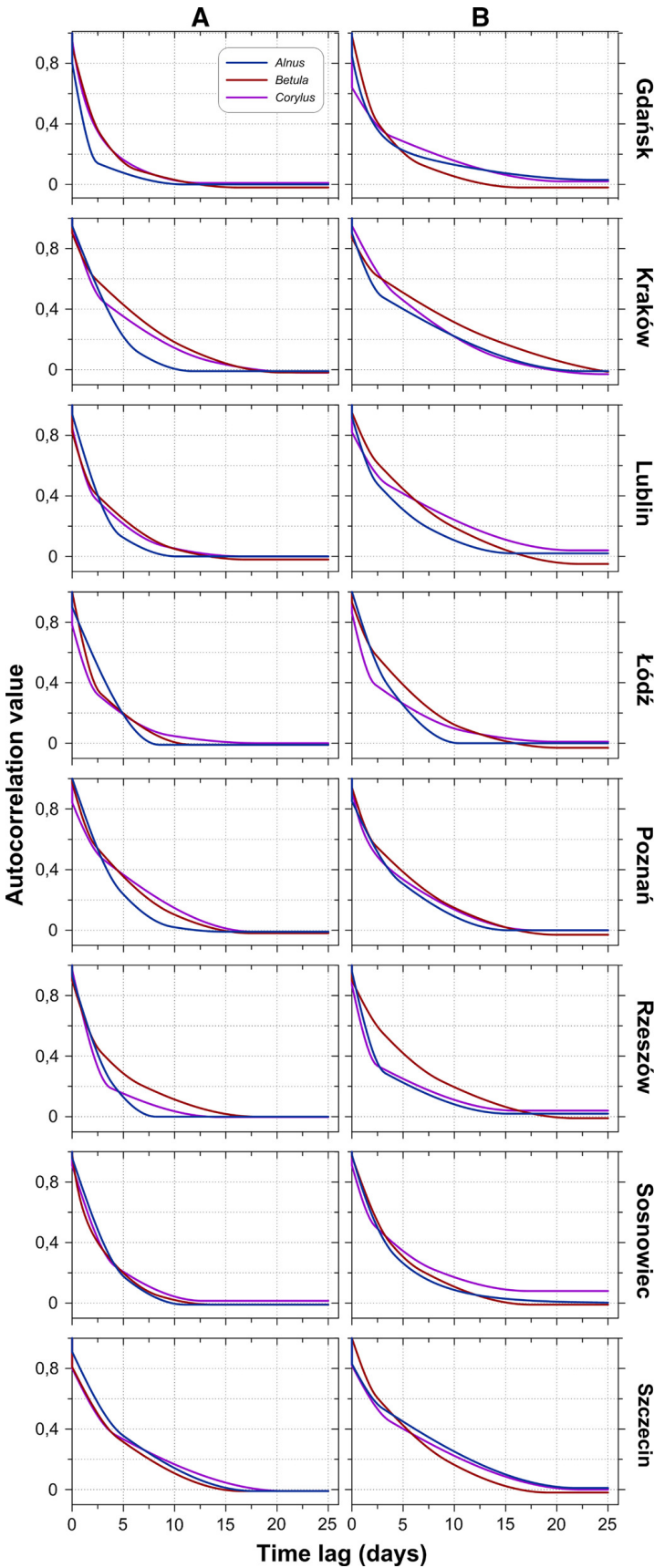various taxa at the same
locations

Fig. 7 Correlation of concentrations of individual pollen taxa on the same day between the locations as a function of distance. The diagrams present linear regression curves and their 95 % confidence intervals (*shaded*), formulae for the models employed, and the significance level (*p* value) of functions. Only outlying pairs of stations are labelled
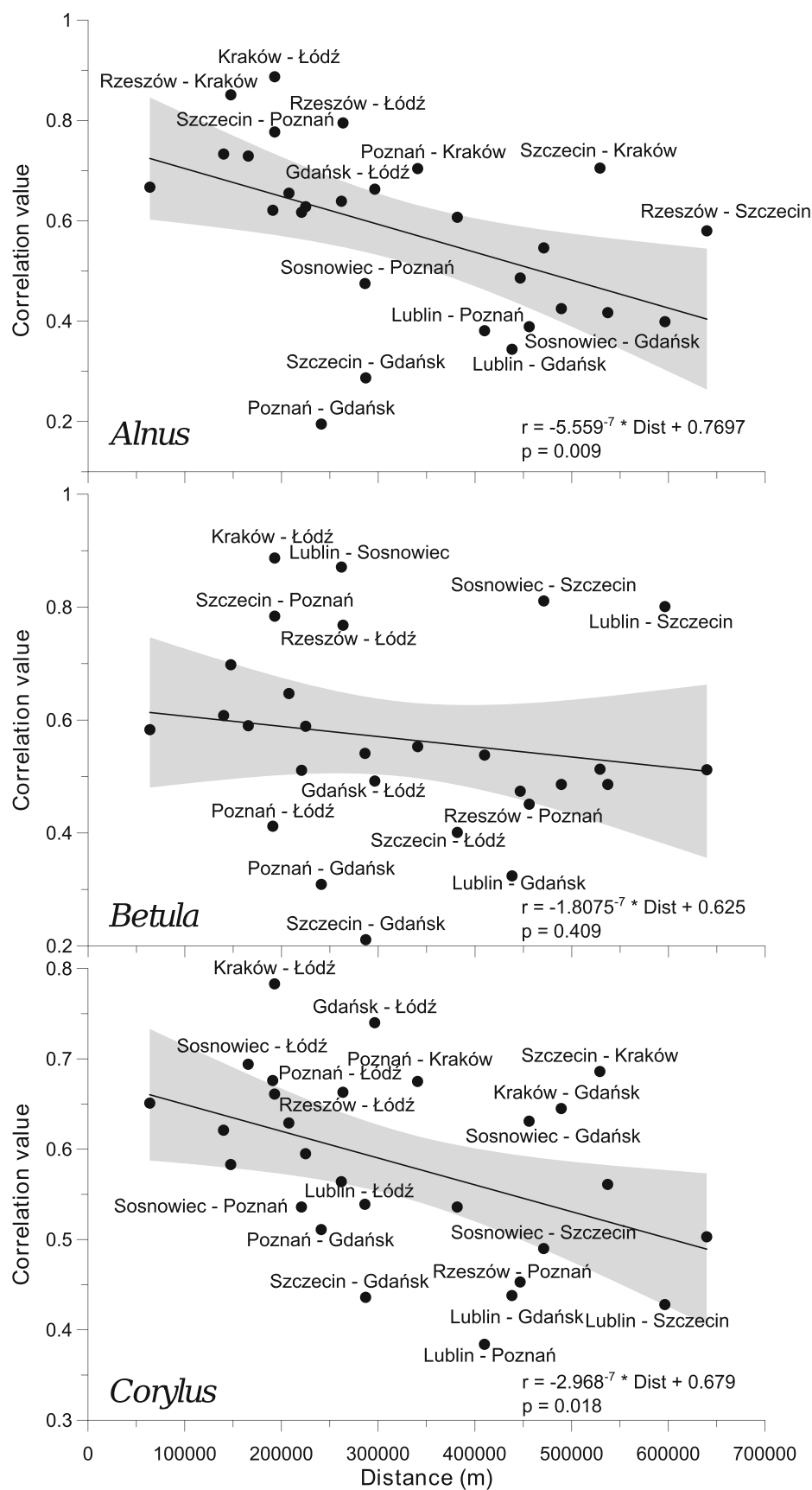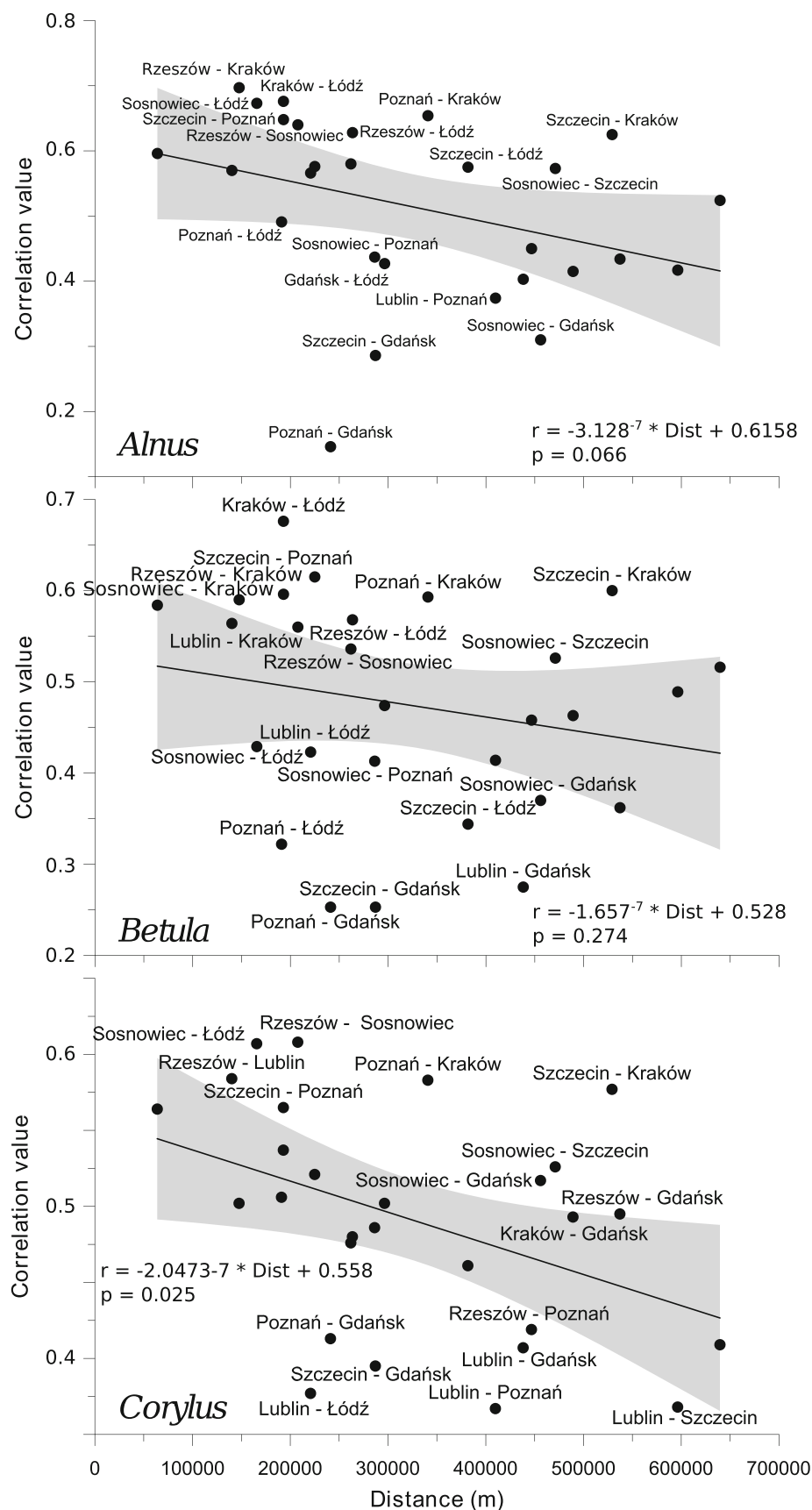
Springer

Fig. 8 Correlation of concentrations of individual pollen taxa with a 1-day lag between the locations as a function of distance. The diagrams present linear regression curves and their 95 % confidence intervals (*shaded*), formulae for the models employed, and the significance level (*p* value) of functions. Only outlying pairs of stations are labelled

of a single weather front. The remaining 50–60 % depends on longer-lasting factors, exceeding 4 days.

Table 4 Frequency of cross-correlation classes at each location

City	Class 1	Class 2	Class 3	Class 4
Gdańsk	4	6	9	2
Kraków	2	8	1	10
Lublin	6	4	4	7
Łódź	5	7	3	6
Poznań	2	8	4	7
Rzeszów	2	7	1	11
Sosnowiec	5	8	0	8
Szczecin	4	10	6	1

According to Kotas et al. (2013), the variability of most of strings with a single air mass type is between 1 and 3 days. The influx of a new air mass has the physical characteristics of which differ from the previous one. An increase in the correlation between pollen counts in pairs of monitoring sites is delayed by 1–3 days (Figs. 9, 10, 11). This relationship could be partly associated with the movement of atmospheric fronts and the distance between the sites.

It was also found that pollen characteristics in Gdańsk differ from the rest of the sites. An average birch pollen season starts in Gdańsk 6 days later than in the rest of the country. The Gdańsk cross-correlograms were heterogeneous and asymmetric. Therefore, concentrations of tree pollen there vary considerably and clearly differ from many of the other monitoring sites. The uniqueness of Gdańsk can
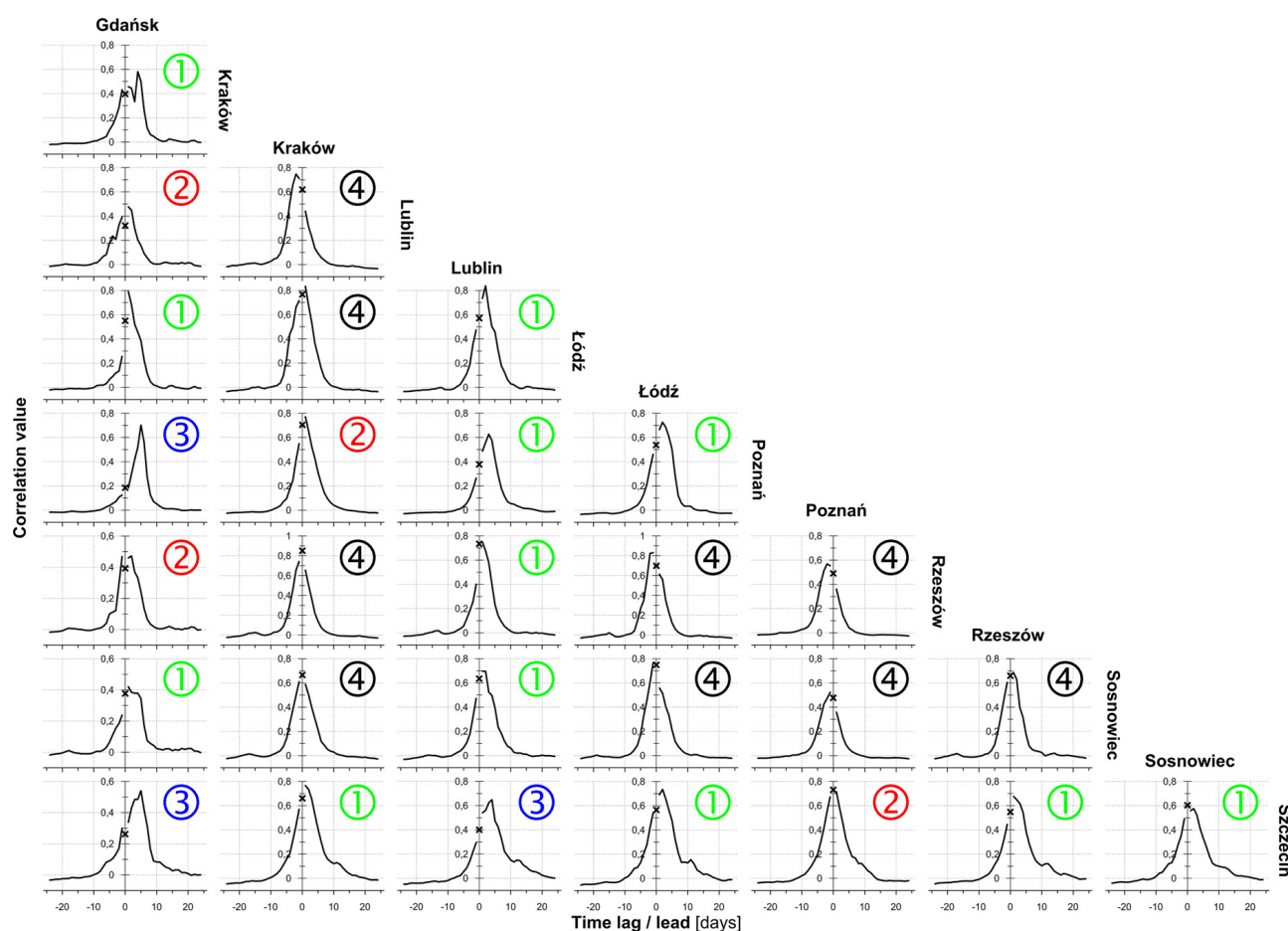

Fig. 9 Diagrams of *Alnus* cross-correlation between locations with lead (*left branch*) and lag (*right branch*) time using all the data. The *black cross* shows the synchronous correlation value (lag/lead = 0 days). A comparison of these charts allows capturing the asymmetric relationship of time-predominant leads or delays. Cross-correlograms are labelled by class number (see Fig. 3 and description on page 5)
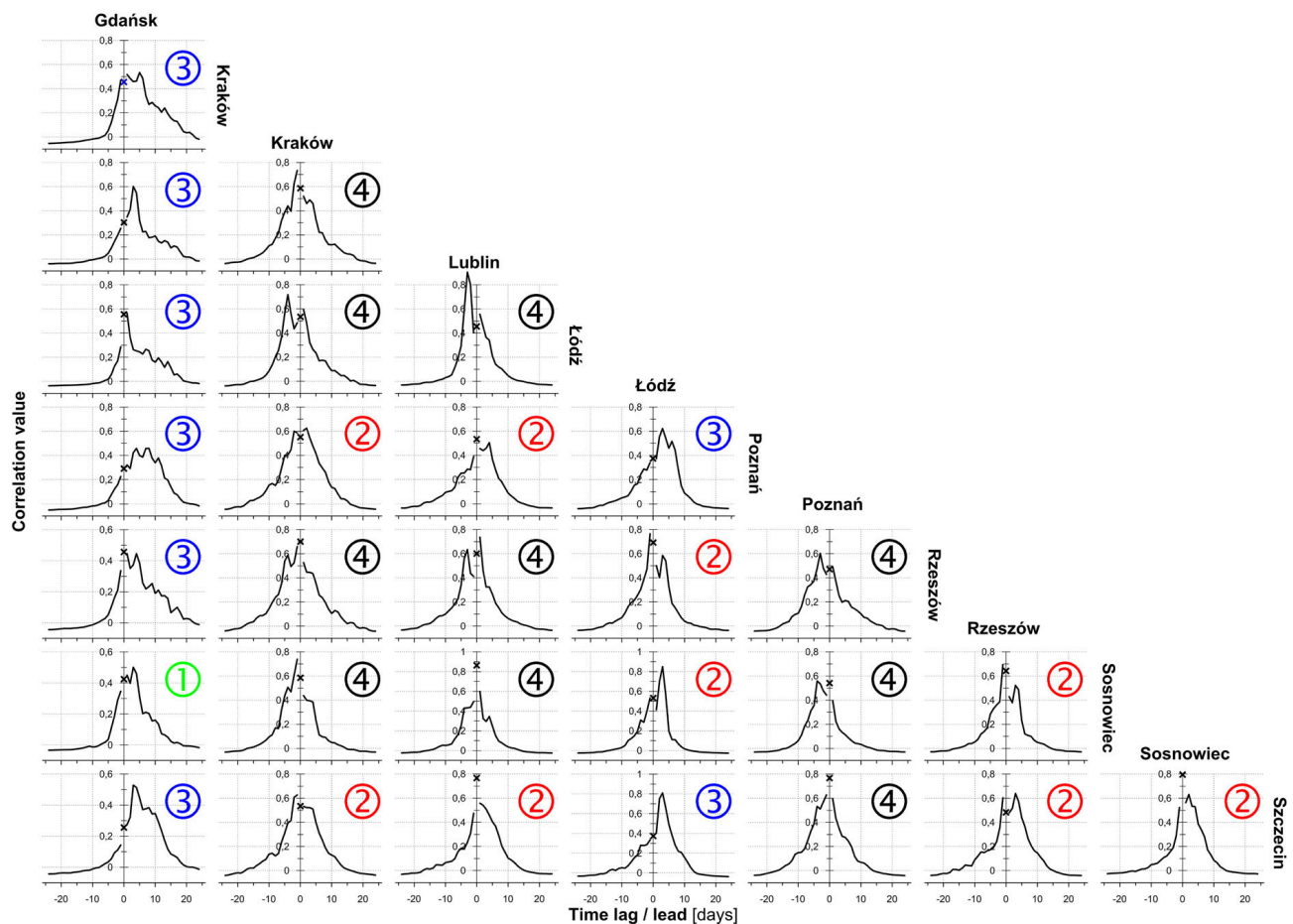
Fig. 10 Diagrams of *Betula* cross-correlation between locations with lead (*left branch*) and lag (*right branch*) time using all the data. The *cross* shows the synchronous correlation value (lag/lead = 0 days). A comparison of these charts allows capturing the asymmetric relationship of time-predominant leads or delays. Cross-correlograms are labelled by class number (see Fig. 3 and description on page 5)

be explained by its coastal location and the impact of the Baltic Sea on the local climate.

The size of pollen grains of the studied taxa is similar (*Alnus*−22 μm × 34 μm, *Betula*−18 μm × 28 μm, *Corylus*−18 μm × 23 μm) (Accorsi et al. 1991). They were found to be among the longest-occurring pollen grains in the air (Dyakowska 1937). Alder, birch, and hazel pollen grains in Poland have similar transport and dispersion conditions. However, an analysis of the cross-correlogram classes shows that the character and range of autocorrelation depend on the location and the taxon. Moreover, the correlation coefficients decrease as the distance between sites increases. As follows from the cross-correlograms classes (Fig. 3), there are three groups of locations connected with the local topography and climate. The first consists of the coastal cities of Gdańsk and Szczecin. Their climate is influenced by Atlantic air masses and the proximity of the Baltic Sea—winters are mild, summers are not very hot, and the weather often changes. The second group is mostly linked with lowlands (Poznań, Łódź) and highlands (Lublin). The last group embraces Kraków, Sosnowiec, and Rzeszów. Those cities are located in southern Poland, and their climate is influenced by the Carpathian Mountains.

The location of monitoring sites in Poland is not random. They are primarily situated in populous cities with aerobiological research facilities. Urban areas modify the climate of their regions as a result of local changes in the properties of the land surface and the atmosphere. Consequently, those areas have a higher temperature than the surroundings. They
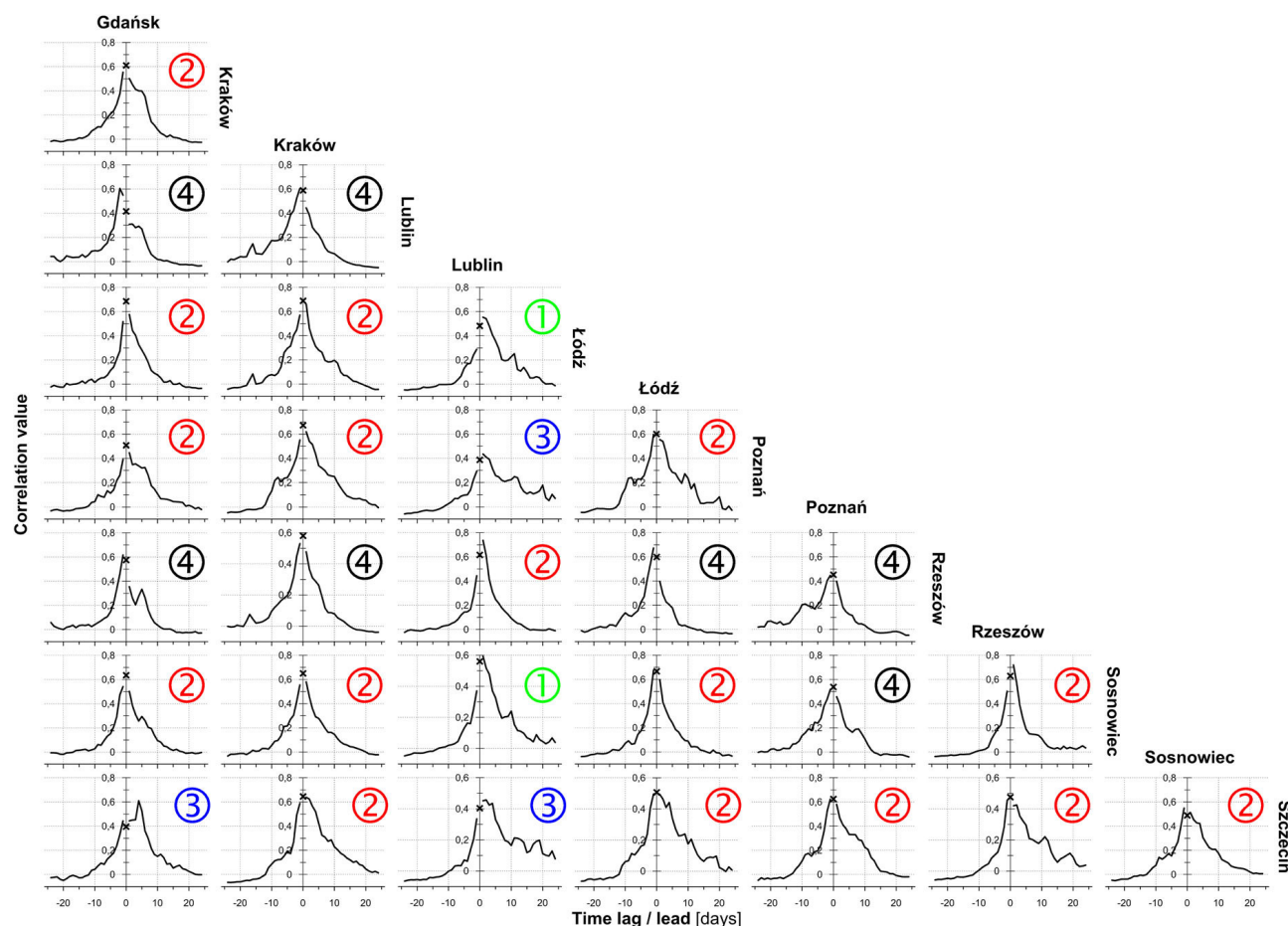
Fig. 11 Diagrams of *Corylus* cross-correlation between locations with lead (*left branch*) and lag (*right branch*) time using all the data. The *cross* shows the synchronous correlation value (lag/lead = 0 days). A comparison of these charts allows capturing the asymmetric relationship of time-predominant leads or delays. Cross-correlograms are labelled by class number (see Fig. 3 and description on page 5)

are often referred to as urban heat islands. The average temperature in large Polish cities is 1–1.2°C higher than in the surrounding areas, and the growing seasons last longer (Szymanowski 2005). Jochner et al. (2012) suggest that in completely urbanised areas, phenological phases start 2.6–7.6 days in advance. Therefore, caution should be exercised when interpolating the results for areas between stations. In this study, data from eight Polish monitoring sites were used, with none in the south-western and north-eastern parts of the country. Myszkowska et al. (2010) suggested higher concentrations of *Alnus* pollen in the north-eastern region of Poland to be caused by its long-distance transport and the dominant westerly direction of wind. An analysis of data from those parts of Poland could reveal other relationships in variations of tree pollen.

5 Conclusion

- Temporal variations in the *Alnus*, *Betula*, and *Corylus* pollen counts seem to be associated with three main groups of factors: (i) diurnal variability and measurement errors, (ii) an exchange of air masses after the passage of a single weather front (every 3.5 days), and (iii) longer-lasting factors
- Due to the recurrence of circulation patterns, an increase in the correlation between pollen counts in pairs of monitoring sites is delayed by 1–3 days
- The start, course, and intensity of the pollen season in Gdańsk differ from the rest of the studied sites, possibly due to its coastal location and the local climate

– Three groups of locations connected with the local topography and climate could be distinguished. The first consists of the coastal cities, the second group is linked with lowlands and highlands, and the last group embraces those cities located in southern Poland.

– It would be worthwhile to analyse the relationship between variations in pollen counts and atmospheric circulation patterns

References

Accorsi, C. A., Bandini, M., Romano, B., Frenguelli, G., & Mincigrucci, G. (1991). Allergenic pollen: Morphology and microscopic photographs. In F. Spieksma, S. Bonini, & G. D'Anato (Eds.), *Allergenic pollen andpollinosis in Europe*. New York: Blackwell Scientific Publications.

Bleines, C., Deraisme, J., Geffroy, F., Jeannée, N., Perseval, S., Rambert, F., et al. (2011). *Isatis technical references*. Isatis release 2011.

Bremer, B., Bremer, K., & Chase, M. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, *161*(2), 105–121.

Chiles, J., & Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*. New Jersey: Wiley Series in Probability and Statistics: Wiley.

Comtois, P. (1998). Statistical analysis of aerobiological data. In P. Comtois, V. Levizzani, & P. Mandrioli (Eds.), *Methods in aerobiology* (pp. 217–258). Bologna: Pitagora Editrice Bologna.

Corden, J., Stach, A., & Millington, W. (2002). A comparison of Betula pollen seasons at two European sites; Derby, United Kingdom and Poznan, Poland (1995–1999). *Aerobiologia*, *1996*, 45–53.

Cowpertwait, P. S., & Metcalfe, A. V. (2010). *Introductory time series with R (Use R!)*. New York: Springer.

Dmochowska, H. (Ed.). (2013). *Concise statistical yearbook of Poland 2013*. Warsaw: Statistical Publishing Establishment.

Dyakowska, J. (1937). *Researches on the rapidity of the falling down of pollen of some trees*. Polska Akademia Umiejętności.

Emberlin, J., Mullins, J., Corden, J., Millington, W., Brooke, M., Savage, M., et al. (1997). The trend to earlier birch pollen seasons in the U.K.: A biotic response to changes in weather conditions? *Grana*, *36*(1), 29–33. doi:10.1080/00173139709362586.

Frenguelli, G. (2003). Basic microscopy, calculating the field of view, scanning of slides, sources of error. *Postepy Dermatologii i Alergologii*, *2*, 227–229.

Garrison, B. A., Koenig, W. D., & Knops, J. M. H. (2006). Spatial synchrony and temporal patterns in acorn production of California Black Oaks. In: *Proceedings of the 6th symposium on Oak woodlands: Today's challenges, tomorrow's opportunities PSW-GTR-217*, pp. 343–356.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation. Applied geostatistics series*. Oxford: Oxford University Press.

Grewling, L., 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.

Grewling, L., Šikoparija, B., Skjøth, C., Radišić, P., Apatini, D., Magyar, D., et al. (2012). Variation in Artemisia pollen seasons in Central and Eastern Europe. *Agricultural and Forest Meteorology*, *160*, 48–59. doi:10.1016/j.agrformet.2012.02.013.

Hájková, L., Nekovář, J., & Richterová, D. (2009). Temporal and spatial variability in allergy-triggering phenological phases of hazel and alder in Czechia. *Folia Oecologica*, *36*(1), 8–19.

Heinzerling, L. M., Burbach, G. J., Edenharter, G., Bachert, C., Bindslev-Jensen, C., Bonini, S., et al. (2009). GA(2)LEN skin test study I: GA(2)LEN harmonization of skin prick testing: Novel sensitization patterns for inhalant allergens in Europe. *Allergy*, *64*(10), 1498–1506. doi:10.1111/j.1398-9995.2009.02093.x.

Hirst, J. M. (1952). An automatic volumetric spore trap. *Annals of Applied Biology*, *39*(2), 257–265. doi:10.1111/j.1744-7348.1952.tb00904.x.

Jochner, S. C., Sparks, T. H., Estrella, N., & Menzel, A. (2012). The influence of altitude and urbanisation on trends and mean dates in phenology (1980–2009). *International Journal of Biometeorology*, *56*(2), 387–394. doi:10.1007/s00484-011-0444-3.

Kaszewski, B. M., Pidek, I. A., Piotrowska, K., & Weryszko-Chmielewska, E. (2008). Annual pollen sums of Alnus in Lublin and Roztocze in the years 2001–2007 against selected meteorological parameters. *Acta Agrobotanica*, *61*(2), 57–64. doi:10.5586/aa.2008.033.

Kluza-Wieloch, M., & Szewczak, J. (2006). Flowering phenology of selected wind pollinated allergenic deciduous tree species. *Acta Agrobotanica*, *59*(1), 309–316.

Kotas, P., Twardosz, R., & Nieckarz, Z. (2013). Variability of air mass occurrence in southern Poland (1951–2010). *Theoretical and Applied Climatology*, *114*(3–4), 615–623. doi:10.1007/s00704-013-0861-9.

León Ruiz, E. J., García Mozo, H., Domínguez Vilches, E., & Galán, C. (2012). The use of geostatistics in the study of floral phenology of Vulpia geniculata (L.) Link. *The Scientific World Journal*, *2012*, 1–19. doi:10.1100/2012/624247.

Liebhold, A., Sork, V., & Peltonen, M. (2003). Peltonen M (2004) Within-population spatial synchrony in mast seeding of North American oaks. *Oikos*, *1*, 156–164.

Lorenc, H. (2005). *Atlas klimatu Polski*. Warsaw: IMGW.

Malkiewicz, M., Klaczak, K., Drzeniecka-Osiadacz, A., Krynicka, J., & Migała, K. (2013). *Types of Artemisia pollen season depending on the weather conditions in Wrocław (Poland), 2002–2011*. Aerobiologia. doi:10.1007/s10453-013-9304-4.

Milewski, W. (Ed.). (2013). *Forests in Poland 2012*. Warsaw: The State Forests Information Centre.

Mothes, N., & Valenta, R. (2004). Biology of tree pollen allergens. *Current Allergy and Asthma Reports*, *4*(5), 384–390.

Myszkowska, D., Jenner, B., Puc, M., Stach, A., Nowak, M., Malkiewicz, M., et al. (2010). Spatial variations in the dynamics of the Alnus and Corylus pollen seasons in Poland. *Aerobiologia*, *26*(3), 209–221. doi:10.1007/s10453-010-9157-z.

Myszkowska, D., Jenner, B., Stępalska, D., & Czarnobilska, E. (2011). The pollen season dynamics and the relationship among some season parameters (start, end, annual total, season phases) in Kraków, Poland, 1991–2008. *Aerobiologia*, *27*(3), 229–238. doi:10.1007/s10453-010-9192-9.

Nilsson, S., & Persson, S. (1981). Tree pollen spectra in the stockholm region (sweden), 1973–1980. *Grana*, *20*(3), 179–182. doi:10.1080/00173138109427661.

Pannatier, Y. (1996). *VARIOWIN software for spatial data analysis in 2D/Yvan Pannatier*. New York: Springer.

Piotrowska, K., & Kubik-Komar, A. (2012). The effect of meteorological factors on airborne Betula pollen concentrations in Lublin (Poland). *Aerobiologia*, *28*(4), 467–479. doi:10.1007/s10453-012-9249-z.

Puc, M. (2003). Characterisation of pollen allergens. *Annals of Agricultural and Environmental Medicine: AAEM*, *10*(2), 143–149.

Puc, M. (2007). The effect of meteorological conditions on hazel (Corylus spp.) and alder (Alnus spp.) pollen concentration in the air of Szczecin. *Acta Agrobotanica*, *60*(2), 65–70. doi:10.5586/aa.2007.032.

Puc, M. (2012). Artificial neural network model of the relationship between Betula pollen and meteorological factors in Szczecin (Poland). *International Journal of Biometeorology*, *56*(2), 395–401. doi:10.1007/s00484-011-0446-1.

R Core Team. (2013). *R: A language and environment for statistical computing*. R foundation for statistical computing, Vienna, Austria. http://www.R-project.org/.

Ranta, H., & Satri, P. (2007). Synchronized inter-annual fluctuation of flowering intensity affects the exposure to allergenic tree pollen in North Europe. *Grana*, *46*(4), 274–284. doi:10.1080/00173130701653079.

Rapiejko, P., Stankiewicz, W., Szczygielski, K., & Jurkiewicz, D. (2007). Progowe stężenie pyłku roślin niezbędne do wywołania objawów alergicznych. *Otolaryngologia Polska*, *61*(4), 591–594. doi:10.1016/S0030-6657(07)70491-2.

Rasmussen, A. (2002). The effects of climate change on the birch pollen season in Denmark. *Aerobiologia*, *18*(3–4), 253–265. doi:10.1023/A:1021321615254.

Rodriguez-Rajo, F. J., Dopazo, A., & Jato, V. (2004). Environmental factors affecting the start of pollen season and concentrations of airborne Alnus pollen in two localities of Galicia (NW Spain). *Annals of Agricultural and Environmental Medicine: AAEM*, *11*(1), 35–44.

Rodriguez-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 by using ARIMA models. *Annals of Agricultural and Environmental Medicine: AAEM*, *13*(1), 25–32.

Rodriguez-Rajo, F. J., Grewling, L., Stach, A., & Smith, M. (2009). Factors involved in the phenological mechanism of Alnus flowering in Central Europe. *Annals of Agricultural and Environmental Medicine: AAEM*, *16*(2), 277–284.

Smith, M., Emberlin, J., Stach, A., Czarnecka-Operacz, M., Jenerowicz, D., & Silny, W. (2007). Regional importance of Alnus pollen as an aeroallergen: A comparative study of Alnus pollen counts from Worcester (UK) and Poznan (Poland). *Annals of Agricultural and Environmental Medicine: AAEM*, *14*, 123–128.

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). *International Journal of Biometeorology*, *52*(4), 311–321. doi:10.1007/s00484-007-0127-2.

Szymanowski, M. (2005). Miejska wyspa ciepła we Wrocławiu [Urban heat island in Wrocław]. In: *Studia Geograficzne— tom 77, Wydawnictwo Uniwersytetu Wrocławskiego*.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*.

Weryszko-Chmielewska, E., Puc, M., & Rapiejko, P. (2001). Comparative analysis of pollen counts of Corylus, Alnus and Betula in Szczecin, Warsaw and Lublin (2000–2001). *Annals of Agricultural and Environmental Medicine: AAEM*, *8*(2), 235–240.

Ziello, C., Sparks, T. H., Estrella, N., Belmonte, J., Bergmann, K. C., Bucher, E., et al. (2012). Changes to airborne pollen counts across Europe. *PloS one 7*(4). doi:10.1371/journal.pone.0034076.

Chapter 4

Forecasting model of *Corylus*, *Alnus*, and *Betula* pollen concentration levels using spatiotemporal correlation properties of pollen count
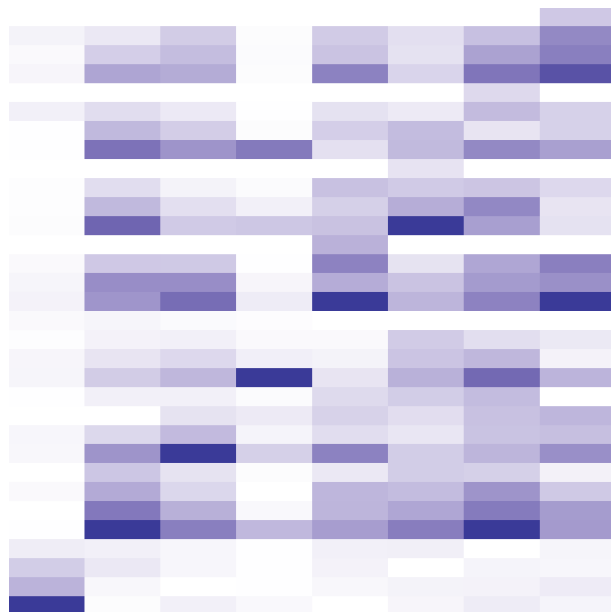
Forecasting model of *Corylus*, *Alnus*, and *Betula* pollen concentration levels using spatiotemporal correlation properties of pollen count

Jakub Nowosad · **Alfred Stach** · **Idalia Kasprzyk** ·
Elżbieta Weryszko-Chmielewska · **Krystyna Piotrowska-Weryszko** ·
Małgorzata Puc · **Łukasz Grewling** · **Anna Pędziszewska** ·
Agnieszka Uruska · **Dorota Myszkowska** · **Kazimiera Chłopek** ·
Barbara Majkowska-Wojciechowska

Abstract The aim of the study was to create and evaluate models for predicting high levels of daily pollen concentration of *Corylus*, *Alnus*, and *Betula* using a spatiotemporal correlation of pollen count. For each taxon, a high pollen count level was established according to the first allergy symptoms during exposure. The dataset was divided into a training set and a test set, using a stratified random split. For each taxon and city, the model was built using a random forest method. *Corylus* models performed poorly. However, the study revealed the possibility of predicting with substantial accuracy the occurrence of days with high pollen concentrations of *Alnus* and *Betula* using past pollen count data from monitoring sites. These results can be used for building (1) simpler models, which require data only from aerobiological monitoring sites, and (2) combined meteorological and aerobiological models for predicting high levels of pollen concentration.

J. Nowosad (✉) · A. Stach
Institute of Geoecology and Geoinformation, Adam Mickiewicz University, Dzięgielowa 27, 61-680 Poznań, Poland
e-mail: nowosad@amu.edu.pl

I. Kasprzyk
Department of Environmental Biology, University of Rzeszów, Zelwerowicza 4, 35-601 Rzeszów, Poland

E. Weryszko-Chmielewska
Department of Botany, University of Life Sciences in Lublin, Akademicka 15, 20-950 Lublin, Poland

K. Piotrowska-Weryszko
Department of General Ecology, University of Life Sciences in Lublin, Leszczyńskiego 58, 20-950 Lublin, Poland

M. Puc
Department of Botany and Nature Conservation, University of Szczecin, Felczaka 3c, 71-412 Szczecin, Poland

Ł. Grewling
Laboratory of Aeropalynology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland

A. Pędziszewska · A. Uruska
Department of Plant Ecology, University of Gdańsk, Wita Stwosza 59, 80-308 Gdańsk, Poland

D. Myszkowska
Department of Clinical and Environmental Allergology, Jagiellonian University Medical College, Śniadeckich 10, 31-531 Kraków, Poland

K. Chłopek
Faculty of Earth Sciences, University of Silesia, Będzińska 60, 41-200 Sosnowiec, Poland

B. Majkowska-Wojciechowska
Department of Immunology, Rheumatology and Allergy, Faculty of Medicine, Medical University, Pomorska 251, 92-215 Łódź, Poland

1 Introduction

Corylus L. (hazel), *Alnus* Mill. (alder), and *Betula* L. (birch) belong to the order Fagales Engl. and the family Betulaceae Gray (Bremer et al. 2009). These trees are very common in the Northern Hemisphere (Kornas and Medwecka-Kornas 2002). The dominant species from this family in Poland are *Alnus glutinosa*, *Alnus incana*, and *Betula pendula*; less common are *Betula pubescens*, *Corylus avellana*, and their cultivars. The start and length of the *Corylus* and *Alnus* pollen seasons are very variable from year to year. Their pollen season in Poland usually begins some time between early February and late March and lasts on average for 30 days (*Corylus*) and 26 days (*Alnus*). The *Betula* pollen season occurs between the middle of April and the middle of May and lasts for approximately 18 days. Its pollen season start and duration are less variable than those of *Corylus* and *Alnus* (Nowosad et al. 2015). Furthermore, the daily and annual pollen counts of *Corylus*, *Alnus*, and *Betula* vary greatly (Nowosad et al. 2015).

Corylus, *Alnus*, and *Betula* pollen are well known for their allergenic properties (Viander and Koivikko 1978), and the occurrence of allergic reactions is connected with pollen concentration levels. According to Rapiejko et al. (2007), the first allergy symptoms are seen during exposure to a concentration of 35 pollen/m^3 of air for *Corylus*, 45 pollen/m^3 of air for *Alnus* , and 20 pollen/m^3 of air for *Betula*. Allergy symptoms in all subjects were noted at concentrations of 80, 85, and 75 pollen/m^3 of air, respectively, for *Corylus*, *Alnus*, and *Betula*. Sensitization rates to tree species of the family Betulaceae in Poland are high: *Corylus*, 22.3 %; *Alnus*, 22.8 %; and *Betula*, 27.7 % (Heinzerling et al. 2009). The pollen allergens from *Corylus*, *Alnus*, and *Betula* are structurally and immunochemically similar. Thus, cross-reactions are very likely between allergens of the *Betula* family (Valenta et al. 1991).

One of aerobiology's objectives is to develop models enabling the prediction of pollen concentration in the air (Rodriguez-Rajo et al. 2006). Forecast models of pollen concentrations have many practical applications. They are highly important for allergy sufferers because predictions can allow them to undertake appropriate treatment. Models could also be useful in agriculture, forestry, and many fields of science. Most of the published results are based on the relationship between pollen season characteristics or on pollen count and meteorological conditions (Bringfelt et al. 1982; Cotos-Yáñez et al. 2004; Castellano-Méndez et al. 2005; Rodriguez-Rajo et al. 2006; Hilaire et al. 2012). Other models have been built based on an operational weather forecast system (Vogel et al. 2008) and the System for Integrated modeLling of Atmospheric coMposition (SILAM) (Sofiev et al. 2013). In Poland, Latałowa et al. (2002) delimited the major meteorological parameters as a basis for future forecast modeling of the atmospheric *Betula* pollen concentration in Gdańsk; Puc (2012) built an artificial neural network model of the relationship between *Betula* pollen and meteorological factors in Szczecin; and Myszkowska (2013) predicted *Betula* pollen season characteristics in Kraków.

Aerobiological surveys have been focused either on the statistical relationship between pollen count and meteorological variables or on a deterministic representation of pollen dispersion. A recent study showed that pollen counts are highly inert, temporally and spatially (Nowosad et al. 2015). Models based on this property would be relatively simple, because they do not require data other than that of pollen concentration. The forecasts would require data only from aerobiological monitoring sites and could be almost completely automated. The aim of this study was to create and evaluate *Corylus*, *Alnus*, and *Betula* pollen concentration level predictions based on previous pollen count values from given sites. To the best of our knowledge, there are no reports in the literature regarding how these kinds of models perform in practice.

2 Materials and methods

2.1 Study area

The studies were conducted in eight cities in Poland (Gdańsk, Kraków, Lublin, Łódź, Poznań, Rzeszów, Sosnowiec, and Szczecin) and covered six years of

measurement (2003–2005, 2009–2011) (Fig. 1; Table 1). The measurements from years 2006–2008 were not available for all of the sites. Therefore, incomplete data were not included.

Poland is a country in Central Europe, extending 649 km from north to south and 689 km from east to west. It is a lowland country with an average elevation of 173 m and a surface area of 312,679 km^2, only 3.1 % of which is higher than an elevation of 500 m. There are five topographic zones in Poland. The order of the zones from north to south is as follows: the Baltic coastal plains, the lake region, the central lowlands, the uplands, and the mountains (Sudeten and Carpathian ranges). Agricultural land covers approximately 60 % of the surface area of the country, and forest, bush, and wooded land cover about 30 %. Built-up and urbanized areas occupy approximately 5 % of the total area (Dmochowska 2013). All of the studied cities are agglomerations surrounded mainly by forests and farmlands. However, the proportion of land use classes around each city is different. Gdańsk

lies on the coast of the central Baltic Sea, with sea to the east and agricultural lands or forests to the west and south; Szczecin is located between Dąbie Lake to the northeast, forests to the southeast, and agricultural areas to the southwest; Poznań and Łódź are surrounded by forests and agricultural areas; Lublin is surrounded by agricultural areas; Rzeszów lies between forests to the north and agricultural areas to the south, east, and west; Kraków and Sosnowiec are distinguished by a larger proportion of urban areas.

Poland has a temperate continental climate. The effects of Atlantic masses of air and the proximity of the Baltic Sea are felt in Gdańsk and Szczecin. Poznań, Łódź, Sosnowiec, and Kraków are located in a transition zone between oceanic and continental air masses. The climate of Rzeszów and Lublin is influenced by continental air masses. In addition, Kraków, Sosnowiec, and Rzeszów lie near the Carpathian Mountains, which affect their climate (Blazejczyk 2006). A two-sample Kolmogorov-Smirnov test yielded no significant differences ($D = 0.08$,

Fig. 1 The location of the sites used for the study of pollen concentration levels prediction

Table 1 Characteristics of the study sites: latitude, longitude, and altitude of the aerobiological monitoring sites; population of the cities; and mean temperatures recorded at local meteorological stations

City	λ (DD)	ϕ (DD)	Altitude (a.s.l.)	Population (in thousands)	Mean temperature (°C) (1983–2012)				
					Annual	January	February	March	April
Gdańsk	18.6131	54.3856	10	460	7.42	−1.40	−1.40	1.78	6.66
Kraków	19.9559	50.0637	212	758	8.47	−1.90	−1.11	3.12	8.94
Lublin	22.5402	51.2437	198	348	7.81	−2.66	−2.22	1.93	8.31
Łódź	19.4748	51.7715	216	719	8.43	−1.62	−1.05	2.80	8.76
Poznań	16.9243	52.4671	91	551	8.96	−0.62	−0.18	3.59	9.08
Rzeszów	22.0160	50.0293	209	182	8.46	−2.10	−1.28	2.87	8.79
Sosnowiec	19.1389	50.2972	252	214	8.47	−1.50	−0.81	3.08	8.72
Szczecin	14.5478	53.4395	30	409	8.96	0.18	0.57	3.74	8.60

p value $= 0.17$) between the daily temperatures for the six years of the study (2003–2005, 2009–2011) and a 30-year time series of measurements (1983–2012).

2.2 Aerobiological data

Daily average pollen counts were measured by a volumetric spore trap of the Hirst design (Hirst 1952), according to the recommendations of the European Aerobiology Society's Working Group on Quality Control Galán et al. (2014). Traps were located 12 m above ground level (Rzeszów) or higher. Two different pollen counting methods were used. Pollen grains were counted along 12 vertical transects using the methods outlined by Stach (2000) (Gdańsk and Rzeszów), or along four horizontal transects using the method recommended by the Spanish Aerobiology Network (Kraków, Lublin, Łódź, Poznań, Sosnowiec, Szczecin) (Galán et al. 2007). Cariñanos and Emberlin (2000) reported that both methods follow similar trends and provide close approximations to the pollen count from the entire slide. The pollen concentration was expressed as the number of grains/m^3 of air per 24 h (Comtois 1998).

2.3 Dataset creation

All the calculations were carried out using R software packages (R Core Team 2014; Kuhn 2014; Liaw and Wiener 2002). The pollen season limits of *Corylus*, *Alnus*, and *Betula* at each location and for each year were calculated using the 90 % method (Nilsson and Persson 1981). In this method, a season starts when 5 % of the total catch has been achieved and ends

when 95 % has been reached. For each taxon, the earliest day of pollen season start and the latest day of pollen season end based on all of the data were used as the temporal scope.

The aim of this work was to forecast the pollen count level of allergenic risk. For each taxon, two levels of concentration were distinguished: low and high. The ranges of concentration level values were different for each taxon. The concentration levels were as follows: *Corylus*, low 0–35 grains/m^3 and high >35 grains/m^3; *Alnus*, low 0–45 grains/m^3 and high >45 grains/m^3; and *Betula*, low 0–20 grains/m^3 and high >20 grains/m^3 (Table 2). Threshold values were based on first symptom values for patients allergic to these taxa (Rapiejko et al. 2007).

2.4 Statistical modeling

By using a stratified random split to divide the datasets, the distribution of the outcome in the training and test sets was preserved. Two subsets were created:

- Training set, used for training a model and choosing its optimal parameters (2/3 of cases)
- Test set, used only to evaluate the model on data not present during previous stages (1/3 of cases)

Most of the machine learning algorithms expect equal instances of each class. Thus, the imbalance between classes can have a significant impact on the quality of the model. The dataset was slightly imbalanced in the case of *Betula*, and highly imbalanced for *Corylus* and *Alnus* (Table 2). Up-sampling was used to reduce this class imbalance. This technique imputes additional data to improve balance across the classes. Training

Table 2 Absolute and relative number of days with given pollen concentration levels for the individual taxa and location

Taxon	City	Low	High
Corylus	Gdańsk	432 (98.63 %)	6 (1.37 %)
Corylus	Kraków	404 (92.24 %)	34 (7.76 %)
Corylus	Lublin	395 (90.18 %)	43 (9.82 %)
Corylus	Łódź	421 (96.12 %)	17 (3.88 %)
Corylus	Poznań	420 (95.89 %)	18 (4.11 %)
Corylus	Rzeszów	413 (94.29 %)	25 (5.71 %)
Corylus	Sosnowiec	406 (92.69 %)	32 (7.31 %)
Corylus	Szczecin	406 (92.69 %)	32 (7.31 %)
Alnus	Gdańsk	344 (89.58 %)	40 (10.42 %)
Alnus	Kraków	334 (86.98 %)	50 (13.02 %)
Alnus	Lublin	310 (80.73 %)	74 (19.27 %)
Alnus	Łódź	332 (86.46 %)	52 (13.54 %)
Alnus	Poznań	308 (80.21 %)	76 (19.79 %)
Alnus	Rzeszów	332 (86.46 %)	52 (13.54 %)
Alnus	Sosnowiec	333 (86.72 %)	51 (13.28 %)
Alnus	Szczecin	292 (76.04 %)	92 (23.96 %)
Betula	Gdańsk	111 (50 %)	111 (50 %)
Betula	Kraków	80 (36.04 %)	142 (63.96 %)
Betula	Lublin	88 (39.64 %)	134 (60.36 %)
Betula	Łódź	96 (43.24 %)	126 (56.76 %)
Betula	Poznań	71 (31.98 %)	151 (68.02 %)
Betula	Rzeszów	105 (47.3 %)	117 (52.7 %)
Betula	Sosnowiec	100 (45.05 %)	122 (54.95 %)
Betula	Szczecin	79 (35.59 %)	143 (64.41 %)

sets were sampled with replacements to create equal class distribution. All of the original training data were left intact, and additional samples were added to the minority classes with replacements. However, the test sets were not changed, since they should reflect the class imbalance. This is important to obtain reliable estimates of a model's performance (Kuhn and Johnson 2013).

Previous research in Poland showed that there is usually an increase—delayed on average by 1–3 days—in the correlation of pollen count between the pairs of monitoring sites (Nowosad et al. 2015). This is due mainly to prevailing winds from the west toward the east in this latitude zone of the Northern Hemisphere (Rossby waves) and the movement of atmospheric fronts. Levels of *Corylus*, *Alnus*, and *Betula* pollen concentration were used as the outcome

data. The independent variables were the previous 4 days' pollen counts from all of the sites:

$$
\begin{aligned}
\text{PollenConcentrationLevel}_{\text{siteA}_t} \sim\ & \text{PollenCount}_{\text{siteA}_{t-1}} \\
& + \text{PollenCount}_{\text{siteA}_{t-2}} + \text{PollenCount}_{\text{siteA}_{t-3}} \\
& + \text{PollenCount}_{\text{siteA}_{t-4}} + \text{PollenCount}_{\text{siteB}_{t-1}} + \cdots \\
& + \text{PollenCount}_{\text{siteH}_{t-4}}
\end{aligned}
\tag{1}
$$

Random forest (Breiman 2001) was used to predict the pollen concentration levels of *Corylus*, *Alnus*, and *Betula*. Preliminary studies showed that a random forest model's performance is comparable to other techniques, such as support vector machines and boosting trees. Furthermore, valuable information about random forest results could be obtained using, for example, variable importance. A random forest model has one tuning parameter: the number of randomly selected predictors to choose from at each split (m_{try}). A total of 100 iterations of the bootstrap were applied as the re-sampling scheme to select the optimal values of the model's tuning parameter. A series of models were fit to the training sets. For each model, the optimal parameter value was obtained based on specificity: the rate that days with high concentration were predicted correctly:

$$
\text{Specificity} = \frac{\text{CP}_{\text{high}}}{\text{All}_{\text{high}}}
\tag{2}
$$

where CP_{high} is the number of correctly predicted days with high concentration and All_{high} is the total number of days of high concentration.

The general effect of predictors on each model was calculated. Variable importance was estimated by looking at the increase in prediction error when data for a given variable were changed, while all the other variables remained constant (Breiman 2002; Liaw and Wiener 2002). Afterward, the variable importance was scaled to values between 0 and 100.

2.5 Evaluation of the models' performance

The final 24 models (3 taxa × 8 cities) were applied to generate predictions based on the test sets. Model predictions were then compared with the observed data in the test sets. A confusion matrix, unweighted Kappa statistic, sensitivity, specificity, and balanced accuracy were used to describe the performance of the models. The Kappa statistic is:

$$\text{Kappa} = \frac{O - E}{1 - E} \qquad (3)$$

where O is the observed accuracy and E is the accuracy expected to be achieved based on the marginal totals of the confusion matrix. The Kappa statistic ranges from -1 to 1. A value of 0 indicates no agreement between the observed and predicted classes, while a value of 1 indicates perfect agreement. Negative values rarely occur and indicate that "the prediction is in the opposite direction of the truth" (Kuhn and Johnson 2013). The sensitivity is defined as the rate that days with low concentration are predicted correctly:

$$\text{Sensitivity} = \frac{\text{CP}_{\text{low}}}{\text{All}_{\text{low}}} \qquad (4)$$

where CP_{low} is the number of correctly predicted days with low concentration and All_{low} is the total number of days with low concentration.

Balanced accuracy helps to reduce the impact of imbalanced classes on a model's evaluation. It is defined as follows:

$$\text{Balanced accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \qquad (5)$$

Additionally, Mann–Whitney U test (Mann and Whitney 1947) was used to compare the weather conditions (precipitation and temperature) between true and false predictions of low and high pollen concentration levels of *Corylus*, *Alnus*, and *Betula*.

3 Results

3.1 Model

The start of the earliest *Corylus* pollen season was on day 36 of the year, and its latest season end was on day 110 of the year. For *Alnus*, these start and end dates ranged from days 46 to 111 of the year, and for *Betula* they ranged from days 97 to 135 of the year. Data only from these periods were used for model creation and evaluation (Fig. 2).

Twenty-four final models were created. Table 3 shows a summary of model results for training sets. Most (13 of 16) *Corylus* and *Alnus* models had a specificity value equal to 1. Only in the cases of

Lublin, Poznań, and Szczecin was the *Alnus* model's specificity slightly lower: 0.99, 0.99, and 0.97, respectively. Kappa statistics were also very high for the *Alnus* and *Corylus* models, with values ranging from 0.90 (for *Alnus* in Szczecin) to 0.99 (for *Corylus* in Gdańsk and Łódź). Model performance values based on the *Betula* training datasets were lower. However, in the majority of the models, the specificity values were still very high: between 0.81 for Gdańsk and 0.91 for Kraków. Kappa values varied noticeably, from 0.61 for Gdańsk to 0.84 for Kraków.

3.2 Variable importance

Variable importance for *Corylus*, *Alnus*, and *Betula* models shared similar temporal and spatial properties. The values of pollen counts from one day before were the most important variable, while the values from 4 days before were the least. Moreover, variables from the same site were the most important in the majority of the models.

For six of eight of the *Corylus* models, the most important variable was the pollen count from one day (Gdańsk, Lublin, Sosnowiec, and Łódź) or 2 days (Poznań) before. Only in the case of models of pollen concentration levels in Kraków and Rzeszów was the most important independent variable from a different city: Sosnowiec in both cases. In *Corylus* models, the low importance of pollen concentration inputs from 4 days before was noticeable (Figs. 3, 6).

Variable importance for *Alnus* models was the least uniform. Only in half of the models (Lublin, Poznań, Rzeszów, and Szczecin) was the most important variable from the same city as the output. Also, in four of eight models (Poznań, Łódż, Kraków, Sosnowiec), the most important variable was the pollen count in Poznań from one day before. Exceptionally, the most important variable for predicting pollen concentration levels in Gdańsk was data from Sosnowiec. The low impact of variables from 4 days before was also apparent (Figs. 4, 6).

Variable importance for *Betula* models showed some regularities. In six of eight models, the most important variables were pollen concentration from a day before at the same site as the outcome. In addition, the pollen count values at the same site from 2 to 3 days before had a visible influence. The order of variable importance was different in two models. The most important variable for the *Betula* model in

Fig. 2 Pollen count of
Corylus, *Alnus*, *Betula* by
day of year for all of the
analyzed sites in years
2003–2005 and 2009–2011
on a logarithmic scale.
Vertical lines indicate the
temporal scope of analysis
for each taxon. *Horizontal
lines* separate the two pollen
concentration levels of low
and high: (35 g/m^3 for
Corylus, 45 g/m^3 for *Alnus*,
20 g/m^3 for *Betula*)
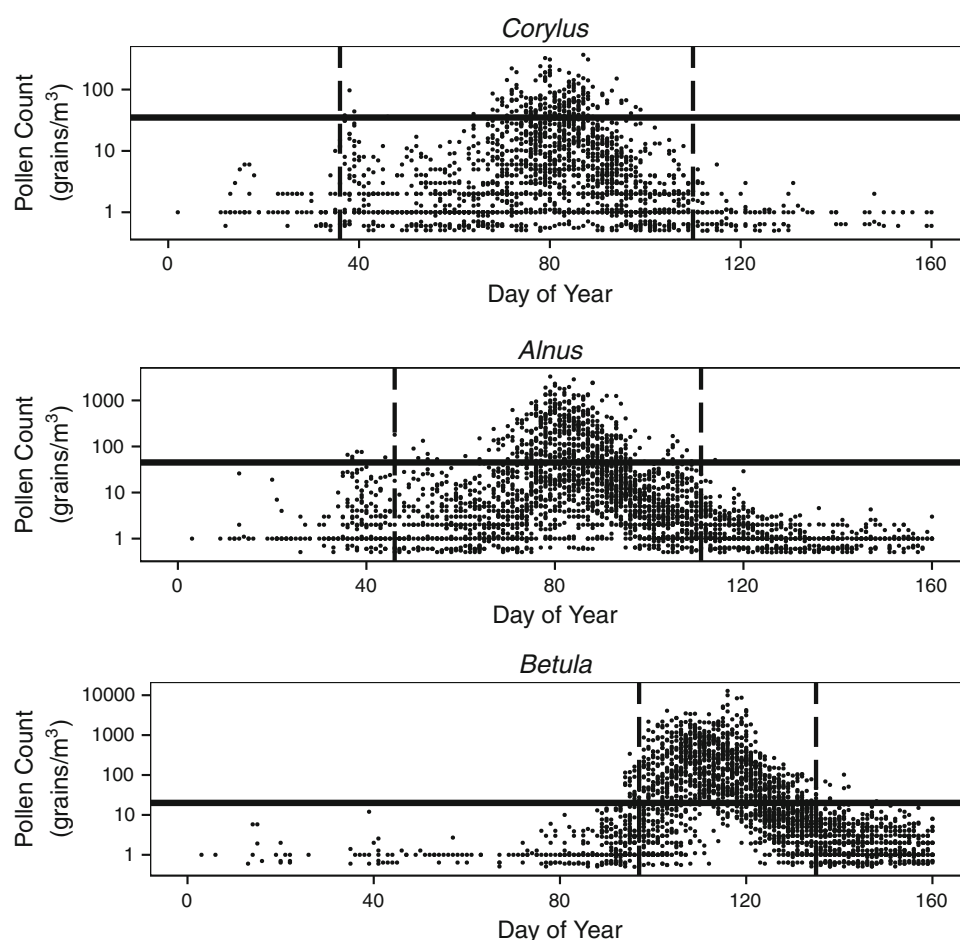
Sosnowiec was the pollen concentration value in Kraków from one day before; and the pollen concentration value in Poznań from one day before was the most important variable for the Szczecin model. Moreover, models for Gdańsk and Lublin were built based mainly on the values from the same site. In the rest of the models, many independent variables were important. Inputs from Gdańsk also had a small impact on the rest of the models. In addition, for most of the models, variables from 4 days before had either little importance or no importance at all (Figs. 5, 6).

3.3 Performance of the models

A confusion matrix, Kappa statistic, sensitivity, specificity, and balanced accuracy were used to evaluate the models on the test sets (Fig. 7; Table 4). *Corylus* models showed the lowest Kappa values. In two cases (Gdańsk and Poznań), Kappa statistics were equal to 0. Only the model of Lublin had a substantial Kappa value (0.68).; it correctly predicted 10 of 14 cases with a high pollen concentration (Fig. 7; Table 4).

The Kappa statistics also proved important for *Alnus* models, with an average value of 0.7. The minimum Kappa value was for the Gdańsk model (0.6), and the maximum was for the Łódź model (0.86). However, most of the *Alnus* models had a specificity lower than the *Betula* models. Only models for Lublin and Łódź specificity exceeded 0.8 (0.88, 0.82, respectively). The lowest specificity was found for the model for Gdańsk: 0.54. For the test set, the model of pollen concentration group of Gdańsk predicted correctly only 7 of 13 cases of high pollen concentration (Fig. 7; Table 4).

Based on th given criteria, models of *Betula* were the most reliable. Their average Kappa was 0.73, with a minimum value of 0.62 for Szczecin and a maximum of 0.81 for Kraków and Sosnowiec. Moreover, all of the specificity values for *Betula* models exceeded 0.8. In all of the *Betula* models, specificity values were higher than the Kappa statistic (Fig. 7; Table 4).

Table 3 A summary of *Corylus*, *Alnus*, and *Betula* models results for training set at each location

Taxon	City	Kappa	Accuracy	Sensitivity	Specificity
Corylus	Gdańsk	0.99	1.00	0.99	1.00
Corylus	Kraków	0.97	0.99	0.97	1.00
Corylus	Lublin	0.96	0.98	0.96	1.00
Corylus	Łódź	0.99	1.00	0.99	1.00
Corylus	Poznań	0.98	0.99	0.98	1.00
Corylus	Rzeszów	0.98	0.99	0.98	1.00
Corylus	Sosnowiec	0.97	0.98	0.97	1.00
Corylus	Szczecin	0.98	0.99	0.98	1.00
Alnus	Gdańsk	0.94	0.97	0.94	1.00
Alnus	Kraków	0.97	0.98	0.97	1.00
Alnus	Lublin	0.92	0.96	0.93	0.99
Alnus	Łódź	0.94	0.97	0.95	1.00
Alnus	Poznań	0.91	0.96	0.93	0.99
Alnus	Rzeszów	0.94	0.97	0.95	1.00
Alnus	Sosnowiec	0.96	0.98	0.96	1.00
Alnus	Szczecin	0.90	0.95	0.93	0.97
Betula	Gdańsk	0.61	0.81	0.81	0.81
Betula	Kraków	0.84	0.92	0.94	0.91
Betula	Lublin	0.82	0.91	0.92	0.90
Betula	Łódź	0.81	0.90	0.91	0.90
Betula	Poznań	0.81	0.91	0.94	0.88
Betula	Rzeszów	0.77	0.89	0.88	0.90
Betula	Sosnowiec	0.71	0.86	0.85	0.87
Betula	Szczecin	0.74	0.87	0.91	0.84

True and false predictions were compared with temperature and precipitation. The results indicated that the rainfall was connected with a false prediction of high level in Alnus (p value = 0.0018) and Betula (p value = 0.0000002) models. However, this relation was not found in Corylus (p value = 0.12) models. Additionally, true and false predictions were compared to the day-to-day changes in precipitation and temperature. The results showed that the final models were robust to changes in the precipitation; however, predictions of high level of Corylus (p value = 0.0016), Alnus (p value = 0.01), and Betula (p value = 0.001) are sensitive to the changes in temperature.

4 Discussion

One of the main goals in aerobiological models is to predict pollen concentration levels which can trigger the onset of allergic symptoms. Stepwise multiple regression (Bringfelt et al. 1982; Myszkowska 2013), additive logistic models, partially linear models (Cotos-Yáñez et al. 2004), artificial neural networks (Castellano-Méndez et al. 2005), ARIMA models (Rodriguez-Rajo et al. 2006), and stochastic gradient boosting (Hilaire et al. 2012) have been used in aerobiological studies aimed at pollen concentration prediction. Most pollen predictive modeling studies have focused on the impact of meteorological variables (such as temperature, humidity, precipitation, wind direction, and speed), on pollen season start and duration, and on pollen concentration (Bringfelt et al. 1982; Cotos-Yáñez et al. 2004; Castellano-Méndez et al. 2005; Rodriguez-Rajo et al. 2006; Hilaire et al. 2012; Myszkowska 2013). Only Castellano-Méndez et al. (2005) attempted to forecast the level of allergenic risk associated with *Betula* using previous *Betula* pollen and meteorological information. However, to the best of our knowledge, empirical predictive models have not used pollen count values from other sites before now. Previous research shows that
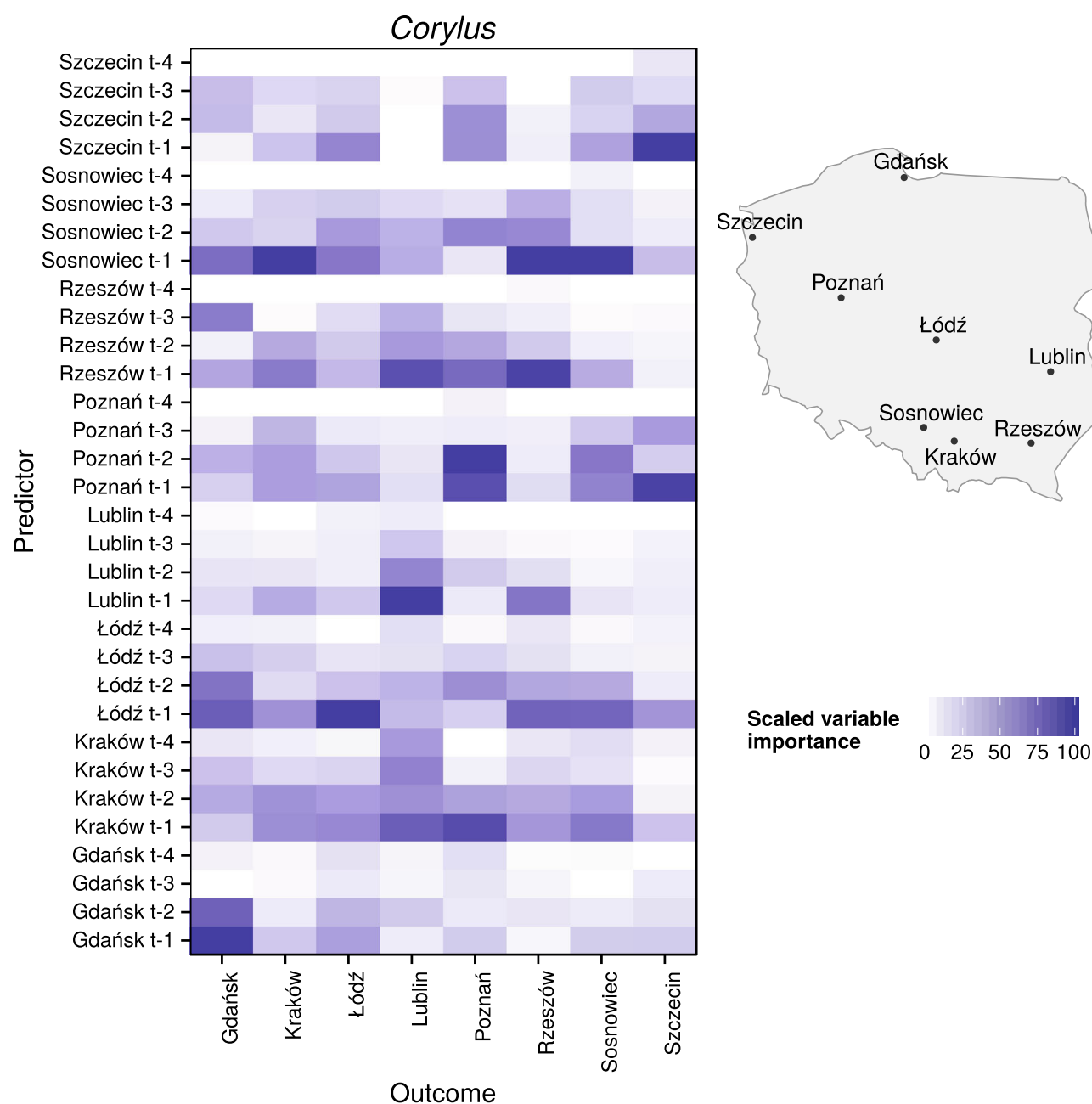
Fig. 3 Scaled variable importance of each predictor (pollen count data with 1-day lag ($t − 1$), 2-day lag ($t − 2$), 3-day lag ($t − 3$), 4-day lag ($t − 4$) at given sites) for *Corylus* in each location. For better spatial relations recognition, a schematic map with measurement sites was provided

Corylus, *Alnus*, and *Betula* pollen concentration are correlated not only in time, but also in space (Nowosad et al. 2015).

Nowadays, daily pollen concentration data are a result of manual pollen counting. Thus, information about the pollen count from the previous day is not available fast enough to be used for predicting levels of pollen concentration. However, there are many efforts to create a semiautomatic and automatic systems for counting airborne pollen (Boucher et al. 2002; Holt and Bennett 2014). As a result of these studies, it should be possible to obtain information about the pollen data from the day before quickly enough to be used by a forecast system.

In this study, random forest (Breiman 2001) was used to predict the pollen concentration level of
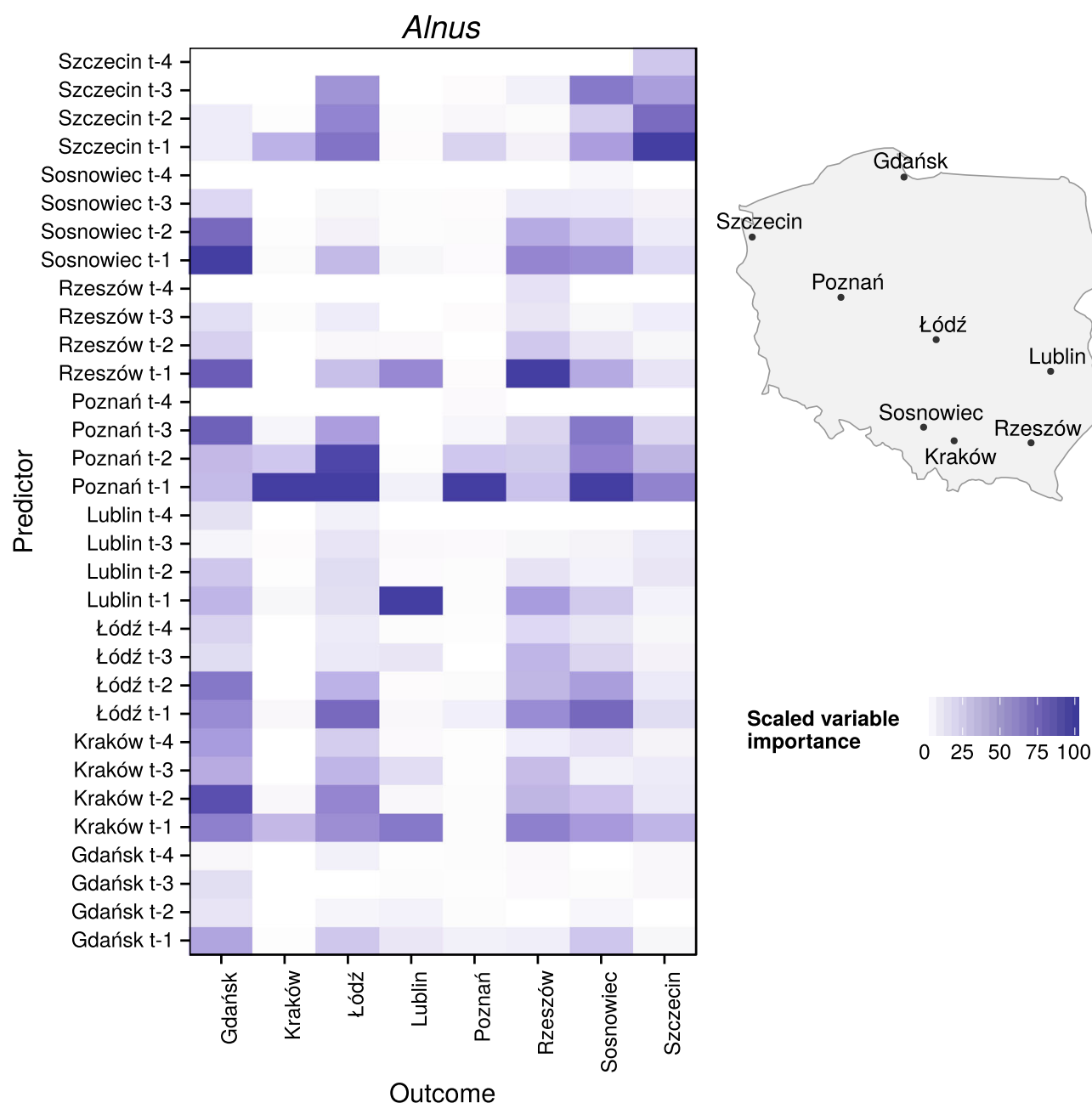
Fig. 4 Scaled variable importance of each predictor (pollen count data with 1-day lag ($t - 1$), 2-day lag ($t - 2$), 3-day lag ($t - 3$), 4-day lag ($t - 4$) at given sites) for *Alnus* in each location. For better spatial relations recognition, a schematic map with measurement sites was provided

Corylus, *Alnus*, and *Betula* using a spatiotemporal correlation of pollen count values at the given sites. The use of random forest presents a distinct advantage with respect to classical statistical methods: the ability to process complex, nonlinear relationships between predictors (Recknagel 2001). The algorithm of random forest is based on the ensemble of a large number of decision trees (Breiman 2001). Consequently, random forest has the advantage of tree-based models over artificial neural networks or support vector machines: interpretability (Geurts et al. 2009). A random forest model can be explained by visualization of decision trees or by using measures of variable importance.

The models of *Alnus* and *Betula* had, at the least, considerable values of model evaluation statistics. More than 81 % of events with high *Betula* pollen concentration could be predicted at each of the given
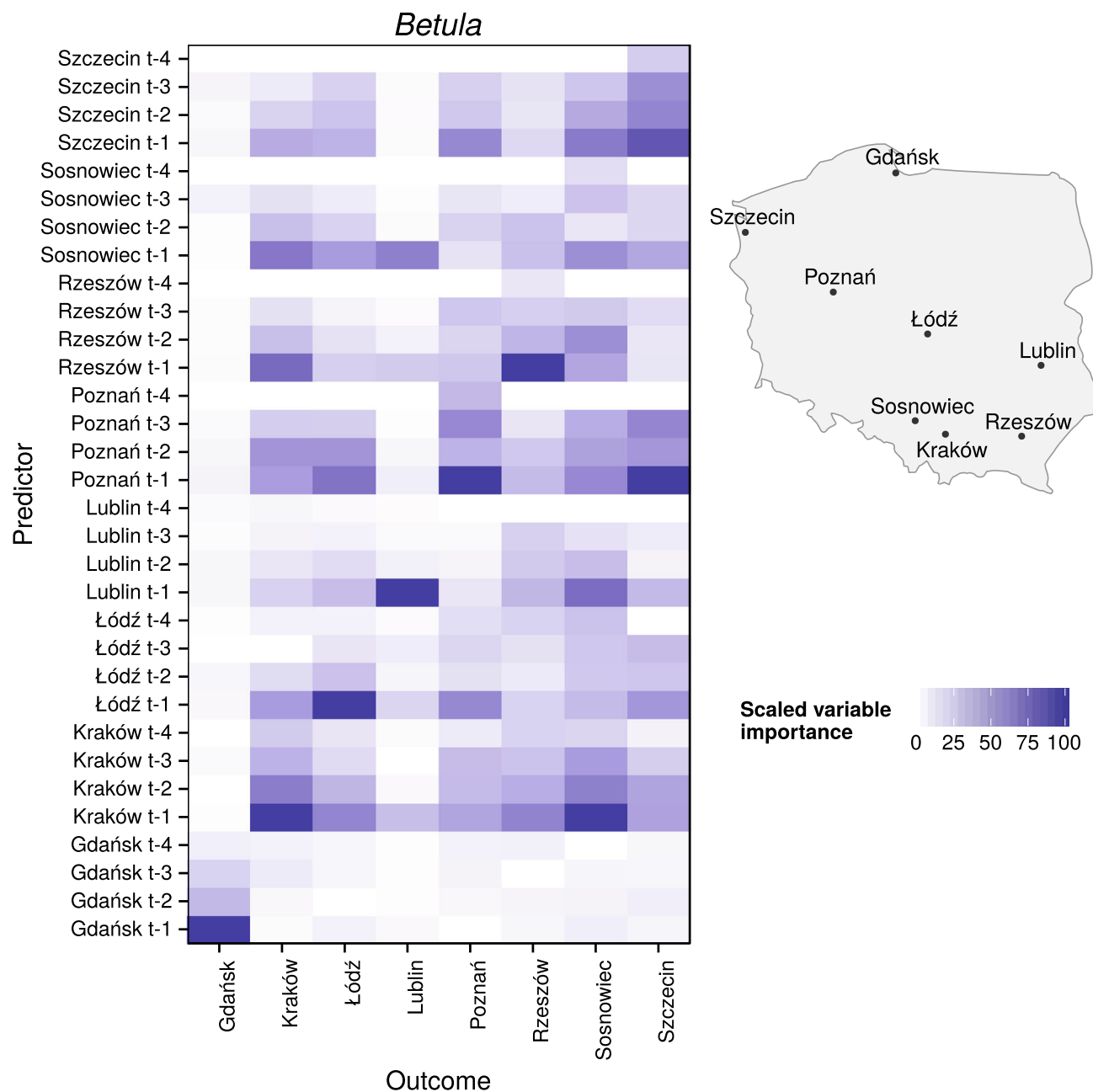
Fig. 5 Scaled variable importance of each predictor (pollen count data with 1-day lag ($t - 1$), 2-day lag ($t - 2$), 3-day lag ($t - 3$), 4-day lag ($t - 4$) at given sites) for *Betula* in each location. For better spatial relations recognition, a schematic map with measurement sites was provided

sites. The models for *Corylus* showed low values of performance statistics in most cases. There are a few possible explanations for this. Firstly, there were insufficient events with high pollen concentration levels in the training/test sets and small values of *Corylus* pollen count generally. In the years 2003–2005 and 2009–2011, high Corylus pollen counts occurred only between 6 and 43 times (26 on average),

and *Corylus* pollen concentration was lower than 19 grains/m^3 on 90 % of the analyzed days. Secondly, *Corylus* models could be highly overfit, possibly due to a relatively short time series. The start and course of *Corylus* pollen season strongly depends on the type of habitat. *Corylus* pollen season starts sooner in sunny locations and cities, where only a few degrees above zero are enough to start the pollination. On the

Fig. 6 Variations of the variables importance for each lag for *Corylus*, *Alnus*, and *Betula* models. The same site: only data from the same site. The other sites: all data except from the same site, and all data
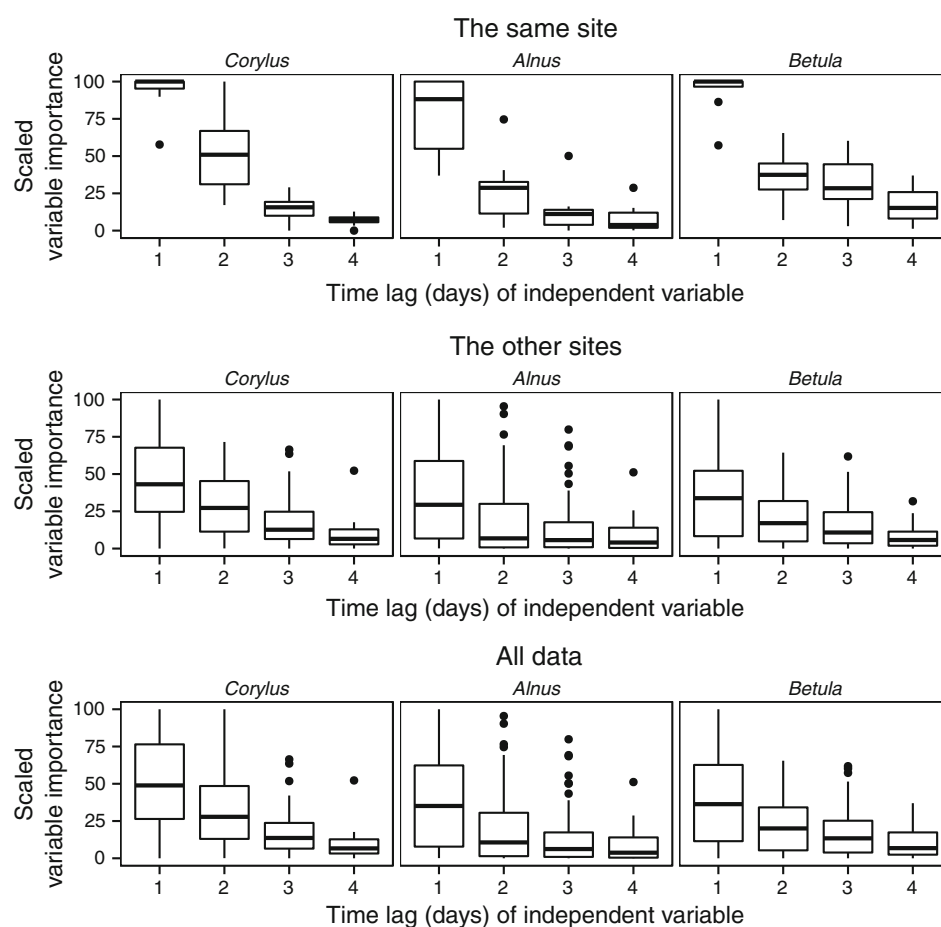
contrary, *Corylus* pollen season starts later in the sunless sites or forests. Therefore, its pollen season lasts longer than *Alnus* or *Betula* (Puc and Kasprzyk 2013). The pollen season of *Alnus* starts on average one weeks after *Corylus* in Poland (Nowosad et al. 2015). *Alnus* is also less common in the cities, and its start of flowering is less changeable between trees (Bugała 2000; Puc and Kasprzyk 2013). As a result, the relationship between *Corylus* pollen data in different cities is more complex than *Alnus* or *Betula*. Longer time series could reveal more information about *Corylus* spatiotemporal properties. Therefore, the variable importance of *Corylus* models should be interpreted with caution.

In this study, the model for *Betula* correctly forecasted between 81 and 95 % of the days with high pollen concentrations in the test set. This model performance is similar to the previous work of Castellano-Méndez et al. (2005), who used artificial neural networks for predicting whether Betula pollen concentrations exceed certain thresholds—20, 30, 70, and 80 g/m^3—using previous pollen and

meteorological information. The artificial neural networks model predicted between 83 and 100 % of over-level pollen days on the validation set (years 2000 and 2001). Thus, the models based on previous pollen counts from several sites could serve as an alternative to models based on pollen and meteorological data from a single analyzed site.

The importance of independent variables showed a clear temporal and spatial dependency. In 23 of 24 models, variables from a day before had the largest impact. Moreover, input from the same site as the output was the most important in 16 of 24 models. In *Alnus* models, a high impact of variables from Poznań on Łódź, Sosnowiec, and Kraków was observed. This relationship was also found for *Betula* models. At the same time, in *Betula* models, aside from those at Szczecin and Sosnowiec, the input values from the nearest sites had clear importance; for example, Rzeszów and Sosnowiec for Kraków; Poznań for Łódź; Szczecin and Łódź for Poznań; and Poznań for Szczecin. One possible explanation is that neighboring stations have similar weather conditions and therefore

Fig. 7 Confusion matrices and specificity/sensitivity for test sets of pollen concentration level prediction for *Corylus*, *Alnus*, and *Betula* at each location

similar pollen emission. The other explanation is influence of long-distance transport caused by the dominant westerly direction of winds in Poland. After a one-day lag, the variable importance noticeably decreases. It was found that variables from 4 days before show the least importance, with average values of 10.52, 8.58, and 8.67 % for *Corylus*, *Alnus*, and *Betula*, respectively, and the highest values did not exceed 50 %. This could be associated with the inflow

of a new air mass with different physical characteristics from the previous one. In Poland, the variability of most strings of days with a single air mass type was determined to be between 1 and 3 days (Kotas et al. 2013). It should be noted that variable importance is not a measure of causation.

Pollen production and dispersion is affected by many factors: regional flora, land use, vegetation structure, topoclimate, and weather conditions. Pollen

Table 4 A summary of *Corylus*, *Alnus*, and *Betula* models results for test set at each location

Taxon	City	Kappa	Balanced accuracy	Sensitivity	Specificity
Corylus	Gdańsk	0.00	0.50	1.00	0.00
Corylus	Kraków	0.54	0.76	0.97	0.55
Corylus	Lublin	0.68	0.84	0.97	0.71
Corylus	Łódź	0.23	0.59	0.99	0.20
Corylus	Poznań	0.00	0.50	1.00	0.00
Corylus	Rzeszów	0.49	0.79	0.96	0.62
Corylus	Sosnowiec	0.49	0.74	0.97	0.50
Corylus	Szczecin	0.41	0.65	0.99	0.30
Alnus	Gdańsk	0.60	0.76	0.98	0.54
Alnus	Kraków	0.74	0.84	0.98	0.71
Alnus	Lublin	0.71	0.89	0.90	0.88
Alnus	Łódź	0.86	0.91	0.99	0.82
Alnus	Poznań	0.72	0.86	0.95	0.76
Alnus	Rzeszów	0.66	0.83	0.96	0.71
Alnus	Sosnowiec	0.67	0.81	0.97	0.65
Alnus	Szczecin	0.65	0.84	0.90	0.77
Betula	Gdańsk	0.76	0.88	0.92	0.84
Betula	Kraków	0.81	0.92	0.96	0.88
Betula	Lublin	0.72	0.86	0.86	0.87
Betula	Łódź	0.67	0.84	0.88	0.81
Betula	Poznań	0.75	0.88	0.83	0.92
Betula	Rzeszów	0.67	0.83	0.77	0.90
Betula	Sosnowiec	0.81	0.90	0.85	0.95
Betula	Szczecin	0.62	0.81	0.77	0.85

concentration in air cannot be described as a linear effect of the impact of these factors. Despite the fact that *Alnus* and *Betula* models had substantial prediction quality, some of the events of low or high pollen level were wrongly classified. This is connected mainly with unusual events, such as no pollen or low pollen concentration at monitoring sites on the days before and high pollen count at a given site. Or it may be the opposite: low pollen concentration at a given site and high pollen count at monitoring sites in the preceding days. The occurrence of these situations can be partially explained by the influence of atmospheric conditions. The pollen concentration level could be low during the rainfall and high on the next day with a dry weather. The rapid day-to-day temperature changes also can be the cause of the models errors. Additionally, the occurrence of wrongly classified cases could probably be explained by other factors, such as random local events or changes in scale smaller than those analyzed.

Gdańsk distinguished itself from the other sites. Days with high pollen concentration of *Corylus*, *Alnus*, and *Betula* were less frequent there. In most of the models, independent variables from Gdańsk had little or no importance. Moreover, the prediction quality for Gdańsk was the lowest in the *Corylus* and *Alnus* models. As has been reported previously (Nowosad et al. 2015), Gdańsk has different pollen characteristics from other Polish sites. Its northern, coastal location has an impact on the local climate, and the start of the growing season is usually delayed there.

5 Conclusions

In this study, data from eight Polish monitoring sites over six years were used. The final 24 models are not necessarily the best ones in terms of prediction quality. The *Corylus* models performed poorly, which could be

a mixed result of (1) insufficient events with high pollen concentration level in the training/test set, (2) highly overfit models, and (3) fast change of pollen autocorrelation drop.

On the other hand, the study has clearly shown that it is possible to predict the occurrence of days with high pollen concentration of *Alnus* and *Betula* using past pollen count data from monitoring sites. For these taxa, random forest models offer capabilities for forecasting pollen concentration levels, with substantial accuracy. The models are an alternative to pollen concentration models based on weather conditions, and they show promise as a useful source of information on high pollen concentration levels for allergists and their patients. It would thus be worthwhile to combine two groups of independent variables—meteorological and aerobiological—from several sites to improve models for predicting pollen concentrations which exceed threshold values. An analysis of longer time periods or a denser monitoring network could also result in better model quality, especially in the case of *Corylus*.

References

Blazejczyk, K. (2006). Climate and bioclimate of Poland. In Degórski EM (Ed.). *Natural and human environment of Poland. A geographical overview* (pp. 31–48). Warsaw: Polish Academy of Sciences, Inst. of Geography and Spatial Organization Polish Geographical Society.

Boucher, A., Hidalgo, P. J., Thonnat, M., Belmonte, J., Galan, C., Bonton, P., et al. (2002). Development of a semi-automatic system for pollen recognition. *Aerobiologia*, *18*(3), 195–201. doi:10.1023/A:1021322813565.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi:10.1023/A:1010933404324.

Breiman, L. (2002). Manual on setting up, using, and understanding random forests v3.1.

Bremer, B., Bremer, K., & Chase, M. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, *161*(2), 105–121.

Bringfelt, B., Engström, I., & Nilsson, S. (1982). An evaluation of some models to predict airborne pollen concentration from meteorological conditions in Stockholm, Sweden. *Grana*, *21*(1), 59–64. doi:10.1080/00173138209427680.

Bugała, W. (2000). Drzewa i krzewy (Trees and shrubbery). Państwowe Wydawnictwo Rolnicze i Leśne PWRiL (in Polish).

Cariñanos, P., & Emberlin, J. (2000). Comparison of two pollen counting methods of slides from a Hirst type volumetric trap. *Aerobiologia*, *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. *International Journal of Biometeorology*, *49*(5), 310–6. doi:10.1007/s00484-004-0247-x.

Comtois, P. (1998). Statistical analysis of aerobiological data. In P. Mandrioli, P. Comtois, & V. Levizzani (Eds.), *Methods in aerobiology*. Bologna: Pitagora Editrice.

Cotos-Yáñez, T. R., Rodríguez-Rajo, F. J., & Jato, M. V. (2004). Short-term prediction of Betula airborne pollen concentration in Vigo (NW Spain) using logistic additive models and partially linear models. *International Journal of Biometeorology*, *48*(4), 179–85. doi:10.1007/s00484-004-0203-9.

Dmochowska, H. (Ed.). (2013). *Concise Statistical Yearbook of Poland 2013*. Poland: Statistical Publishing Establishment.

Galán, C., Cariñanos, P., Alcázar, P., & Dominguez-Vilches, E. (2007). *Spanish aerobiology network (REA) management and quality manual*. Servicio de Publicaciones Universidad de Córdoba. Tech. rep., ISBN 978-84-690-6353-8.

Galán, C., Smith, M., Thibaudon, M., Frenguelli, G., Oteros, J., Gehrig, R., et al. (2014). Pollen monitoring: Minimum requirements and reproducibility of analysis. *Aerobiologia*, *30*(4), 385–395. doi:10.1007/s10453-014-9335-5.

Geurts, P., Irrthum, A., & Wehenkel, L. (2009). Supervised learning with decision tree-based methods in computational and systems biology. *Molecular Biosystems*, *5*, 1593–1605. doi:10.1039/b907946g.

Heinzerling, L. M., Burbach, G. J., Edenharter, G., Bachert, C., Bindslev-Jensen, C., Bonini, S., et al. (2009). GA(2)LEN skin test study I: GA(2)LEN harmonization of skin prick testing: Novel sensitization patterns for inhalant allergens in Europe. *Allergy*, *64*(10), 1498–1506. doi:10.1111/j.1398-9995.2009.02093.x.

Hilaire, D., Rotach, M. M. W., & Clot, B. (2012). Building models for daily pollen concentrations. *Aerobiologia*, *28*(4), 499–513. doi:10.1007/s10453-012-9252-4.

Hirst, J. M. (1952). An automatic volumetric spore trap. *Annals of Applied Biology*, *39*(2), 257–265. doi:10.1111/j.1744-7348.1952.tb00904.x.

Holt, K. A., & Bennett, K. D. (2014). Principles and methods for automated palynology. *New Phytologist*, *203*(3), 735–742. doi:10.1111/nph.12848.

Kornas, J., & Medwecka-Kornas, A. (2002). *Geografia roślin (Plant geography)*. Warszawa: Wydawnictwo Naukowe PWN. (in Polish).

Kotas, P., Twardosz, R., & Nieckarz, Z. (2013). Variability of air mass occurrence in southern Poland (1951–2010). *Theoretical and Applied Climatology*, *114*(3–4), 615–623. doi:10.1007/s00704-013-0861-9.

Kuhn, M. (2014). caret: Classification and regression training. http://cran.r-project.org/package=caret

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York: Springer.

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

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, *2*(3), 18–22.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. doi:10.1214/aoms/1177730491.

Myszkowska, D. (2013). Prediction of the birch pollen season characteristics in Cracow, Poland using an 18-year data series. *Aerobiologia*, *29*(1), 31–44. doi:10.1007/s10453-012-9260-4.

Nilsson, S., & Persson, S. (1981). Tree pollen spectra in the Stockholm region (Sweden), 1973–1980. *Grana*, *20*(3), 179–182. doi:10.1080/00173138109427661.

Nowosad, J., Stach, A., Kasprzyk, I., Latałowa, M., Puc, M., Myszkowska, D., et al. (2015). Temporal and spatiotemporal autocorrelation of daily concentrations of Alnus, Betula, and Corylus pollen in Poland. *Aerobiologia*, *31*, 159–177. doi:10.1007/s10453-014-9354-2.

Puc, M. (2012). Artificial neural network model of the relationship between Betula pollen and meteorological factors in Szczecin (Poland). *International Journal of Biometeorology*, *56*(2), 395–401. doi:10.1007/s00484-011-0446-1.

Puc, M., & Kasprzyk, I. (2013). The patterns of Corylus and Alnus pollen seasons and pollination periods in two Polish cities located in different climatic regions. *Aerobiologia*, *29*, 495–511. doi:10.1007/s10453-013-9299-x.

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.r-project.org/

Rapiejko, P., Stankiewicz, W., Szczygielski, K., & Jurkiewicz, D. (2007). Progowe stężenie pyłku roślin niezbędne do wywołania objawów alergicznych (Threshold pollen count necessary to evoke allergic symptoms). *Otolaryngologia Polska*, *61*(4), 591–594. doi:10.1016/S0030-6657(07)70491-2.

Recknagel, F. (2001). Applications of machine learning to ecological modelling. *Ecological Modelling*, *146*, 303–310. doi:10.1016/S0304-3800(01)00316-7.

Rodriguez-Rajo, F., Rodríguez-Rajo, F. J., Valencia-Barrera, R. M., Vega-Maray, A. M., Suárez, F. J., Fernández-González, D., et al. (2006). Prediction of airborne Alnus pollen concentration by using ARIMA models. *Annals of Agricultural and Environmental Medicine*, *13*(1), 25–32.

Sofiev, M., Siljamo, P., Ranta, H., Linkosalo, T., Jaeger, S., Rasmussen, A., et al. (2013). A numerical model of birch pollen emission and dispersion in the atmosphere. Description of the emission module. *International Journal of Biometeorology*, *57*(1), 45–58. doi:10.1007/s00484-012-0532-z.

Stach, A. (2000). Variation in pollen concentration of the most allergenic taxa in Poznań (Poland), 1995–1996. *Aerobiologia*, *16*(1), 63–68. doi:10.1023/A:1007603019797.

Valenta, R., Breiteneder, H., Pettenburger, K., Breitenbach, M., Rumpold, H., Kraft, D., et al. (1991). Homology of the major birch-pollen allergen, I, with the major pollen allergens of alder, hazel, and hornbeam at the nucleic acid level as determined by cross-hybridization. *Journal of Allergy and Clinical Immunology*, *87*(3), 677–682. doi:10.1016/0091-6749(91)90388-5.

Viander, M., & Koivikko, A. (1978). The seasonal symptoms of hyposensitized and untreated hay fever patients in relation to birch pollen counts: correlations with nasal sensitivity, prick tests and RAST. *Clinical & Experimental Allergy*, *8*(4), 387–396.

Vogel, H., Pauling, A., & Vogel, B. (2008). Numerical simulation of birch pollen dispersion with an operational weather forecast system. *International Journal of Biometeorology*, *52*(8), 805–814. doi:10.1007/s00484-008-0174-3.

Chapter 5

Spatiotemporal models for predicting high pollen concentration level of *Corylus*, *Alnus*, and *Betula*

CrossMark
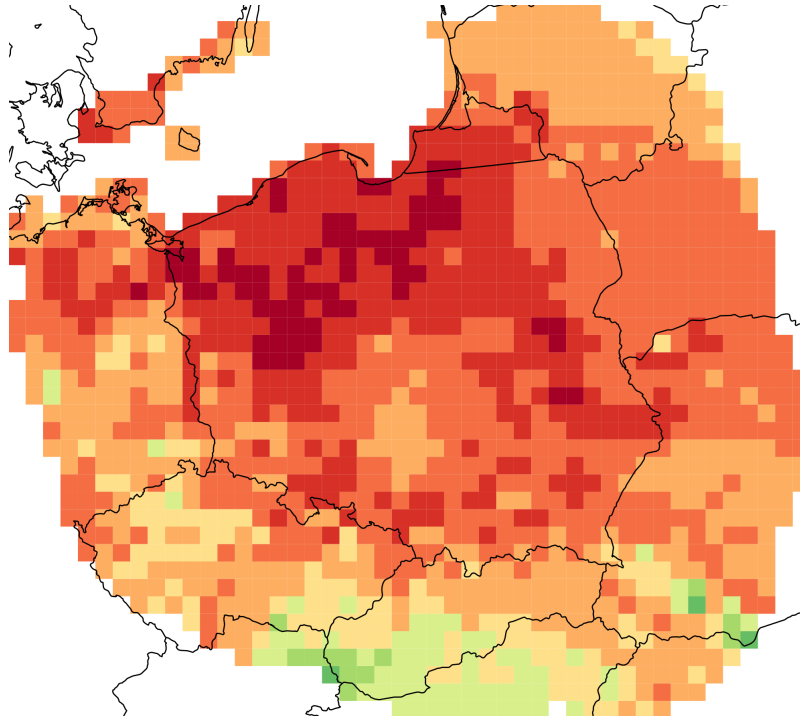
Spatiotemporal models for predicting high pollen concentration level of *Corylus*, *Alnus*, and *Betula*

Jakub Nowosad[1]

Abstract *Corylus*, *Alnus*, and *Betula* trees are among the most important sources of allergic pollen in the temperate zone of the Northern Hemisphere and have a large impact on the quality of life and productivity of allergy sufferers. Therefore, it is important to predict high pollen concentrations, both in time and space. The aim of this study was to create and evaluate spatiotemporal models for predicting high *Corylus*, *Alnus*, and *Betula* pollen concentration levels, based on gridded meteorological data. Aerobiological monitoring was carried out in 11 cities in Poland and gathered, depending on the site, between 2 and 16 years of measurements. According to the first allergy symptoms during exposure, a high pollen count level was established for each taxon. An optimizing probability threshold technique was used for mitigation of the problem of imbalance in the pollen concentration levels. For each taxon, the model was built using a random forest method. The study revealed the possibility of moderately reliable prediction of *Corylus* and highly reliable prediction of *Alnus* and *Betula* high pollen concentration levels, using preprocessed gridded meteorological data. Cumulative growing degree days and potential evaporation proved to be two of the most important predictor variables in the models. The final models predicted not only for single locations but also for continuous areas. Furthermore, the proposed modeling framework could be used to predict high pollen concentrations of *Corylus*, *Alnus*, *Betula*, and other taxa, and in other countries.

Introduction

Corylus L. (hazel), *Alnus* Mill. (alder), and *Betula* L. (birch) are considered to be among the most important sources of allergic pollen in the temperate zone of the Northern Hemisphere (D'Amato et al. 2007). According to Heinzerling et al. (2009), approximately 21–24 % of Europeans are sensitized to tree pollen from the Betulaceae family. These rates in Poland are 22.3, 22.8, and 27.7 %, respectively, for *Corylus*, *Alnus*, and *Betula* (Heinzerling et al. 2009). There are also high levels of cross-reactivity between *Corylus*, *Alnus*, and *Betula* (Ebner et al. 1995). As a consequence, *Corylus* and *Alnus* pollination can lead to more marked clinical symptoms during a *Betula* pollen season (D'Amato et al. 2007).

Pollen concentration in the air is the resultant of many factors of different temporal and spatial variability. The spatial distribution of the taxa and phytosociological and habitat relationships mainly affect the temporal variability and intensity of pollen seasons. Moreover, meteorological factors have an impact not only on the production and release but also on the dispersal of tree pollen grains. Previous studies found a relationship between the temperature in the preceding year and the annual pollen sum (Latałowa et al. 2002; Rasmussen 2002). The influence of air temperature on pollen concentration has often been reported (Rodríguez-Rajo et al. 2004; Puc 2007, 2012; Kizilpinar et al. 2011). The impact of other meteorological parameters, such as precipitation, wind speed, and humidity, has also been reported (Latałowa et al. 2002; Puc 2007, 2012). In

✉ Jakub Nowosad
  nowosad@amu.edu.pl

[1] Institute of Geoecology and Geoinformation, Adam Mickiewicz University, Dzięgielowa 27, 61-680 Poznań, Poland

🖉 Springer

addition, recent studies have shown that the temporal variations in *Corylus*, *Alnus*, and *Betula* pollen counts are related to three groups of factors. The temporal span of these factors are (i) daily, (ii) approximately 3.5 days, and (iii) more than 15 days (Nowosad et al. 2015).

Spatial analyses in aerobiology mainly involve the following: the comparison between two or more different localizations (Stach et al. 2008; Puc and Kasprzyk 2013; Sauliene et al. 2014); the description of spatial variation of pollen season properties or pollen concentrations (Emberlin et al. 2002; Rieux et al. 2008; Myszkowska et al. 2010; Nowosad et al. 2015); or the investigation of pollen transportation using back trajectories (Skjoth et al. 2008, 2009; Veriankaite et al. 2009; Rojo and Pérez-Badia 2015). There have been only a few studies in which spatial models of *Betula* pollen count were built. Vogel et al. (2008) included the parametrisation of the emissions of *Betula* pollen into a non-hydrostatic mesoscale model. Sofiev et al. (2013) used the SILAM dispersion model to create a *Betula* pollen emission model. According to the author's knowledge, spatial models of *Corylus* and *Alnus* pollen concentration have not been reported.

The main aim of this study was to develop spatiotemporal predictive models of *Corylus*, *Alnus*, and *Betula* pollen concentration levels, using preprocessed gridded meteorological data. Based on the final models, it is possible to predict pollen concentration levels, not only in aerobiological monitoring sites but also at unsampled locations.

Materials and methods

The development of spatiotemporal predictive models of *Corylus*, *Alnus*, and *Betula* pollen concentration levels was a main goal of this study. For each taxon, the workflow was as follows. Aerobiological data were split into a training set and two test sets, whilst gridded meteorological data were preprocessed. Using data from the training set, optimal model parameters were estimated. Final model performance was obtained by comparison of model prediction with true pollen concentration levels from both of the test sets. Afterwards, the final model and processed gridded meteorological data were used to create spatiotemporal predictions for every available day for each grid cell in the study area (Fig. 1). All the calculations were carried out using R (R Core Team 2014) and R packages (Liaw and Wiener 2002; Pebesma and Bivand 2005; Wickham 2009; Kuhn 2015). The workflow is described in detail in the subsections below.
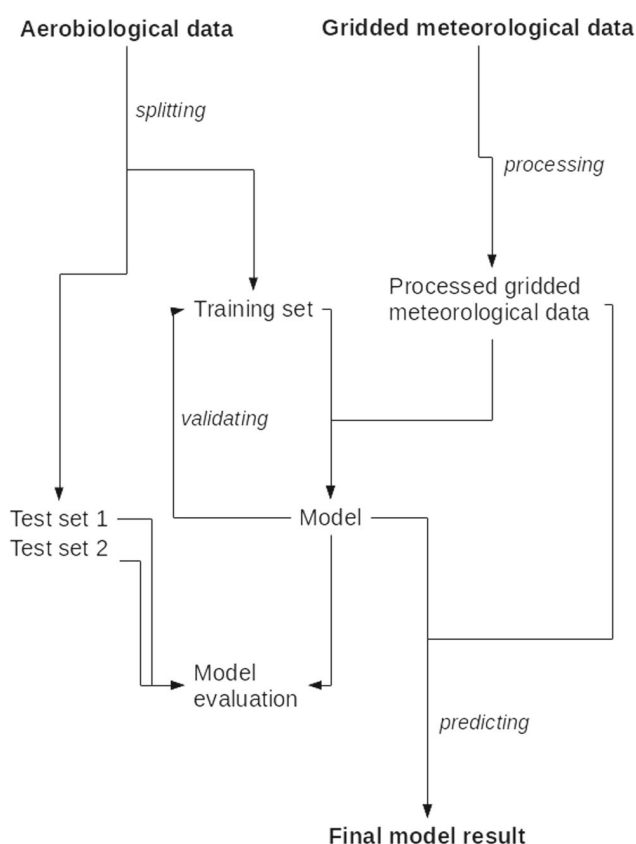

Fig. 1 A flowchart of the processes for the predictive mapping of the pollen concentration levels

Aerobiological data

The monitoring of the concentrations of *Corylus*, *Alnus*, and *Betula* pollen was conducted in 11 cities in Poland: Bydgoszcz, Gdańsk, Kraków, Lublin, Łódź, Olsztyn, Poznań, Rzeszów, Siedlce, Sosnowiec, and Szczecin. The aerobiological studies covered between 2 (Siedlce) and 16 (Poznań) years of measurements (Table 1).Daily pollen concentrations were measured using the recommendations of the European Aerobiology Society's Working Group on Quality Control (Galán et al. 2014).

The *Corylus*, *Alnus*, and *Betula* pollen season limits at each location, and each year was calculated using the 99 % method. According to this method, the onset of the pollen season was determined when 0.5 % of the total annual pollen count was noted, whereas the end of the season was determined when 99.5 % of pollen grains were recorded. The maximum range of the *Corylus* pollen season, using all the data, was between 6 and 150 days of the year; for *Alnus*,

Table 1 Population and area of the cities with the aerobiological monitoring sites; longitude, latitude, altitude, and the studied years of the aerobiological monitoring sites

Site	Population (in thousands)	Area (km^2)	λ (DD)	ϕ (DD)	Altitude (a.s.l.)	Studied years
Bydgoszcz	361	176	18.13	53.14	51	2009–2011
Gdańsk	460	262	18.61	54.39	12	1998–2005, 2009–2011
Kraków	758	327	19.96	50.06	207	1998–2005, 2009–2011
Łódź	719	293	19.47	51.77	216	2003–2005, 2009–2011
Lublin	348	147	22.54	51.24	194	2001–2005, 2009–2011
Olsztyn	175	88	20.49	53.78	132	2009–2011
Poznań	551	262	16.92	52.47	93	1996–2011
Rzeszów	182	116	22.02	50.03	201	1997–2005, 2009–2011
Siedlce	76	32	22.31	52.18	147	2010–2011
Sosnowiec	214	91	19.14	50.30	253	2001–2011
Szczecin	409	301	14.55	53.44	28	2002–2011

the number of days ranged from 14 to 145 days; for *Betula*, from 35 to 164 days. Models were built and evaluated using the data from these periods (Fig. 2).

Based on first symptom values for patients allergic to each taxon, two levels of concentration (low and high) were distinguished (Rapiejko et al. 2007). The limits were set at 35 grains/m^3 for *Corylus*, 45 grains/m^3 for *Alnus*, and 20 grains/m^3 for *Betula* (Fig. 2).

Grid data

AGRI4CAST Interpolated Meteorological Data (Baruth et al. 2007) were used as the main input data. The AGRI4CAST database is a collection of daily meteorological parameters from weather stations interpolated to a 25 × 25 km grid and contains data from 1975 to 2014. For the purpose of this study, grid data were restricted to the area of Poland and a zone of 200 km around Polish borders. The buffer value was based on the longest distance between the nearest aerobiological sites (Szczecin and Poznań): approximately 200 km.

Data split

The models were designed to predict pollen concentration levels (i) in the aerobiological monitoring sites, and (ii) in sites without aerobiological monitoring. For this purpose, data were split into the following three sets:

– Training set, which contained 2/3 of the data from eight cities in Poland (Gdańsk, Kraków, Lublin, Olsztyn, Poznań, Rzeszów, Siedlce, and Szczecin). The data

were split randomly based on the dates available in this study.
– First test set, which contained the remaining 1/3 of the data from the same eight cities (Gdańsk, Kraków, Lublin, Olsztyn, Poznań, Rzeszów, Siedlce, and Szczecin).
– Second test set, which contained data from Bydgoszcz, Łódź, and Sosnowiec.

Predictor variables

The pollen concentration level at each site and on each day from the training set was used as an outcome variable. The daily meteorological variables were preprocessed. Afterwards, preprocessed meteorological variables from the grid cells corresponding to the location of aerobiological sites were selected as predictor variables (Table 2):

– The average monthly temperatures for each month over the previous year for each site
– Four- and 16-day averages, calculated for each of the meteorological parameters. The temporal span of these factors was based on a recent study which showed that the temporal variations in *Corylus*, *Alnus*, and *Betula* pollen counts are related to factors that change (i) diurnally, (ii) approximately every 3.5 days, and (iii) in more than 15 days (Nowosad et al. 2015). These values were then lagged by 1 day
– Cumulated growing degree days (GDD), lagged by 1 day
– Longitude, latitude, and altitude of grid cell

Fig. 2 Pollen count of *Corylus*, *Alnus*, and *Betula* for all of the analyzed sites and all years, on a logarithmic scale. *Vertical lines* indicate the temporal scope of analysis for each taxon. *Horizontal lines* separate the two pollen concentration levels of low and high: 35 grains/m^3 for *Corylus*, 45 grains/m^3 for *Alnus*, and 20 grains/m^3 for *Betula*
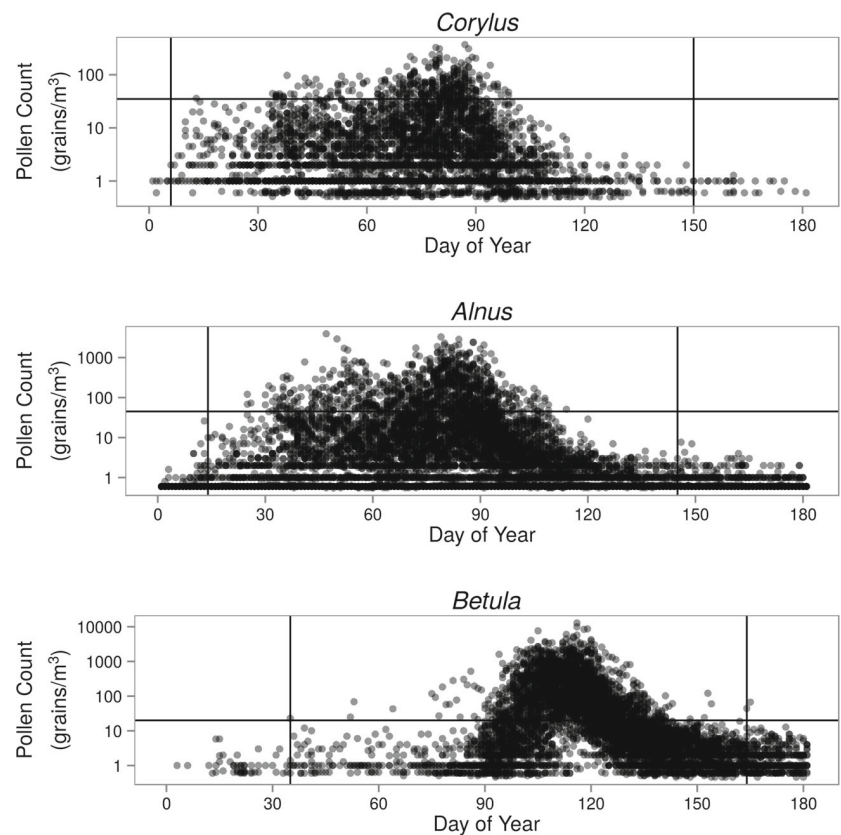
For each day and each grid cell, growing degree days (GDD) were calculated as follows:

$$\text{Daily GDD} = \frac{T_{\max} + T_{\min}}{2} - T_{\text{base}} \qquad (1)$$

where $T_{\max}$ is the daily maximum temperature, $T_{\min}$ is the daily minimum temperature, and T_{base} is the base temperature.

GDDs were accumulated by adding the number of degree days that accumulated each day from January 1. The base temperature was designated as 5 °C, which is the standard threshold temperature for growth in temperate species (Dahl et al. 2013). If the daily maximum temperature is not higher than the base temperature, then no degree days accumulate.

Model development

Random forest (Breiman 2001) was used to spatiotemporally predict the pollen level of *Corylus*, *Alnus*, and *Betula*. For classification tasks, it is an ensemble of unpruned classification trees. The prediction is made by aggregating the prediction of the ensemble. The random forest algorithm uses two parameters: ntree (the number of trees) and mtry (the number of input variables randomly chosen at each split). In this study, ntree was set to 500, while optimal

values of mtry were obtained by using 100 repetitions of ten-fold cross-validation on the training set.

The random forest algorithm focuses on overall accuracy and, consequently, does not work well for imbalanced data. The *Corylus*, *Alnus*, and *Betula* pollen concentration levels were highly imbalanced. In the period analyzed, the proportion between high and low levels was 330 to 13,182 for *Corylus*, 966 to 11,348 for *Alnus*, and 2104 to 9933 for *Betula*. In this study, an optimizing probability threshold technique was applied (Kuhn and Johnson 2013). This approach determines alternative cutoffs for the predicted probabilities. Using resampling, 20 different threshold values were tried on the training sets. Optimal threshold values were obtained by minimizing the distance between obtained sensitivity (Sens), specificity (Spec), positive predictive value (Ppv), negative predictive value (Npv), and the best possible performance (Fig. 3). In all cases, the best possible performance was equal to 1.

$$\text{Distance} = \sqrt{\begin{array}{c}(1 - \text{Sens})^2 + (1 - \text{Spec})^2 \\ + (1 - \text{Ppv})^2 + (1 - \text{Npv})^2\end{array}} \qquad (2)$$

A permutation importance (mean decrease in accuracy) was used to determine input variable importance (Breiman 2001; Liaw and Wiener 2002).

Table 2 Explanation of the predictor variable abbreviations used in spatiotemporal modeling of *Corylus*, *Alnus*, and *Betula* pollen concentration levels

Abbreviation	Predictor variable name	Unit
TAVG_JANUARY_PREVYEAR	Average monthly temperature for January in the preceding year	°C
TAVG_FEBRUARY_PREVYEAR	Average monthly temperature for February in the preceding year	°C
TAVG_MARCH_PREVYEAR	Average monthly temperature for March in the preceding year	°C
TAVG_APRIL_PREVYEAR	Average monthly temperature for April in the preceding year	°C
TAVG_MAY_PREVYEAR	Average monthly temperature for May in the preceding year	°C
TAVG_JUNE_PREVYEAR	Average monthly temperature for June in the preceding year	°C
TAVG_JULY_PREVYEAR	Average monthly temperature for July in the preceding year	°C
TAVG_AUGUST_PREVYEAR	Average monthly temperature for August in the preceding year	°C
TAVG_SEPTEMBER_PREVYEAR	Average monthly temperature for September in the preceding year	°C
TAVG_OCTOBER_PREVYEAR	Average monthly temperature for October in the preceding year	°C
TAVG_NOVEMBER_PREVYEAR	Average monthly temperature for November in the preceding year	°C
TAVG_DECEMBER_PREVYEAR	Average monthly temperature for December in the preceding year	°C
TMAX_4DAYS_AVG_1DAYLAG	Average maximum temperature in preceding 4 days	°C
TMAX_16DAYS_AVG_1DAYLAG	Average maximum temperature in preceding 16 days	°C
TMIN_4DAYS_AVG_1DAYLAG	Average minimum temperature in preceding 4 days	°C
TMIN_16DAYS_AVG_1DAYLAG	Average minimum temperature in preceding 16 days	°C
VAPORPRESSURE_4DAYS_AVG_1DAYLAG	Average vapor pressure in preceding 4 days	hPa
VAPORPRESSURE_16DAYS_AVG_1DAYLAG	Average vapor pressure in preceding 16 days	hPa
WINDSPEED_4DAYS_AVG_1DAYLAG	Average wind speed in preceding 4 days	m/s
WINDSPEED_16DAYS_AVG_1DAYLAG	Average wind speed in preceding 16 days	m/s
PRECIPITATION_4DAYS_AVG_1DAYLAG	Average daily precipitation in the preceiding 4 days	mm
PRECIPITATION_16DAYS_AVG_1DAYLAG	Average daily precipitation in the preceiding 16 days	mm
EVAPORATION_4DAYS_AVG_1DAYLAG	Average potential evaporation in the preceding 4 days	mm/day
EVAPORATION_16DAYS_AVG_1DAYLAG	Average potential evaporation in the preceding 16 days	mm/day
RADIATION_4DAYS_AVG_1DAYLAG	Average total global radiation in the preceding 4 days	KJ/m^2/day
RADIATION_16DAYS_AVG_1DAYLAG	Average total global radiation in the preceding 16 days	KJ/m^2/day
GDD_1DAYLAG	Cummulated growing degree days (GDD) lagged by one day	GDD
LONGITUDE	Grid cell longitude	degrees
LATITUDE	Grid cell latitude	degrees
ALTITUDE	Average altitude of grid cell	m a.s.l.

Evaluation of the models performance

The accuracy of a model is not an appropriate measure of performance of prediction with highly imbalanced data. Instead, Kappa statistic, sensitivity, specificity, positive predictive value, and negative predictive value were used to evaluate the performance of the models:

$$\text{Kappa} = \frac{O - E}{1 - E} \qquad (3)$$

where O is the observed accuracy, and E is the accuracy expected to be achieved based on the marginal totals of the confusion matrix. The Kappa statistic values range from -1 to 1. A value of 1 indicates perfect agreement between the observed and predicted classes; a value of 0 indicates no

agreement; negative values indicate that the predicted class is the opposite of the reference class (Kuhn and Johnson 2013).

$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad (4)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \qquad (5)$$

$$\text{Positive predictive value} = \frac{TP}{TP + FP} \qquad (6)$$

$$\text{Negative predictive value} = \frac{TN}{TN + FN} \qquad (7)$$

where TP are the true positives (high levels predicted correctly), FP are false positives (high levels incorrectly

Fig. 3 Resampled values of sensitivity, specificity, positive predictive value, negative predictive value, and the numerical distance between those values for each taxon model
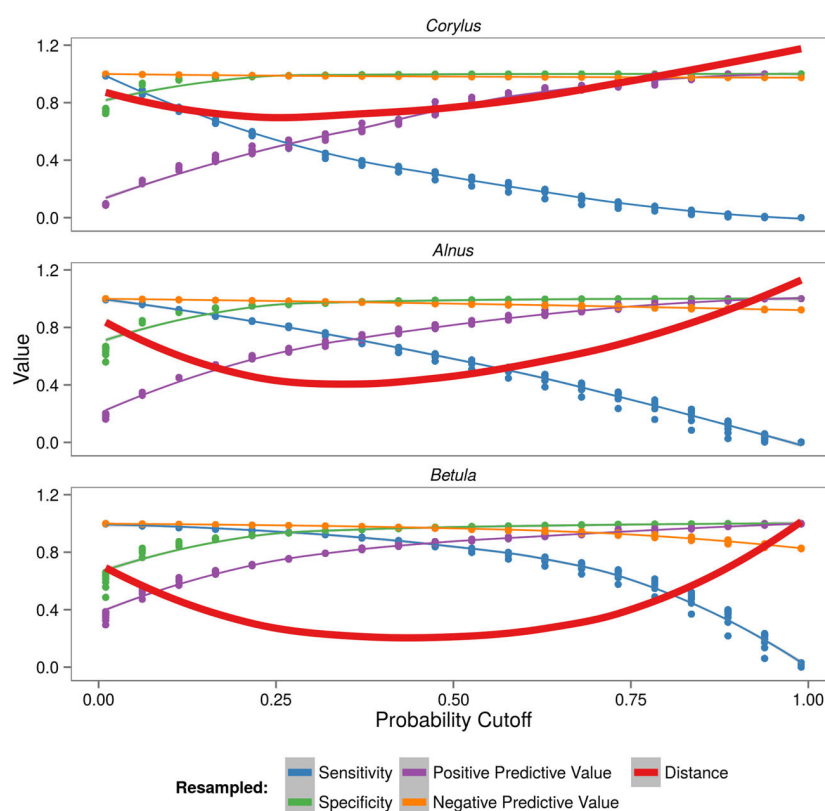
predicted), TN are true negatives (low levels correctly predicted), and FN are false negatives (low levels incorrectly predicted).

Corylus, *Alnus*, and *Betula* models were evaluated on two test sets. Firstly, the temporal modelsŠ performance was determined by comparison between true pollen concentration levels and predictions on the first test set. Secondly, models' predictions were compared with true pollen concentration levels from Bydgoszcz, Łódź, and Sosnowiec (the second test set). Data from these cities were not used for model creation. Thus, the evaluation was used to determinate spatial quality of the models.
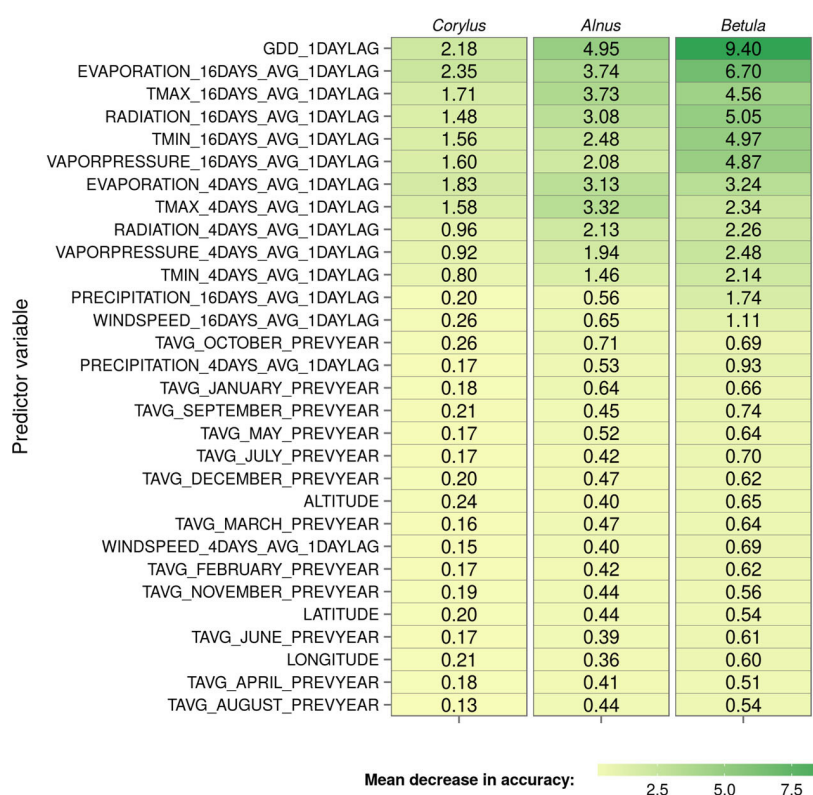
Results

Probability threshold

For each taxon, the final model had different probability thresholds. In the *Corylus* models, the probability threshold dividing low and high pollen concentration levels was optimized to 0.22. In the other taxa, the class imbalance was less severe, and thus the optimal probability threshold value was higher: 0.32 for *Alnus* and 0.42 for *Betula* (Fig. 3).

Variable importance

Using a random forest model, it is possible to examine predicted class probabilities for each variable in the dataset (Fig. 4). Figure 5 shows the relationship between the probability of high pollen concentration levels and the values of the four most important variables on the training set of the *Corylus*, *Alnus*, and *Betula* models. Cumulated growing degree days (GDD) was the most important variable in the *Alnus* and *Betula* models and the second-most important variable in the *Corylus* model. For each taxa, the highest probability of high pollen concentration level had a different range of GDD values. In the *Corylus* model, there was more than 0.5 probability of a high pollen concentration level when the GDD value was between 3 and 77. The highest probability (0.79) was connected with a GDD of 32. The range of probabilities and GDD values was slightly different in the *Alnus* model. A probability higher than 0.5 of high pollen concentration levels occurred when GDD was between 15 and 90. The peak of probability was 0.82 for a GDD value of 44. In the *Betula* model, the optimal value of GDD was between 94 and 323, with the highest peak (0.93) for a GDD of 183. Moreover, the distribution of high pollen concentration level probability in the *Betula* model

Fig. 4 Variable importance of each input variable for *Corylus*, *Alnus*, and *Betula* models. The variables are showed by the mean value of variable importance for all of the taxa in descending order

Predictor variable	Corylus	Alnus	Betula
GDD_1DAYLAG	2.18	4.95	9.40
EVAPORATION_16DAYS_AVG_1DAYLAG	2.35	3.74	6.70
TMAX_16DAYS_AVG_1DAYLAG	1.71	3.73	4.56
RADIATION_16DAYS_AVG_1DAYLAG	1.48	3.08	5.05
TMIN_16DAYS_AVG_1DAYLAG	1.56	2.48	4.97
VAPORPRESSURE_16DAYS_AVG_1DAYLAG	1.60	2.08	4.87
EVAPORATION_4DAYS_AVG_1DAYLAG	1.83	3.13	3.24
TMAX_4DAYS_AVG_1DAYLAG	1.58	3.32	2.34
RADIATION_4DAYS_AVG_1DAYLAG	0.96	2.13	2.26
VAPORPRESSURE_4DAYS_AVG_1DAYLAG	0.92	1.94	2.48
TMIN_4DAYS_AVG_1DAYLAG	0.80	1.46	2.14
PRECIPITATION_16DAYS_AVG_1DAYLAG	0.20	0.56	1.74
WINDSPEED_16DAYS_AVG_1DAYLAG	0.26	0.65	1.11
TAVG_OCTOBER_PREVYEAR	0.26	0.71	0.69
PRECIPITATION_4DAYS_AVG_1DAYLAG	0.17	0.53	0.93
TAVG_JANUARY_PREVYEAR	0.18	0.64	0.66
TAVG_SEPTEMBER_PREVYEAR	0.21	0.45	0.74
TAVG_MAY_PREVYEAR	0.17	0.52	0.64
TAVG_JULY_PREVYEAR	0.17	0.42	0.70
TAVG_DECEMBER_PREVYEAR	0.20	0.47	0.62
ALTITUDE	0.24	0.40	0.65
TAVG_MARCH_PREVYEAR	0.16	0.47	0.64
WINDSPEED_4DAYS_AVG_1DAYLAG	0.15	0.40	0.69
TAVG_FEBRUARY_PREVYEAR	0.17	0.42	0.62
TAVG_NOVEMBER_PREVYEAR	0.19	0.44	0.56
LATITUDE	0.20	0.44	0.54
TAVG_JUNE_PREVYEAR	0.17	0.39	0.61
LONGITUDE	0.21	0.36	0.60
TAVG_APRIL_PREVYEAR	0.18	0.41	0.51
TAVG_AUGUST_PREVYEAR	0.13	0.44	0.54

Mean decrease in accuracy: 2.5 5.0 7.5

was noticeably right-skewed. The 16-day average Penman potential evaporation from a free water surface also proved to be a highly important variable in all of the models. The optimal values of potential evaporation for predicting high pollen concentration levels were similar in the *Corylus* and *Alnus* models. In these models, potential evaporation values were between 0.6 and 1.7 (*Corylus*) and between 0.6 and 2 (*Alnus*) for at least 0.5 probability of high pollen concentration level. In the *Betula* model, optimal values of potential evaporation for high pollen concentration levels were between 1.7 and 3.6. In contrast, average monthly temperatures for the preceding year and spatial variables such as latitude, longitude, and altitude were the least important variables in the *Corylus*, *Alnus*, and *Betula* models (Fig. 4).
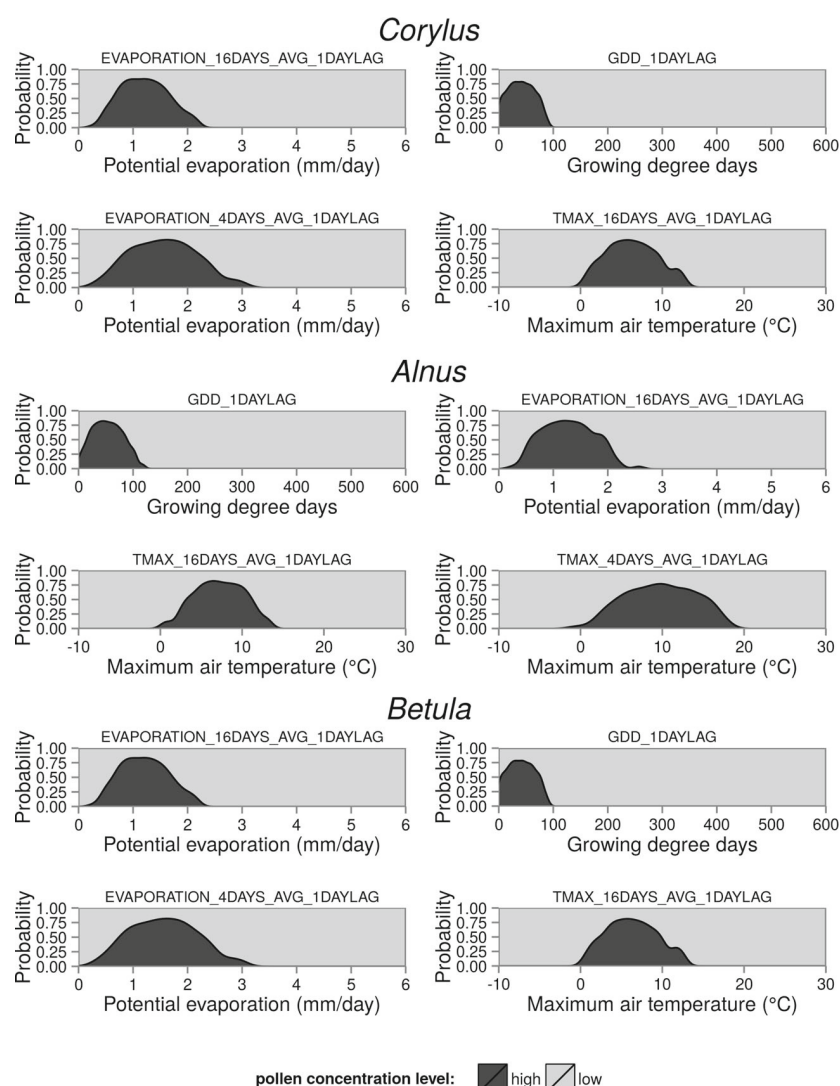
Performance of the models

The comparison of the models' performance on the test sets showed several trends (Fig. 6). In general, the *Betula* model produced the best prediction on both test sets, while the *Corylus* model was the least accurate. Moreover, there was a difference between the quality of the temporal and spatial models. For each taxon, the model was better in predicting the pollen concentration level on the first test set. The differences between sensitivity and positive predictive value were visible between the first and second test set. All of the models were very accurate in predicting low pollen concentration levels. Specificity ranged between 0.96 for the *Betula* model on the second test set and 0.99 for the *Corylus* model on the first test set. The average model specificity was 0.97 for both the first and second test set.

Temporal performance of the models

The positive predictive value of the *Corylus* model prediction on the first training set was 0.47. However, it is more important to correctly predict high pollen concentration levels than to misclassify low pollen concentration levels. The *Corylus* model performed reasonably well in predicting high levels of pollen concentration, with a sensitivity of 0.61. The *Alnus* model correctly predicted 203 out of 288 occurrences of days with high pollen concentration (sensitivity = 0.70). Moreover, the *Alnus* model's positive predictive value was distinctly higher than that of *Corylus*. The *Betula* model showed the best performance on the first test set. The model correctly classified approximately 88 % of days with high pollen concentration levels. The Kappa

Fig. 5 Relationship between probability of high pollen concentration level and the values of the four most important variables on the training set of the *Corylus*, *Alnus*, and *Betula* models

statistic value was 0.83, indicating a very high fit for the model.

Spatial performance of the models

The spatial quality was distinctly different in the models of each taxon. The *Corylus* model showed the lowest predictive capability. The model correctly predicted high pollen concentration levels in 40 out of 81 days (sensitivity = 0.49). The same model incorrectly classified 40 cases as high levels of pollen. The *Alnus* model performed better on the second test set. Both the models' sensitivity and positive predictive value were clearly higher: 0.61 and 0.59, respectively. The *Alnus* model correctly predicted 110 occurrences of high pollen concentration levels and misclassified 76 cases as high level. On the second test set, the performance

of the *Betula* model was found to be the best. The model Kappa statistic was 0.80, the sensitivity was 0.87, and the positive predictive value was 0.81. High *Betula* pollen concentration levels were correctly predicted in 394 of 451 cases. At the same time, only 94 days were incorrectly classified as high level.

Predictive maps

The *Corylus*, *Alnus*, and *Betula* models were built using processed variables from gridded meteorological data. Thus, it was possible to predict the probability of high pollen concentration levels for each cell and each date in the available data. Figure 7 show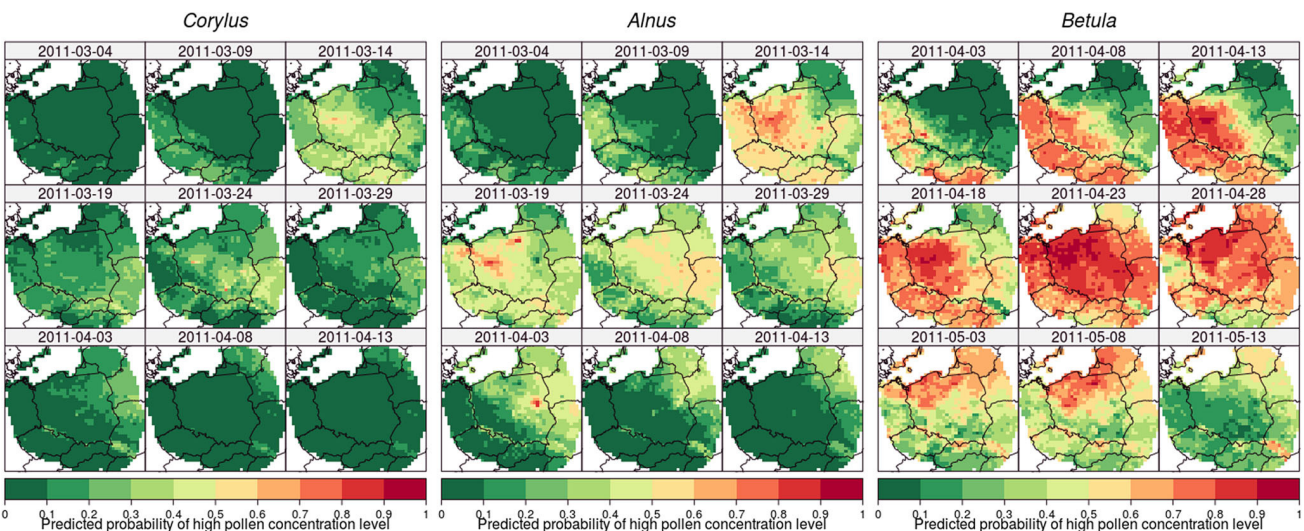s examples of the *Corylus*, *Alnus*, and *Betula* models' prediction for nine regularly distributed days in the year 2011.

Fig. 6 Relation between observed and predicted days with low and high concentration levels for *Corylus*, *Alnus*, and *Betula* pollen and prediction errors for the first and the second test set

Discussion

Corylus, *Alnus*, and *Betula* pollen have an enormous impact on the quality of life and the productivity of allergy sufferers. Therefore, these tree pollen are the origin of significant

social and financial burdens. Many aerobiological studies have been conducted in response to this problem, some of which have tried to build predictive models of pollen concentration (Bringfelt et al. 1982; Cotos-Yáñez et al. 2004; Castellano-Méndez et al. 2005; Rodriguez-Rajo et al. 2006;

Fig. 7 Examples of *Corylus*, *Alnus*, and *Betula* models' prediction for nine regularly distributed days in the year 2011, based on the data for Poland and the area within 200 km of the Polish border

Vogel et al. 2008; Hilaire et al. 2012; Puc 2012; Sofiev et al. 2013). As a result of these studies, it is possible to predict high pollen concentration levels with considerable accuracy in the analyzed sites. However, in many countries, the aerobiological network is not dense, and therefore, it is not possible to predict pollen counts in unsampled locations. In this study, gridded meteorological data were used as predictor variables to build a model of high *Corylus*, *Alnus*, and *Betula* pollen concentration levels for spatially continuous areas of Poland.

The days with high pollen concentration levels of the analyzed taxa occur rarely. This property should be taken into consideration when building predictive models. This study used a novel technique of obtaining an optimal threshold, by minimizing the distance between sensitivity, specificity, positive predictive value, and negative predictive value and the best possible performance. Preliminary studies showed that, in this two-class problem, the optimizing probability threshold technique surpasses other strategies for overcoming class imbalances, such as upsampling and downsampling.

In the *Corylus*, *Alnus*, and *Betula* models, cumulated growing degree days lagged by 1 day proved to be one of the most important variables. The fully growth competent buds need stimulation before they can burst; therefore, the occurrence of temperatures above a certain base level is required (Dahl et al. 2013). Secondly, most of the 16-day averages of meteorological factors (daily potential evaporation, total global radiation, vapor pressure, minimum temperature, and maximum temperature) showed high values of the variable importance. Previous studies showed that the readiness to flower is dependent on, inter alia, light intensity, and evaporation (Pacini and Hesse 2004; Dahl et al. 2013). In contrast, the importance of the preceding years average monthly temperatures for each month and grid cell longitude, latitude, and altitude had little impact on the model. Although the studies of Latałowa et al. (2002) and Rasmussen (2002) found a relationship between the annual pollen sum and the mean temperature in the preceding year, this relationship has a small influence on the daily pollen concentration.

The *Corylus*, *Alnus*, and *Betula* models varied in terms of predictive quality. The *Corylus* model predicted correctly approximately 55 % of the high pollen concentration levels on the test sets. This model misclassification could be connected with very rare (330 cases, about 2.5 % of the analyzed days) occurrences of high *Corylus* pollen concentration levels. *Corylus* inflorescences produce about two times fewer pollen grains than *Alnus* inflorescences (Piotrowska 2008). Thus, a dataset of longer time periods or a denser monitoring network could result in a more precise

model. The *Alnus* model performed better, with correct prediction of approximately 2/3 of high pollen concentration levels on the test sets. The problem of class imbalance was less severe in the *Alnus* dataset. Nonetheless, *Alnus* (and *Corylus*) pollen seasons are highly changeable from year to year. In addition, the location of aerobiological monitoring sites influences the variability of the pollen count of these taxa (Nowosad et al. 2015). The *Betula* model had the best values of model evaluation statistics. Almost 88 % of high pollen concentration levels were correctly predicted on the test sets. The negative impact of class imbalance was modest due to the relatively frequent occurrence of high *Betula* pollen concentration levels. Moreover, the *Betula* pollination period is relatively short and less changeable, and therefore, the *Betula* pollen count is more predictable.

The predictive quality of the *Betula* model is comparable to previous work. Castellano-Méndez et al. (2005) created a neutral network model for prediction of the risk of pollen concentration values exceeding a given level, using pollen and meteorological data. That model was built and validated in only one location: the city of Santiago, Spain. In contrast, in this study data from 11 aerobiological sites, as well as gridded meteorological data, were used in the process of model creation; therefore, model prediction should be verifiable in substantial areas surrounding the aerobiological monitoring sites.

Relationships between pollen concentration in the air and meteorological factors are complex and strongly nonlinear. Thus, classical statistical models, such as logistic regression or linear discriminant analysis, tend to perform poorly. Machine learning techniques could find patterns in nonlinear, noisy data, and generate prediction with relatively high accuracy (Recknagel 2001). Some of the most often used machine learning methods include nonlinear classification models (e.g., neural networks and support vector machines) and tree-based models (e.g., classification trees and random forest) (Kuhn and Johnson 2013). Random forest proved to give more accurate prediction than single tree models (Breiman 2001). In addition, this technique was compared to neural networks, and support vector machines require minimal preprocessing of the data. However, none of the single modeling techniques work best for every problem (Wolpert 1996). Therefore, it would be worthwhile to compare performance of different machine learning models for predicting pollen concentration.

Prediction errors of the *Corylus*, *Alnus*, and *Betula* models are the result of a combination of numerous factors: (i) omission of some non-meteorological predictors, (ii) influence of medium- and long-range pollen transport, and (iii) temporal and spatial uncertainty of pollen data. Meteoro-

logical conditions are also not the only factor that influence pollen concentration values. After the same meteorological conditions, high or low pollen concentration levels of *Corylus*, *Alnus*, and *Betula* can be observed on different occasions. Pollen concentration in the air is a result of nonlinear interactions between many factors, such as the land cover, topography, and human impact (Piotrowska and Kubik-Komar 2012). Taking into account the proportion of the analyzed taxa in the local vegetation could positively influence model quality. Previous studies also showed that most of recorded airborne pollen comes from local sources (Adams-Groom et al. 2002; Damialis et al. 2005). Nevertheless, medium- and long-range transport is also often recorded, as pollen grains are found hundreds or thousands of kilometers away from their source (Damialis et al. 2005; Ranta et al. 2006). In the *Corylus*, *Alnus*, and *Betula* models, the effect of long-range transport is not included. Moreover, uncertainty in the results of models could be connected with several characteristics of the data. The results of aerobiological monitoring are not the exact values of the pollen concentration of the surrounding area but are subject to errors from various sources, such as device, preparation of the sampling surface and slides, and slide analysis (Gottardini et al. 2009). In addition, there is diurnal variation in the number of pollen grains in the air (Galán et al. 1991; Skjoth et al. 2008). It is estimated that approximately 10 % of *Corylus*, *Alnus*, and *Betula* pollen count variations can be due to diurnal fluctuations and measurement errors (Nowosad et al. 2015). Only 11 aerobiological monitoring sites, which are not randomly distributed in Poland, were used in this study. The sites are located mainly in large cities, where the local climate is modified by human activities. These cities are significantly warmer than the surrounding rural areas, on average by 0.8–1.3 °C (Szymanowski 2005). Furthermore, the local airflow and turbulence are affected by buildings and non-building structures (Emberlin and Norris-Hill 1991). As a result, the deposition patterns in cities are different from those in the countryside (Emberlin and Norris-Hill 1991; Gonzalo-Garijo et al. 2006). Moreover, given the lack of sites in mountainous areas, caution should be exercised when using prediction models in those areas. In the long term, it will be valuable to add monitoring sites in remote rural areas, as well as in mountainous areas.

The modeling framework used in this study can be used as the basis for further research. The models are built based on meteorological factors and could be easily implemented in other countries. Moreover, it would be worthwhile to analyze the possibility of improving the models' quality by utilizing non-meteorological parameters, such as the distribution of tree species and local land use.

Conclusions

– In this study, the probability of high pollen concentration levels of *Corylus*, *Alnus*, and *Betula* was predicted using preprocessed gridded meteorological data. The result of the models could be used for prediction in continuous areas rather than just in single locations

– The models built allow moderately reliable predictions of high pollen concentration levels of *Corylus* and highly reliable predictions of high levels of *Alnus* and *Betula* pollen

– Temporal verifiability was higher than spatial verifiability in each of the *Corylus*, *Alnus*, and *Betula* models

– Average monthly temperatures for the preceding year were not very important for the results of the models

– Cumulated growing degree days was one of the most important variables in the *Corylus*, *Alnus*, and *Betula* models. In addition, sixteen-day averages of potential evaporation, total global radiation, vapor pressure, minimum temperature, and maximum temperature were important variables for the models.

– Spatial variables such as latitude, longitude, and altitude had little impact on the models

– The modeling framework could be applied in predicting high pollen concentrations of the different pollen taxa in the study sites and also in other areas

References

Adams-Groom B, Emberlin J, Corden J, Millington W, Mullins J (2002) Predicting the start of the birch pollen season at London, Derby and Cardiff, United Kingdom, using a multiple regression model, based on data from 1987 to 1997. Aerobiologia 18(2):117–123. doi:10.1023/A:1020698023134

Baruth B, Genovese G, Leo O (2007) GCMS version 9.2—User manual and technical documentation. Tech. rep.

Breiman L (2001) Random Forests. Mach Learn 45(1):5–32. doi:10.1023/A:1010933404324

Bringfelt B, Engström I, Nilsson S (1982) An evaluation of some models to predict airborne pollen concentration from meteorological conditions in Stockholm, Sweden. Grana 211:59–64. doi:10.1080/00173138209427680

Castellano-Méndez M, Aira MJ, 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(5):310–6. doi:10.1007/s00484-004-0247-x

Cotos-Yáñez TR, Rodríguez-Rajo FJ, Jato MV (2004) Short-term prediction of Betula airborne pollen concentration in Vigo (NW Spain) using logistic additive models and partially linear models. Int J Biometeorol 48(4):179–85. doi:10.1007/s00484-004-0203-9

Dahl A, Galán C, Hajkova L, Pauling A, Sikoparija B, Smith M, Vokou D (2013). In: Sofiev M, Bergmann KC (eds) The onset, course and intensity of the pollen season. Springer, Netherlands, Dordrecht, pp 29–70. doi:10.1007/978-94-007-4881-1_3

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 62(9):976–990. doi:10.1111/j.1398-9995.2007.01393.x

Damialis A, Gioulekas D, Lazopoulou C, Balafoutis C, Vokou D (2005) Transport of airborne pollen into the city of Thessaloniki: the effects of wind direction, speed and persistence. Int J Biometeorol 49(3):139–45. doi:10.1007/s00484-004-0229-z

Ebner C, Hirschwehr R, Bauer L, Breiteneder H, Valenta R, Ebner H, Kraft D, Scheiner O (1995) Identification of allergens in fruits and vegetables: IgE cross-reactivities with the important birch pollen allergens Bet v 1 and Bet v 2 (birch profilin). J Allergy Clin Immunol 95:962–969. doi:10.1016/S0091-6749(95)70096-X

Emberlin J, Norris-Hill J (1991) Spatial variation of pollen deposition in North London. Grana 30:190–195. doi:10.1080/00173139109427798

Emberlin J, Detandt 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. doi:10.1007/s00484-002-0139-x

Galán C, Tromo R, Cuevas J, Infante F, Domínguez E (1991) Theoretical daily variation patterns of airborne pollen in the South-West of Spain. Grana 30:201–209

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(4):385–395. doi:10.1007/s10453-014-9335-5

Gonzalo-Garijo M., Tormo-Molina R, Muñoz Rodríguez aF, Silva-Palacios I (2006) Differences in the spatial distribution of airborne pollen concentrations at different urban locations within a city. J Investig Allergol Clin Immunol 16(1):37–43

Gottardini E, Cristofolini F, Cristofori A, Vannini A, Ferretti M (2009) Sampling bias and sampling errors in pollen counting in aerobiological monitoring in Italy. J Environ Monit 11(4):751–755. doi:10.1039/b818162b

Heinzerling LM, Burbach GJ, Edenharter G, Bachert C, Bindslev-Jensen C, Bonini S, Bousquet J, Bousquet-Rouanet L, Bousquet PJ, Bresciani M, Bruno A, Burney P, Canonica GW, Darsow U, Demoly P, Durham S, Fokkens WJ, Giavi S, Gjomarkaj M, Gramiccioni C, Haahtela T, Kowalski ML, Magyar P, Muraközi G, Orosz M, Papadopoulos NG, Röhnelt C, Stingl G, Todo-Bom A, von Mutius E, Wiesner A, Wöhrl S, Zuberbier T (2009) GA(2)LEN skin test study I: GA(2)LEN harmonization of skin prick testing: novel sensitization patterns for inhalant allergens in Europe. Allergy 64(10):1498–506. doi:10.1111/j.1398-9995.2009.02093.x

Hilaire D, Rotach MMW, Clot B (2012) Building models for daily pollen concentrations. Aerobiologia 28(4):499–513. doi:10.1007/s10453-012-9252-4

Kizilpinar I, Civelek E, Tuncer A, Dogan C, Karabulut E, Sahiner UM, Yavuz ST, Sackesen C (2011) Pollen counts and their relationship to meteorological factors in Ankara, Turkey during 2005-2008. Int J Biometeorol 55:623–631. doi:10.1007/s00484-010-0363-8

Kuhn M (2015) caret: Classification and Regression Training. https://github.com/topepo/caret/

Kuhn M, Johnson K (2013) Applied predictive modeling. Springer New York

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

Liaw A, Wiener M (2002) Classification and Regression by random-Forest. R News 2(3):18–22

Myszkowska D, Jenner B, Puc M, Stach A, Nowak M, Malkiewicz M, Chłopek K, Uruska A, Rapiejko P, Majkowska-Wojciechowska B, Weryszko-Chmielewska E, Piotrowska K, Kasprzyk I (2010) Spatial variations in the dynamics of the Alnus and Corylus pollen seasons in Poland. Aerobiologia 26(3):209–221. doi:10.1007/s10453-010-9157-z

Nowosad J, Stach A, Kasprzyk I, Latałowa M, Puc M, Myszkowska D, Weryszko-Chmielewska E, Piotrowska-Weryszko K, Chłopek K, Majkowska-Wojciechowska B, Others, Grewling L, Latałowa M, Puc M, Myszkowska D, Weryszko- Chmielewska E, Piotrowska-Weryszko K, Chłopek K, Majkowska-Wojciechowska B, Uruska A (2015) Temporal and spatiotemporal autocorrelation of daily concentrations of Alnus, Betula, and Corylus pollen in Poland. Aerobiologia 31:159–177. doi:10.1007/s10453-014-9354-2

Pacini E, Hesse M (2004) Cytophysiology of pollen presentation and dispersal. Flora - Morphology, Distribution, Functional Ecology of Plants 199(4):273–285. doi:10.1078/0367-2530-00156

Pebesma E, Bivand R (2005) Classes and methods for spatial data in R. R News 5(2):9–13

Piotrowska K (2008) Ecological features of flowers and the amount of pollen released in Corylus avellana (L.) and Alnus glutinosa (L.) Gaertn. Acta Agrobotanica 61(1):33–39

Piotrowska K, Kubik-Komar A (2012) The effect of meteorological factors on airborne Betula pollen concentrations in Lublin (Poland). Aerobiologia 28:467–479. doi:10.1007/s10453-012-9249-z

Puc M (2007) The effect of meteorological conditions on hazel (Corylus spp.) and alder (Alnus spp.) pollen concentration in the air of Szczecin. Acta Agrobotanica 60(2):65–70. doi:10.5586/aa.2007.032

Puc M (2012) Artificial neural network model of the relationship between Betula pollen and meteorological factors in Szczecin (Poland). Int J Biometeorol 56(2):395–401. doi:10.1007/s00484-011-0446-1

Puc M, Kasprzyk I (2013) The patterns of Corylus and Alnus pollen seasons and pollination periods in two Polish cities located in different climatic regions. Aerobiologia 29:495–511. doi:10.1007/s10453-013-9299-x

R Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.r-project.org/

Ranta H, Kubin E, Siljamo P, Sofiev M, Linkosalo T, Oksanen A, Bondestam K (2006) Long distance pollen transport cause

problems for determining the timing of birch pollen season in Fennoscandia by using phenological observations. Grana 45(4):297–304. doi:10.1080/00173130600984740

Rapiejko P, Stankiewicz W, Szczygielski K, Jurkiewicz D (2007) Progowe stężenie pyłku róslin niezbędne do wywołania objawów alergicznych (Threshold pollen count necessary to evoke allergic symptoms). Otolaryngologia Polska 61(4):591–594. doi:10.1016/S0030-6657(07)70491-2

Rasmussen A (2002) The effects of climate change on the birch pollen season in Denmark. Aerobiologia 18(3-4):253–265. doi:10.1023/A:1021321615254

Recknagel F (2001) Applications of machine learning to ecological modelling. Ecol Model 146:303–310. doi:10.1016/S0304-3800(01)00316-7

Rieux C, Personnaz MB, Thibaudon M (2008) Spatial variation of airborne pollen over south-east France: Characterization and implications for monitoring networks management. Aerobiologia 24:43–52. doi:10.1007/s10453-007-9081-z

Rodriguez-Rajo F, Rodríguez-Rajo FJ, Valencia-Barrera RM, Vega-Maray AM, Suárez FJ, Fernández-González D, Jato V (2006) Prediction of airborne Alnus pollen concentration by using ARIMA models. Ann Agric Environ Med 13(1):25–32

Rodríguez-Rajo FJ, Dopazo A, Jato V (2004) Environmental factors affecting the start of pollen season and concentrations of airborne Alnus pollen in two localities of Galicia (NW Spain). Ann Agric Environ Med 11:35–44

Rojo J, Pérez-Badia R (2015) Spatiotemporal analysis of olive flowering using geostatistical techniques. Sci Total Environ 505:860–869. doi:10.1016/j.scitotenv.2014.10.022

Sauliene I, Sukiene L, Severova E, Kalnina L (2014) Comparison of Alnus, Corylus, Betula pollen seasons in Riga, Moscow and Vilnius. Aerobiologia 30:423–433. doi:10.1007/s10453-014-9341-7

Skjoth CA, Sommer J, Brandt Jr, Hvidberg M, Geels C, Hansen KM, Hertel O, Frohn LM, Christensen JH (2008) Copenhagen—a significant source of birch (Betula) pollen? Int J Biometeorol 52:453–462. doi:10.1007/s00484-007-0139-y

Skjoth CA, Smith M, Brandt J, Emberlin J (2009) Are the birch trees in Southern England a source of Betula pollen for North London? Int J Biometeorol 53(1):75–86. doi:10.1007/s00484-008-0192-1

Sofiev M, Siljamo P, Ranta H, Linkosalo T, Jaeger S, Rasmussen A, Rantio-Lehtimaki A, Severova E, Kukkonen J (2013) A numerical model of birch pollen emission and dispersion in the atmosphere. Description of the emission module. Int J Biometeorol 57(1):45–58. doi:10.1007/s00484-012-0532-z

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 52:311–321. doi:10.1007/s00484-007-0127-2

Szymanowski M (2005) Interactions between thermal advection in frontal zones and the urban heat island of Wroclaw, Poland. Theor Appl Climatol 82(3-4):207–224. doi:10.1007/s00704-005-0135-2

Veriankaite L, Siljamo P, Sofiev M, Sauliene I, Kukkonen J (2009) Modelling analysis of source regions of long-range transported birch pollen that influences allergenic seasons in Lithuania. Aerobiologia 26(1):47–62. doi:10.1007/s10453-009-9142-6

Vogel H, Pauling A, Vogel B (2008) Numerical simulation of birch pollen dispersion with an operational weather forecast system. International Journal of Biometeorology 52(8):805–814. doi:10.1007/s00484-008-0174-3

Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer New York

Wolpert DH (1996) The Existence of A Priori Distinctions Between Learning Algorithms. Neural Comput 8(7):1391–1420. doi:10.1162/neco.1996.8.7.1391

Chapter 6

Conclusions and the way forward

The objective of this thesis was to develop spatiotemporal predictive models of *Corylus*, *Alnus*, and *Betula* pollen concentration levels in the air in Poland. The research aims were achieved through the determination of temporal and spatiotemporal autocorrelation characteristics of *Corylus*, *Alnus*, and *Betula* pollen count (the third chapter), the creation and evaluation of *Corylus*, *Alnus*, and *Betula* pollen concentration levels model using past pollen count data from monitoring sites (the fourth chapter), and the development and application of a framework for the predictive mapping of the pollen concentration levels using preprocessed gridded meteorological data (the fifth chapter).

The number of pollen grains in the air varies in both space and time. Thus, autocorrelation analysis is important for documenting the patterns and changes in pollen concentration properties. The study provided an understanding of several aspects of spatial and space–time variations of *Corylus*, *Alnus*, and *Betula* pollen counts. Pollen correlation values were moderately to strongly autocorrelated with next-day values. A few changes in temporal autocorrelation were observed, with the most important being after approximately 3.5 and 15 days. The first autocorrelation decline could be connected with an exchange of air masses after the passage of a single weather front. The second decline could be connected with (i) the length of the upward or downward phase of pollen concentrations in a pollen season or (ii) the length of pollen season. In addition, autocorrelation modeling

helped to reveal that roughly 10% of the temporal variations were due to diurnal variability and measurement errors. *Corylus*, *Alnus*, and *Betula* pollen concentration values were also spatially correlated, with an average value of 0.58. Moreover, a dependence between correlation values and the distance between pairs of monitoring sites was evident, especially for *Corylus* and *Alnus*. The modeling results showed the differences in pollen count variation between taxa at these pairs of sites. The *Alnus* cross-correlation course was straightforward and the signal reduced to zero after 10 days on average. In the case of *Corylus* and *Betula*, the properties of space–time variation were more complex. Changes in cross-correlation were asymmetrical and the range was longer, sometimes even more than 20 days. At numerous pairs of monitoring sites, the peak of correlation was shifted by 1–3 days, which could be partially connected with prevailing winds from the west towards the east in this latitude zone of the Northern Hemisphere, the movement of atmospheric fronts, and the distance between sites. Lastly, autocorrelation analysis led to the identification of three spatial domains with homogeneous pollen concentration properties: (i) coastal cities; (ii) lowlands and highlands; (iii) cities in southern Poland with climate influenced by the Carpathian Mountains.

The spatiotemporal correlation properties of pollen count were used to develop and evaluate models for predicting high levels of *Corylus*, *Alnus*, and *Betula* pollen concentration. The models for *Alnus* and *Betula* had, at the least, considerable reliability. On the other hand, the models for *Corylus* had a low predictive quality. High levels of this latter taxon occurred rarely, in less than 10% of the analyzed pollen season. The relationship between *Corylus* pollen counts in different cities also proved to be more complex than that between *Alnus* or *Betula* counts. The properties of *Corylus* pollen seasons depend strongly on the type of habitat. Its pollen season lasts longer than those of *Alnus* and *Betula*, while the number of *Corylus* pollen grains in the air is substantially lower. Errors in the models were compared with temperature and precipitation. Rainfall was significantly connected with a false prediction of high levels of *Alnus* and *Betula*. Analysis of errors also showed that the models of *Corylus*, *Alnus*, and *Betula* were sensitive to changes in temperature. In the models studied, independent variables showed a clear temporal and spatial depen-

dency. The variables from the day before had the largest impact on pollen levels, and the importance of input from the same site as the output was clear. In most cases, the values from the nearest sites were also important. This could be explained either by similar weather conditions at neighboring stations or by the influence of long-distance transport. The variable importance noticeably decreased over time, which could be associated with the inflow of new air masses.

The models for predicting high levels of *Alnus* and *Betula* pollen concentration using the spatiotemporal correlation properties of pollen count had a substantial accuracy; however, their results could be used only in cities with an aerobiological monitoring station. Therefore, alternative models based on preprocessed gridded meteorological data were created. A novel technique for obtaining an optimal threshold was applied: minimizing the distance between sensitivity, specificity, positive predictive value, negative predictive value, and the best possible performance. Preliminary studies showed that this technique surpasses other strategies, such as upsampling and downsampling, for overcoming class imbalances in this two-class problem. Cumulated growing degree days and 16-day averages of potential evaporation, total global radiation, vapor pressure, minimum temperature, and maximum temperature were the most important variables for the models. On the other hand, spatial variables (latitude, longitude, altitude) and average monthly temperature for the preceding year had little impact on the models. In spite of the fact that the model for *Corylus* was again the least accurate, it could predict approximately half of the days with high pollen concentration levels. The performance of the *Alnus* model was better, with a correct prediction of more than 60% of the days with high pollen concentration levels. The *Betula* model performed best, correctly predicting about 88% of the days with high pollen concentration levels. In all of the cases, temporal verifiability was higher than spatial verifiability. However, this difference was less than 0.5% in *Betula* model. These models could be used for prediction in single locations as well as in continuous areas. As part of the study, a new modeling framework was developed, which can be applied to predict pollen concentration levels of different taxa and also in other areas.

The results presented provide a new insight into the temporal and spatiotemporal proper-

ties of *Corylus*, *Alnus*, and *Betula* pollen concentrations in the air, and they demonstrate the potential for predicting the pollen concentration levels of these taxa. Furthermore, many opportunities for research have arisen from this work. Firstly, the pollen data used in the creation and validation of the models contain measurement errors. Owing to these errors, forecast accuracy will never exceed a certain threshold. As a consequence, there is a need to determine (i) the accuracy threshold in aerobiological models, and (ii) the influence of the errors in the data on the errors in forecasts. Moreover, it is necessary not only to maintain, but also to extend aerobiological monitoring in Poland. Longer time series of data would reveal more information about the spatiotemporal properties of the taxa analyzed, particularly *Corylus*. New monitoring sites in remote rural, coastal, and mountainous areas would fill the gaps in our knowledge and positively affect predictive models. A denser monitoring network would also reveal the changes in pollen concentration on a local and regional scale. Most airborne pollen grains come from local sources. Therefore, another possibility would be to create detailed distribution maps of *Corylus*, *Alnus*, and *Betula*. This data could be used as a predictor in pollen concentration models. In addition, it remains to be further clarified whether the random forest method used in this study is the most robust method for predicting pollen concentrations of *Corylus*, *Alnus*, and *Betula*. This technique should be compared with other predictive models. Finally, predictions based on models of *Corylus*, *Alnus*, and *Betula* pollen concentration levels should be made available to researchers, allergists, allergy sufferers, and other interested parties. Thus, the models need to be deployed as a web service, monitored, and improved.

References

Accorsi, C. et al., 1991. Allergenic pollen: morphology and microscopic photographs. In G. D'Amato, F. T. M. Spieksma, & S. Bonini, eds. *Allergenic pollen and pollinosis in Europe*. Wiley-Blackwell.

Alcazar, P. et al., 1999. Effects of sampling height and climatic conditions in aerobiological studies. *Journal of Investigational Allergology and Clinical Immunology*, 9(4), pp.253–261.

Baruth, B., Genovese, G. & Leo, O., 2007. *GCMS version 9.2 - User manual and technical documentation*, Luxembourg: Publications Office of the European Union.

Berger, U., Kmenta, M. & Bastl, K., 2014. Individual pollen exposure measurements: are they feasible? *Current opinion in allergy and clinical immunology*, 14(3), pp.200–5.

Berggren, B., Nilsson, S. & Boëthius, G., 1995. Diurnal Variation of Airborne Birch Pollen at Some Sites in Sweden. *Grana*, 34(4), pp.251–259.

Boucher, A. et al., 2002. Development of a semi-automatic system for pollen recognition. *Aerobiologia*, 18(3), pp.195–201.

Bremer, B., Bremer, K. & Chase, M., 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, 161(2), pp.105–121.

Bringfelt, B., Engström, I. & Nilsson, S., 1982. An evaluation of some models to predict airborne pollen concentration from meteorological conditions in Stockholm, Sweden. *Grana*, 21(1), pp.59–64.

Cariñanos, P. & Emberlin, J., 2000. Comparison of two pollen counting methods of slides from a Hirst type volumetric trap. *Aerobiologia*, 16, pp.339–346.

Castellano-Méndez, M. et al., 2005. Artificial neural networks as a useful tool to predict the risk level of Betula pollen in the air. *International Journal of Biometeorology*, 49(5), pp.310–6.

Comtois, P., 1998. Statistical analysis of aerobiological data. In P. Mandrioli, P. Comtois, & V. Levizzani, eds. *Methods in aerobiology*. Pitagora Editrice: Bologna.

Comtois, P. & Mandrioli, P., 1997. Pollen capture media: a comparative study. *Aerobiologia*, 13(149), pp.149–154.

Comtois, P., Alcazar, P. & Néron, D., 1999. Pollen counts statistics and its relevance to precision. *Aerobiologia*, 15(1), pp.19–28.

Cotos-Yáñez, T.R., Rodríguez-Rajo, F.J. & Jato, M.V., 2004. Short-term prediction of Betula airborne pollen concentration in Vigo (NW Spain) using logistic additive models and partially linear models. *International Journal of Biometeorology*, 48(4), pp.179–85.

Czernecki, B., 2013. *Zastosowanie technik numerycznego modelowania do rozpoznania lokalnych warunków anemologicznych Słowińskiego Parku Narodowego (Numerical modeling techniques for local wind field recognition over the Słowiński National Park)*. PhD thesis.

Dahl, A. & Strandhede, S.-o., 1996. Predicting the intensity of the birch pollen season. *Aerobiologia*, 12,

pp.97–106.

Dahl, A. et al., 2013. The onset, course and intensity of the pollen season. In M. Sofiev & K.-C. Bergmann, eds. *Allergenic pollen*. Dordrecht: Springer Netherlands, pp. 29–70.

Damialis, A. et al., 2005. Transport of airborne pollen into the city of Thessaloniki: the effects of wind direction, speed and persistence. *International Journal of Biometeorology*, 49(3), pp.139–45.

Di-Giovanni, F. & Kevan, P.G., 1991. Factors affecting pollen dynamics and its importance to pollen contamination: a review. *Canadian Journal of Forest Research*, 21(8), pp.1155–1170.

Dmochowska, H. ed., 2013. *Concise Statistical Yearbook of Poland 2013*, Warsaw: Statistical Publishing Establishment.

Domaszewicz, B. ed., 2013. *Forestry 2013*, Warsaw: Zakład Wydawnictw Statystycznych.

D'Amato, G. et al., 2007. Allergenic pollen and pollen allergy in Europe. *Allergy*, 62(9), pp.976–990.

Emberlin, J. et al., 2002. Responses in the start of Betula (birch) pollen seasons to recent changes in spring temperatures across Europe. *International Journal of Biometeorology*, 46, pp.159–170.

Emberlin, J., Savage, M. & Woodman, R., 1993. Annual variations in the concentrations of Betula pollen in the London area, 1961–1990. *Grana*, 32(6), pp.359–363.

Emberlin, J. et al., 2007. Changes in the pollen seasons of the early flowering trees Alnus spp. and Corylus spp. in Worcester, United Kingdom, 1996-2005. *International Journal of Biometeorology*, 51(3), pp.181–91.

Frenguelli, G. et al., 1991. The influence of air temperature on the starting dates of the pollen season of Alnus and Populus. *Grana*, 30, pp.196–200.

Galán, C. & Dominguez-Vilches, E., 1997. The capture media in aerobiological sampling. *Aerobiologia*, 13, pp.155–160.

Galán, C. et al., 2007. *Spanish aerobiology network (REA) management and quality manual. Servicio de Publicaciones Universidad de Córdoba*, ISBN 978-84-690-6353-8.

Galán, C. et al., 2014. Pollen monitoring: minimum requirements and reproducibility of analysis. *Aerobiologia*, 30(4), pp.385–395.

Galán, C. et al., 1991. Theoretical daily variation patterns of airborne pollen in the South-West of Spain. *Grana*, 30, pp.201–209.

Gottardini, E. et al., 2009. Sampling bias and sampling errors in pollen counting in aerobiological monitoring in Italy. *Journal of Environmental Monitoring*, 11(4), pp.751–755.

Gregory, P.H., 1961. *The Microbiology of the Atmosphere*, New York: Interscience.

Heinrich, J. et al., 2002. Trends in prevalence of atopic diseases and allergic sensitization in children in Eastern Germany. *European Respiratory Journal*, 19, pp.1040–1046.

Heinzerling, L.M. et al., 2009. GA(2)LEN skin test study I: GA(2)LEN harmonization of skin prick testing: novel sensitization patterns for inhalant allergens in Europe. *Allergy*, 64(10), pp.1498–1506.

Helbig, N. et al., 2004. Numerical modelling of pollen dispersion on the regional scale. *Aerobiologia*, 3, pp.3–19.

Hilaire, D., Rotach, M.M.W. & Clot, B., 2012. Building models for daily pollen concentrations. *Aerobiologia*, 28(4), pp.499–513.

Hirst, J.M., 1952. An automatic volumetric spore trap. *Annals of Applied Biology*, 39(2), pp.257–265.

Holt, K.A. & Bennett, K.D., 2014. Principles and methods for automated palynology. *New Phytologist*,

203(3), pp.735–742.

Ipsen, H. et al., 1985. Immunochemical Characterization of Reference Alder (Alnus glutinosa) and Hazel (Corylus avellana) Pollen Extracts and the Partial Immunochemical Identity between the Major Allergens of Alder, Birch and Hazel Pollens. *Allergy*, 40(7), pp.510–518.

Jäger, S., Spieksma, E.T.M. & Nolard, N., 1991. Fluctuations and trends in airborne concentrations of some abundant pollen types, monitored at Vienna, Leiden, and Brussels. *Grana*, 30(May 2012), pp.309–312.

Kasprzyk, I. et al., 2001. Diurnal variation of chosen airborne pollen at five sites in Poland. *Aerobiologia*, 17(4), pp.327–345.

Köble, R. & Seufert, G., 2001. Novel Maps for Forest Tree Species in Europe. In *Proceedings of the 8th European Symposium on the Physico-Chemical Behaviour of Air Pollutants: "A Changing Atmosphere!"* Torino (Italy).

Laaidi, M., 2001. Regional variations in the pollen season of Betula in Burgundy: two models for predicting the start of the pollination. *Aerobiologia*, 17, pp.247–254.

Latałowa, M., Miętus, M. & Uruska, A., 2002. Seasonal variations in the atmospheric Betula pollen count in Gdańsk (southern Baltic coast) in relation to meteorological parameters. *Aerobiologia*, 18, pp.33–43.

Latałowa, M. et al., 2005. Diurnal patterns of airborne pollen concentration of the selected tree and herb taxa in Gdańsk (northern Poland). *Grana*, 44(3), pp.192–201.

Leventin, E., Rogers, C. & Hall, S., 2000. Comparison of pollen sampling with a Burkard spore trap and a Tauber trap in a warm temperate climate. *Grana*, 39, pp.294–302.

Lorenc, H. ed., 2005. *Atlas klimatu Polski (Climatic Atlas of Poland)*, Warsaw: IMGW.

Mahura, A.G. et al., 2007. Elevated birch pollen episodes in Denmark: contributions from remote sources. *Aerobiologia*, 23(3), pp.171–179.

Mandrioli, P. & Ariatti, A., 2001. Aerobiology: future course of action. *Aerobiologia*, 17(1), pp.1–10.

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*, 21(2), pp.115–124.

Mullins, J. & Emberlin, J., 1997. Sampling pollens. *Journal of Aerosol Science*, 28(3), pp.365–370.

Myszkowska, D., 2013. Prediction of the birch pollen season characteristics in Cracow, Poland using an 18-year data series. *Aerobiologia*, 29(1), pp.31–44.

Myszkowska, D. & Piotrowicz, K., 2009. Birch (Betula L.) pollen seasons in Cracow in 1991-2008 in relation to meteorological conditions. *Acta Agrobotanica*, 62(2), pp.67–75.

Myszkowska, D. et al., 2010. Spatial variations in the dynamics of the Alnus and Corylus pollen seasons in Poland. *Aerobiologia*, 26(3), pp.209–221.

Norris-Hill, J. & Emberlin, J., 1991. Diurnal Variation of Pollen Concentration in the Air of North-Central London. *Grana*, 3134(February).

Obtułowicz, K., Szczepanek, K. & Szczeklik, A., 1990. The value of pollen count for diagnosis and therapy of pollen allergy in Poland. *Grana*, 29(4), pp.318–320.

Oikonen, M.K. et al., 2005. The start of the birch pollen season in Finnish Lapland: separating non-local from local birch pollen and the implication for allergy sufferers. *Grana*, 44(3), pp.181–186.

Oteros, J. et al., 2013. Quality control in bio-monitoring networks, Spanish Aerobiology Network. *Science of The Total Environment*, 443, pp.559–565.

Pacini, E. & Hesse, M., 2004. Cytophysiology of pollen presentation and dispersal. *Flora - Morphology,*

Distribution, Functional Ecology of Plants, 199(4), pp.273–285.

Pedersen, B.V. & Moseholm, L., 1993. Precision of the daily pollen count. Identifying sources of variation using variance component models. *Aerobiologia*, 9(1), pp.15–26.

Piotrowska, K. & Kubik-Komar, A., 2012. The effect of meteorological factors on airborne Betula pollen concentrations in Lublin (Poland). *Aerobiologia*, 28(4), pp.467–479.

Puc, M., 2012. Artificial neural network model of the relationship between Betula pollen and meteorological factors in Szczecin (Poland). *International Journal of Biometeorology*, 56(2), pp.395–401.

Puc, M., 2007. The effect of meteorological conditions on hazel (Corylus spp.) and alder (Alnus spp.) pollen concentration in the air of Szczecin. *Acta Agrobotanica*, 60(2), pp.65–70.

Puc, M. & Kasprzyk, I., 2013. The patterns of Corylus and Alnus pollen seasons and pollination periods in two Polish cities located in different climatic regions. *Aerobiologia*, 29, pp.495–511.

Ranta, H. & Satri, P., 2007. Synchronized inter-annual fluctuation of flowering intensity affects the exposure to allergenic tree pollen in North Europe. *Grana*, 46(4), pp.274–284.

Ranta, H. et al., 2006. Long distance pollen transport cause problems for determining the timing of birch pollen season in Fennoscandia by using phenological observations. *Grana*, 45(4), pp.297–304.

Rapiejko, P., 1995. Monitoring aeroalergenów w Polsce (Aeroallergens monitoring in Poland). In R. Śpiewak, ed. *Pyłki i pyłkowica: Aktualne problemy.* Lublin: Instytut Medycyny Wsi, pp. 13–19.

Rapiejko, P. et al., 2007. Progowe st żenie pyłku roślin niezb dne do wywołania objawów alergicznych (Threshold pollen count necessary to evoke allergic symptoms). *Otolaryngologia Polska*, 61(4), pp.591–594.

Rieux, C., Personnaz, M.B. & Thibaudon, M., 2008. Spatial variation of airborne pollen over south-east France: Characterization and implications for monitoring networks management. *Aerobiologia*, 24, pp.43–52.

Rodriguez-Rajo, F. et al., 2006. Prediction of airborne Alnus pollen concentration by using ARIMA models. *Annals of Agricultural and Environmental Medicine*, 13(1), pp.25–32.

Sauliene, I. et al., 2014. Comparison of Alnus, Corylus, Betula pollen seasons in Riga, Moscow and Vilnius. *Aerobiologia*, 30, pp.423–433.

Scheifinger, H. et al., 2013. Allergenic Pollen: A Review of the Production, Release, Distribution and Health Impacts. In M. Sofiev & K.-C. Bergmann, eds. Dordrecht: Springer Netherlands, pp. 71–126.

Siljamo, P. et al., 2013. A numerical model of birch pollen emission and dispersion in the atmosphere. Model evaluation and sensitivity analysis. *International Journal of Biometeorology*, 57(1), pp.125–36.

Siljamo, P. et al., 2008. Sources, impact and exchange of early-spring birch pollen in the Moscow region and Finland. *Aerobiologia*, 24(4), pp.211–230.

Singh, A.B. & Mathur, C., 2012. An aerobiological perspective in allergy and asthma. *Asia Pacific Allergy*, 2(3), p.210.

Skjøth, C. et al., 2008. An inventory of tree species in Europe-An essential data input for air pollution modelling. *Ecological Modelling*, 217, pp.292–304.

Skjøth, C. et al., 2009. Are the birch trees in Southern England a source of Betula pollen for North London? *International Journal of Biometeorology*, 53(1), pp.75–86.

Skjøth, C. et al., 2008. Copenhagen - A significant source of birch (Betula) pollen? *International Journal of Biometeorology*, 52, pp.453–462.

Skjøth, C., Šikoparija, B. & Jager, S., 2012. Pollen Sources. In M. Sofiev & K. C. Bergmann, eds. *Allergenic pollen: A review of the production, release, distribution and health impacts.* Dordrecht: Springer

Science+Business Media, pp. 9–27.

Skjøth, C.a. et al., 2007. The long-range transport of birch (Betula) pollen from Poland and Germany causes significant pre-season concentrations in Denmark. *Clinical and Experimental Allergy*, 37, pp.1204–1212.

Sofiev, M. et al., 2013. Airborne pollen transport. In *Allergenic pollen*. Springer, pp. 127–159.

Sofiev, M. et al., 2006. Towards numerical forecasting of long-range air transport of birch pollen: theoretical considerations and a feasibility study. *International Journal of Biometeorology*, 50(6), pp.392–402.

Sofiev, M. et al., 2013. A numerical model of birch pollen emission and dispersion in the atmosphere. Description of the emission module. *International Journal of Biometeorology*, 57(1), pp.45–58.

Stach, A., 2000. Variation in pollen concentration of the most allergenic taxa in Poznań (Poland), 1995–1996. *Aerobiologia*, 16(1), pp.63–68.

Stach, A. et al., 2008. Factors that determine the severity of Betula spp. pollen seasons in Poland (Poznań and Krakow) and the United Kingdom (Worcester and London). *International Journal of Biometeorology*, 52(4), pp.311–321.

Stillman, E.C. & Flenley, J.R., 1996. The needs and prospects for automation in palynology. *Quaternary Science Reviews*, 15(1), pp.1–5.

Śpiewak, R., 2003. Historia zastosowań aeropalinologii w medycynie (History of the applications of aeropalynology in medicine). In S. A. Dybova-Jachowicz S., ed. *Palinologia*. Krakow: Institute of Botanics PAN, pp. 61–62.

Toro, F. et al., 1998. Predictive models in aerobiology: data transformation. *Aerobiologia*, 14, pp.179–184.

Toth, I. et al., 2011. Diurnal variation in airborne pollen concentrations of the selected taxa in Zagreb, Croatia. *Collegium antropologicum*, 35 Suppl 2, pp.43–50.

Tutin, T.G., 1993. Corylus L. In T. G. Tutin et al., eds. *Flora Europaea, Volume 1: Psilotaceae to Platanaceae*. Cambridge: Cambridge University Press, p. 71.

Valenta, R. et al., 1991. Homology of the major birch-pollen allergen, I, with the major pollen allergens of alder, hazel, and hornbeam at the nucleic acid level as determined by cross-hybridization. *Journal of Allergy and Clinical Immunology*, 87(3), pp.677–682.

Veriankaite, L. et al., 2009. Modelling analysis of source regions of long-range transported birch pollen that influences allergenic seasons in Lithuania. *Aerobiologia*, 26(1), pp.47–62.

Vogel, H., Pauling, A. & Vogel, B., 2008. Numerical simulation of birch pollen dispersion with an operational weather forecast system. *International Journal of Biometeorology*, 52(8), pp.805–814.

Walters, M., 1993. Betula L. In T. G. Tutin et al., eds. *Flora Europaea, Volume 1: Psilotaceae to Platanaceae*. Cambridge: Cambridge University Press, pp. 69–70.

Weger, L.A. et al., 2013. Impact of Pollen. In *Allergenic pollen*. Springer, pp. 161–215.

Weryszko-Chmielewska, E., Puc, M. & Rapiejko, P., 2001. Comparative analysis of pollen counts of Corylus, Alnus and Betula in Szczecin, Warsaw and Lublin (2000-2001). *Annals of Agricultural and Environmental Medicine*, 8(2), pp.235–240.

World Allergy Organization, 2013. *White Book on Allergy: Update 2013, Executive Summary* R. Pawankar et al., eds.,

Woś, A., 2010. *Klimat Polski w drugiej połowie XX wieku (Climate of Poland in the second half of the 20th century)*, Poznan: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.

Zuberbier, T. et al., 2014. Economic burden of inadequate management of allergic diseases in the European Union: a GA(2) LEN review. *Allergy*, 69(10), pp.1275–1279.