Development of deterministic and stochastic models for predicting annual airborne pollen - integrating the recursive properties of masting

Yi-Ting TSENG

2020

Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Annual airborne pollen as an important index

Clinical evidence reveals a general increase in both incidence and prevalence of allergic diseases in industrialized countries over the last half century (Pawankar et al., 2013). The cause of these diseases is not only attributed to air pollutants (D'Amato et al., 2016) and changes in lifestyle (Graham-Rowe, 2011), but also an increase in allergenic pollen concentration (Ziello et al., 2012).

The annual airborne pollen, or seasonal pollen index (SPI), which is the sum of daily pollen concentrations over the year, is one of the most important indices describing pollen dynamics. It reflects both the duration of the pollen season and the dispersed condition as it is the overall outcome of the pollen dispersal of a year. Estimation of the annual airborne pollen and its interannual variability not only satisfies the needs of allergy stakeholders, but also benefits the prediction of crop yields and the impact of climate change on plants.

1.2 Determining factors of annual airborne pollen

Analysis of the annual airborne pollen of anemophilous trees indicates that airborne pollen concentrations vary considerably from year to year. This variation reflects the plant flowering intensity which is regulated by two factors: weather conditions during flower formation and the inner resource for assimilation.

For example, male flowers in birch trees initiate formation between May and July prior to the flowering year, depending on the climate zone (Caesar and Macdonald, 1983; Caesar and Macdonald, 1984; Dahl and Strandhede, 1996; Ranta et al., 2008). Hence, the flowering intensity is closely related to weather conditions of the previous summer (Ranta et al., 2008;

Yasaka et al., 2009), and is dependent particularly on temperature and sunlight (Matthews, 1955; Norton and Kelly, 1988), which affect the assimilation efficiency for the generation of carbohydrates during male flower formation (Dahl and Strandhede, 1996).

Flowering intensity is also affected by the endogenous flowering rhythm. This effect can be seen in anemophilous trees and other perennial plants which are capable of producing large quantities of flowers and seeds simultaneously, as well as synchronizing the variation in reproduction output. It is commonly known as masting behavior (Kelly and Sork, 2002; Koenig and Ashley, 2003; Koenig and Knops, 2000; Nilsson and Wastljung, 1987). This phenomenon explains why mast year occurs only once every several years even though favorable weather conditions for flower development, pollination, fertilization, and resource storage are available (Masaka and Maguchi, 2001).

The combination of meteorological factors and biological rhythm is therefore known to affect the flowering intensity. Considering both of these factors, it can be concluded that the quantity of airborne pollen dispersed and the meteorological conditions of the previous year are necessary to define the quantity of airborne pollen dispersed in a given flowering year.

1.3 Inadequacy in the study of annual airborne pollen

There is a considerable lack of research conducted on annual airborne pollen. Limited studies have been carried out in a small number of areas, since the monitoring sites of airborne biological matters are only established in few places and usually lack long-term observation. In addition, the sum of airborne pollen cannot be predicted in regions without an airborne pollen observatory because apart from the meteorological conditions, the quantity of airborne pollen dispersed in the previous year is also necessary to define the quantity of airborne pollen dispersed in a given flowering year.

1.4 Objectives of the thesis

This study is aimed at (i) developing and validating two alternative models for predicting annual airborne pollen for birch trees that is widely distributed at high latitudes in the Northern Hemisphere and associated with high allergenicity, (ii) examining the applicability of the proposed models in other regions, and (iii) discussing future prospects and limitations of these models.

The two proposed models, deterministic and stochastic, are based on recursive properties of masting where, the preceding annual airborne pollen (P_{i-1}) and meteorological conditions (M_{i-1}) are necessary to define the annual airborne pollen in a given flowering year (P_i). The deterministic algorithm in Chapter 3 was derived by substituting variable P_{i-1}repeatedly, turning it into a multiple regression model predicting the annual airborne pollen using only the meteorological conditions of previous years. The stochastic model, in Chapter 5, was the first application of a stochastic model to pollen concentration prediction. It was based on the hidden Markov model which integrated the stochastic process of mast flowering and meteorological conditions influencing flower formation to predict the annual airborne pollen level. The proposed models were developed and validated using the longitudinal data of flat region in Hokkaido, Japan and the applicability of the deterministic model was clarified and discussed to regions sharing similar latitudes with different topological features in Switzerland.

In contrast to previous studies, the two models can predict the annual airborne pollen using only meteorological data. Hence, the annual airborne pollen for the subsequent season can be still predicted without pollen data of the current season.

1.5 Structure of the thesis

The thesis has been divided into seven chapters. Following the introduction, Chapter 2 details a comprehensive literature review on the issues of allergenic pollen, factors regulating flowering intensity, and the prediction model of the annual airborne pollen. Chapter 3 outlines the development of the algorithm predicting annual airborne pollen and Chapter 4 discusses the applicability of the proposed algorithm to different regions with various topographic properties. An alternative approach to the deterministic model, the hidden Markov model, is proposed in Chapter 5. In Chapter 6, the limitations and future prospects of the proposed predicted models are discussed. Chapter 7 concludes the main findings of the thesis.

Chapter 2
Literature Review

2.1 Allergenic pollen and prevalence of pollinosis

Allergenic pollen has long been an issue of great concern for more than a century. Pollinosis was first described by Bostock (1819) as follows, “About the beginning or middle of June in every year ... A sensation of heat and fullness is experienced in the eyes .... To this succeeds irritation of the nose producing sneezing. To the sneezing are added a further sensation of tightness of the chest, and a difficulty of breathing”. Allergenic pollen started receiving considerable attention in the early 1870s, after the first skin prick test, and was recognized as a pathogenic cause for allergic rhinitis, asthma, and conjunctivitis (Blackley, 1873).

Allergenic pollen is one of the most widespread allergens in the world. The geographical distributions of plants play an integral role in the diversity of allergens; for e.g., ragweed (*Ambrosia* L.) is the major source of allergenic pollen in the United States, Europe, and Australia (Thibaudon et al., 2014), as Japanese cedar (*Cryptomeria japonica*) is to Japan (Okuda, 2003) and birch (*Betula* L.) is to Europe and Hokkaido (WHO, 2003). In addition, the prevalence of pollinosis is approximately 29.8% in Japan (Nakae and Baba, 2010), 40% in Europe (D’Amato et al., 2007), and 10–30% in the global population (Pawankar et al., 2013), and is still on the rise. This increase can be attributed to the urban life style as theory “hygiene hypothesis” (Graham-Rowe, 2011), the synergistic effect of airborne pollen with co-stressors such as gaseous and particle-phase air pollutants, and the widespread presence of allergens (Behrendt et al., 1992; D'Amato et al., 2016; Heinrich and Wichmann, 2004). Climate change was also reported to cause adverse impacts on the prevalence of allergic reactions; the effects of which can be seen at the onset, duration, intensity of the pollen season, as well as on pollen allergenicity (Beggs, 2004; D'Amato et al., 2016; D'Amato et al., 2015; Ziello et al., 2012).

2.2 Allergenicity of airborne birch pollen

Betulaceae, including the genera *Betula* (birch), *Corylus* (hazel), and *Alnus* (alder), are the source of allergenic pollen at the Northern Hemisphere temperate zones. They belong to the order Fagales, which has high levels of allergenic cross-reactivity with other members of the Betulaceae family and the grass, causing pollinosis in sensitized individuals during the pollen season. Among the Betulaceae family genera, birch is the major pollen-allergen-producing anemophilous tree in northern and central Europe and is the source of the most widely distributed airborne pollen in that region (D'Amato et al., 2007) as it is efficient colonist of disturbed ground and usually plant as ornamental trees (Madeja et al., 2005). It is also one of the lead causes of pollinosis, with approximately 10–20% of the population estimated to be allergic to birch pollen (Corden et al., 2002; Oei et al., 1986; Spieksma, 1990).

2.3 Airborne pollen monitoring

Airborne pollen monitoring has been instrumental in understanding pollen dispersal dynamics. It enables scientists to compile an annual pollen calendar and examine the relationship between airborne pollen concentrations and the severity of pollinosis symptoms.

The origins of airborne pollen and spore monitoring can be traced back to late 19th century, when "living organisms in the air" (Miquel, 1883) received considerable interest. Scientists began conducting aerobiological observations using low precision instruments built with clothes begs and a glass slide. After the late 1900s, more advanced samplers were introduced, namely conventional and modern devices. Conventional samplers includes gravitational Durham samplers (Durham, 1946) and volumetric Burkard samplers (Hirst, 1952). They both require intense labor work and skill to count the number of pollen grains. However, their simplicity and accuracy make them a popular choice in Japan and other countries of the world. Modern samplers comprise of the latest devices with automatic counting, such as KH-3000-01 (Kawashima et al., 2007) and can identify pollen species using an installed algorithm (e.g., Rapid-E (Šaulienė et al., 2019) and BAA500 (Oteros et al., 2015)). The comparisons of the samplers are summarized in Table 2- 1.

Despite the importance of monitoring airborne pollen, it is not as predominant compared to non-biological air pollutants. The monitoring networks of the non-biological (chemical) components usually are publicly funded and air quality data are open to the public. However, except for only a few countries like Japan, France, and Switzerland, which have state-owned monitoring networks for airborne biological matters, monitoring networks are usually privately owned and the data are not freely available. In addition, many regions lack long-term observation or a lack of any observatories. The information of pollen monitoring network over the world was summarized by Buters et al. (2018).

Table 2- 1 Summarization of conventional and modern airborne pollen samplers.

	Conventional		Modern		
Sampler	Durham (Durham, 1946)	Hirst (Hirst, 1952)	KH-3000-01 (Kawashima et al., 2007)	BA500 (Oteros, 2015)	Rapid-E (Šaulienė et al., 2019)
method	Gravitational method	Volumetric method	Volumetric method	Volumetric method	Volumetric method
Time resolution	Daily	Hourly	Hourly	Hourly	Hourly
Usage condition	Widely used in Japan	Widely used in the world.	Hanako-san in Japan	Germany	Switzerland
Automatic counting?	No	No	Yes (Laser-optics)	Yes (Laser-optics)	Yes (Laser-optics)
Automatic identification?	No	No	No	Yes	Yes

2.4 Annual airborne pollen: continuing concern in epidemiology, plant reproduction, and climate change

Using longitudinal airborne pollen monitored data, three aspects of airborne pollen dynamics can be modeled: the onset and duration of the airborne pollen season (temporal), spatial distribution of airborne pollen (spatial), and airborne pollen concentration in different time resolutions (amount).

The annual airborne pollen has become an aerobiological index of concern ever since the first pollen trap was constructed in the late 1860s. It is the summation of the pollen concentration over a pollen season, which is used to determine the overall dispersal outcome each year.

Predicting the annual airborne pollen concentrations has significant value for allergy stakeholders. Since the triggering of allergic reactions and the severity of pollinosis are highly correlated with the concentration of allergenic pollen (Voukantsis et al., 2010), annual airborne pollen data provides an important reference in decision making for clinical strategies for allergic diseases and management of anti-allergy medicine production.

In addition, the annual airborne pollen can reflect the flowering intensity. Since the availability of airborne pollen partly determines the success of pollination and fertilization (Bogdziewicz et al., 2017b), the annual airborne pollen is used as a bio-indicator for studies on plant reproductive behavior and crop yields of anemophilous plants; such as olive (Galán et al., 2008; Oteros et al., 2013), wine (Cristofolini and Gottardini, 2000; Cunha et al., 2016), oak (Fernández-Martínez et al., 2012; Knapp et al., 2001), and beech (Kasprzyk et al., 2014).

The fluctuation pattern of the annual airborne pollen provides a detailed perspective into the reactions of plants to climate change (Damialis et al., 2007; Garcia-Mozo et al., 2006; Hatfield and Prueger, 2015; Oteros et al., 2014; Recio et al., 2018; Rogers et al., 2006; Zhang et al., 2015; Ziska and Beggs, 2012), as numerous studies have reported an increase in aeroallergens which coincides with the rise in CO_2 and temperature (Kim et al., 2012; Wayne et al., 2002; Ziska and Caulfield, 2000).

2.5 Masting behavior—inter-annual variation of productive effort

2.5.1 Definition

Masting behavior, also known as masting, mast flowering, or mast seeding, is a widespread and extensively studied phenomenon in anemophilous trees and other perennial plants (Nilsson and Wastljung, 1987). It refers to both, a highly variable inter-annual flowers and seeds production among individual plants (CV_i), and intra-annual production synchrony within a plant population (S) (Kelly, 1994; Koenig and Knops, 1998; Koenig and Knops, 2000). Pearse (2016) summarized two levels of hypothesis for masting: ultimate and proximate levels. Ultimate-level hypothesis describes masting from an evolutionary aspect, whereas proximate-levels hypothesis details the mechanisms causing masting, including factors such as the external environmental restrainers and internal resource limitations from producing large amounts of flowers and seeds in some years but not others.

2.5.2 Ultimate-level hypothesis: evolutionary advantages

Ultimate-level hypothesis details the mechanisms of masting that bring about benefits from economies of scale (EOS), which leads to lower cost per surviving offspring. Plants with a masting function are at an advantage and thrive under selection pressures of evolution. Three possible EOSs were proposed: (i) decrease in the probability of an individual organism being consumed (predator satiation) (Janzen, 1971; Silvertown, 1980), (ii) increase pollination efficiency by synchronizing flowering efforts (Moreira et al., 2014; Smith et al., 1990; White et al., 1994) and (iii) enabling surviving offspring to be set into a low-competition and high-nutrient soil, after environmental changes such as a fire (environmental predictions). In ultimate-level, although predator satiation does not affect CV_i and S directly, pollination efficiency and environmental predictions have a positive influence on S and both CV_i and S, respectively (Fig. 2-1).

2.5.3 Proximate-level hypothesis: weather cues and inner resources

Proximate-level hypothesis focuses on the mechanisms causing masting behavior and do not necessarily require ultimate-level benefits. The two main proximate causes of masting are weather cues (Bogdziewicz et al., 2018; Koenig and Knops, 2000; Koenig et al., 2015; Norton and Kelly, 1988; Vacchiano et al., 2017) and variation of inner resources. They were proposed to influence annual seed production in two stages of the plant development cycle: (i) flowering and pollination and (ii) population-level seed production (Fig. 2- 2). The dominance of either of the two stages, in masting, is species dependent (Bogdziewicz et al., 2019). For instance, European beech is a flower masting species and seed production is mainly determined by step (i). On the contrary, sessile oak is a fruit-maturation masting species where variable ripening is determined by step (ii) (Bogdziewicz et al., 2019).

Weather cues

Various hypotheses support the influence of weather cues on masting. In step (i), phenology synchrony hypothesis (② in Fig. 2- 2) supports that weather conditions can affect the flowering synchrony and associated pollination efficiency (Koenig et al., 2015).

In step (ii), pollination dynamics' influence on seed crops can be predicted by two hypotheses. Pollen coupling (density-dependent pollen limitation) (⑥ in Fig. 2- 2) refers to the positive correlation of flower production and pollination efficiency. Under favorable conditions, either high initial flower production (flower masting) or high flower success (fruit maturation masting) will then lead to the mast seed crop (Pearse et al., 2016). On the other hand, poor seed crops are coupled with a low-flowering year due to limited pollen (Ashman et al., 2004; Knight et al., 2005; Satake and Iwasa, 2000). Pollination Moran effect (⑤ in Fig.

2- 2), on the contrary, is a density-independent process and suggests that meteorological conditions, which are commonly shared in large spatial scales (e.g. precipitation), limit the pollination efficiency or transition from flower to fruit (Koenig, 2002; Pearse et al., 2016). Under this condition, flowering and seeding is decoupled, individuals in the population experience similar unfavorable weather conditions, causing failure in pollination or transition from flower to fruit. This is associated poor crop production comparing to an abundant crop production during a mast year (Bogdziewicz et al., 2017a; García-Mozo et al., 2007; Koenig, 2002; Koenig and Knops, 1998; Lyles et al., 2009).

Variation of resources

Variation in flowers and seeds are difficult to explain using only meteorological conditions. Hence, more studies advocate considering resources (e.g. carbon, nitrogen, phosphorous, and potassium) and their limitations in reproduction that align with masting.

Resource-budget model refers to the dynamics (depletion and accumulation) of inner resources where, (i) masting requires more resources than plants can gain in a single year, (ii) resources are accumulated over several years until it reaches threshold for the masting, and (iii) reproduction in the mast years depletes stored resources (⑦ in Fig. 2- 2) resulting in the next low-flowering year.

Experimental tests were implemented to verify the source depletion hypothesis inside the plant. A study conducted by Crone et al. (2009) showed that plants, whose flowers were removed to prevent seed sets in the year, are able to flower again the following year due to larger nonstructural carbohydrate (NSC) stored within the plants.

In proximate-level, weather cues affect the levels of CV_i and S, the two criteria for masting, due to its relation to hormonal regulation for the development in individuals, as well as wider spatial consistency for synchronization in intra-individuals. On the other hand, resources of variation affect the level of CV_i (Fig. 2- 1).

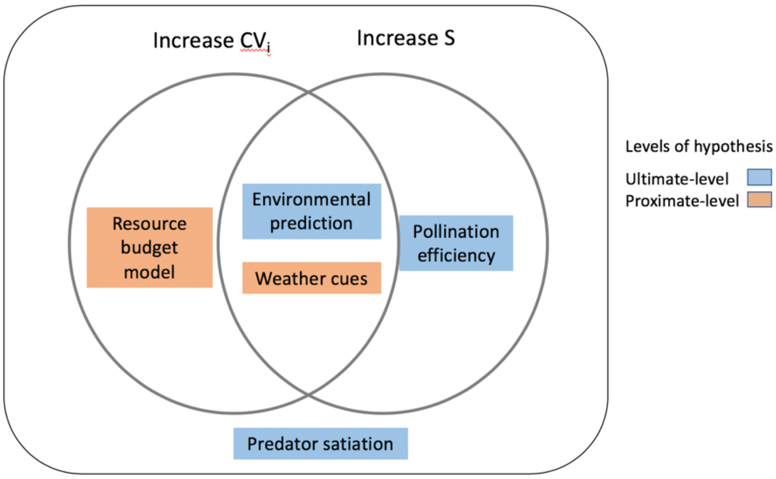


Fig. 2- 1 Direct effects of hypotheses for masting (CV_i : individual variability; S: among-individuals synchrony).

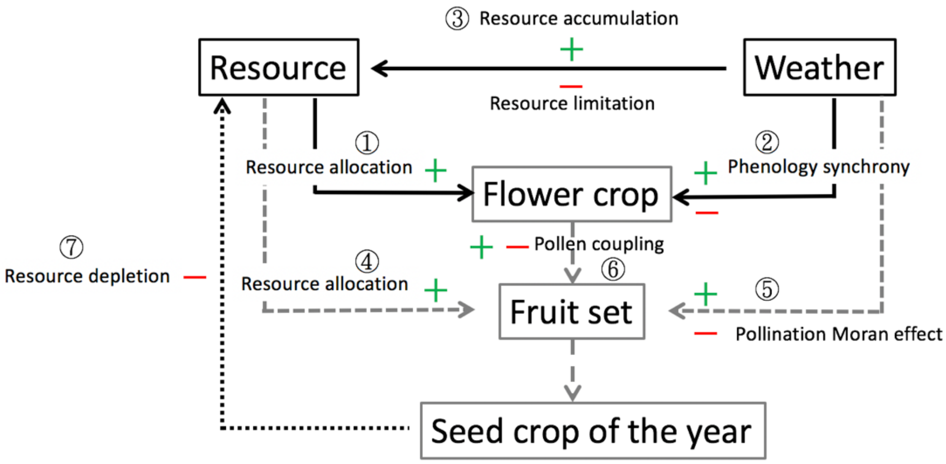


Fig. 2- 2 Conceptual cycle of proximate mechanisms in whole reproduction process and suggested hypotheses.

2.6 Predicting the annual airborne pollen

Factors regulating masting, weather cues, and inner resources, have been applied to studies predicting annual airborne pollen.

2.6.1 Model of annual airborne pollen with only meteorological factors

A number of studies have shown a strong relationship between meteorological conditions and annual airborne pollen. Winter chilling (Lavee, 2007) and rainfall (Galán et al., 2001) are good indicators for the synchronization of flowering in olive; rainfall and temperature were found to be important for pollen season intensity in oak (Garcia-Mozo et al., 2006), especially before or during the spring (Recio et al., 2018); weather conditions prior to the flowering year were revealed to have a strong correlation with flowering intensity in birch (Dahl and Strandhede, 1996; Emberlin et al., 1993; Ranta et al., 2008; Ranta et al., 2005; Rasmussen, 2002; Stach et al., 2008). These studies support the notion that appropriate environmental and weather conditions are capable of inducing flower synchronization leading to higher annual airborne pollen. Pre-seasonal weather parameters are seen as robust references for flowering intensity and annual airborne pollen prediction. Numerous studies have presented the annual airborne pollen forecasting model based on a "resource matching" hypothesis, i.e., considering only weather conditions. Among them, linear regression is the most used (Garcia-Mozo et al., 2014; Latałowa et al., 2002; Ranta et al., 2008; Subiza et al., 1992).

2.6.2 Model of annual airborne pollen with both biological and meteorological factors

Recent studies have highlighted the importance of integrating biological factors to predict annual pollen production, to influence of the nutritional status on flower production (Jochner et al., 2013). Followed the resource-budget model first proposed by Isagi et al. (1997), Masaka and Maguchi (2001) developed the model of masting behavior for birch (*Betula platyphylla var. japonica*) in Sapporo, incorporating biological and meteorological factors a year before flowering year, annual airborne pollen and weather conditions, using multiple linear regression.

Using both biological and meteorological factors, most of the forecasting model of annual airborne pollen were built with multiple linear regression (Ritenberga et al., 2018), with little number of studies using other statistical method such as time series analyses including Seasonal and Trend decomposition using Loess (STL) and the Box-Jenkins method (Garcia-Mozo et al., 2014).

2.7 Stochastic model and hidden Markov model (HMM)

As mentioned in 2.6.2, deterministic approach has to date been employed for predicting annual airborne pollen, in which the forecast is made entirely based on parameters. However, given the complexity of the masting mechanism (which has intrinsic stochastic properties); few attempts have been made to apply a stochastic model that considers the inter-annual airborne pollen variation as a stochastic process.

In contrast to deterministic models, stochastic models provide an alternative approach for representing biological processes by formalizing the notion of uncertainty in the outcome of an experiment. The Markov decision process, a discrete stochastic process, is usually employed to simplify the problems of recurrent events with the assumption that the current state depends only on the previous state with a transition probability (Bailey and Bailey, 1990). Since its introduction by Waggoner and Stephens (1970), the Markov model has been widely employed in plant ecology applications (Binkley, 1980; Van Hulst, 1979). The most ordinary type of Markov model, the stationary Markov model (SMM), has a probability distribution that remains unchanged in the Markov chain over time. It has been applied to describe the yield probabilities of wheat crops annually, since wheat yields in year i can be a function of the yield in year $i-1$ (Bostwick, 1962). However, owing to the limitations of the model structure, the SMM requires that the ecological observations be represented directly by the model states.

As an extension of the Markov model, the hidden Markov model (HMM) provides a framework to overcome the shortcomings of the SMM. The HMM is a doubly stochastic process including an unobservable Markov process and an output sequence depending on it. It is an example of a state-space model involving a discrete hidden state with continuous observation (continuous HMM) or discrete observation (discrete HMM). Its structure allows for the description of hidden processes (such as disturbance regimes) underlying the observed ecological dynamics (community states). Owing to its ability to recover a data sequence that is not observable, the HMM has been used increasingly in speech recognition (Baum et al., 1970), biological sequence analysis (Churchill, 1989; Stultz et al., 1993; White et al., 1994; Yoon, 2009), and a wide range of applications, such as in psychology (Visser et al., 2009), finance (Thomas et al., 2002), and hydrology (Thyer and Kuczera, 2003a; Thyer and Kuczera, 2003b).

The HMM has also become the preferred approach for analyzing ecological problems such as population dynamics, animal movements, and animal behaviors. It provides an approach for describing probabilistic dependence between the latent states and observation, which enables analysis of complex ecological dynamics governed by underlying hidden states that cannot be observed directly (Tucker and Anand, 2005). This enables tracking of the geolocation of animal movements from observable measurements (Jonsen et al., 2013; Langrock et al., 2012; Patterson et al., 2008; Woillez et al., 2016), and the determination of the phenomenology of behaviors from underlying mechanisms (Franke et al., 2004; Franke et al., 2006). The HMM has also been introduced in the field of atmospheric environment to predict the concentration of certain pollutants (Sun et al., 2013a; Sun et al., 2013b; Zhang et al., 2012). The applications mainly utilize the dependence of the doubly embedded stochastic process to predict hidden state sequences from observations. In such cases therefore, the subject of interest (concentration of air pollutants in the near future) is predicted using the previous hidden state (current concentration of air pollutants) and the observed measurements (key meteorological factors and precursors of pollutants).

Chapter 3
A deterministic algorithm for predicting annual airborne pollen

Tseng YT, Kawashima S, Kobayashi S, Takeuchi S, Nakamura K. Algorithm for forecasting the total amount of airborne birch pollen from meteorological conditions of previous years. *Agric. For. Meteorol.* 2018; 249: 35-43.

3.1 Introduction

The variance in pollen production is proved to be correlated with meteorological conditions in the previous summer and resource allocation and depletion (Section 2.5.3). Considering both of these factors, it can be concluded that both the quantity of male flowers produced (reflecting the source depletion) and the meteorological conditions of the previous year are necessary to define the male flowers abundance in a given flowering year. By utilizing the recursive characteristics of the foregoing relation, an algorithm that the annual airborne pollen can be predicted using only the meteorological conditions of multiple previous years was developed using regression model.

Twenty years of airborne birch pollen data (1996–2015) from Hokkaido were used to construct and validate the proposed model. Moreover, from the forecast results, statistical evaluations were used to determine the most appropriate source month for the meteorological data, the most suitable number of years for these observations, and the optimal combination of meteorological elements necessary for accurate forecasts.

3.2 Materials and Methods

3.2.1 Study area

A field experiment was conducted at the Hokkaido Institute of Public Health (43°4’55’’N, 141°20’1’’E; elevation 25 m a.s.l.) in Sapporo, the largest city on the northern Japanese island of Hokkaido (Fig. 3- 1). Because of its coastal location, Sapporo has a humid continental climate with an average annual precipitation of about 1100 mm and a mean annual temperature of about 8.5 °C. In spring, flowers bloom as the winter snow melts. Owing to the wide distribution of birch trees, the period from April to June is known as the tree pollen season, during which the occurrence of pollinosis is prevalent (60% of regional pollinosis cases) (Gotoda et al., 2001; Kobayashi et al., 1998; Takeuchi et al., 1999; Wagatsuma et al., 1972).

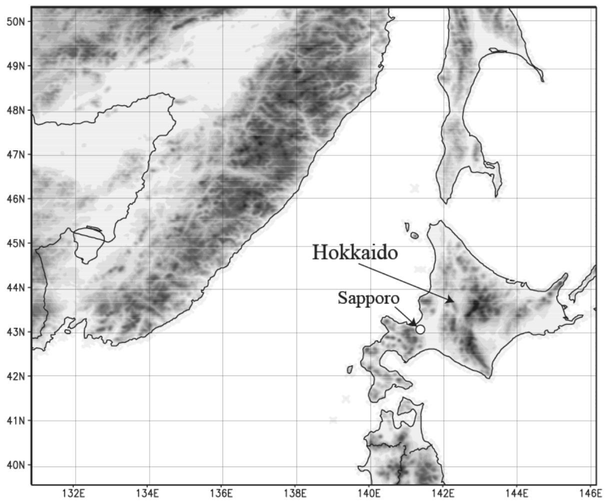


Fig. 3- 1 Location of the experimental observation site (Hokkaido Institute of Public Health) in the city of Sapporo on the northern Japanese island of Hokkaido.

3.2.2 Data

In this study, pollen samples were obtained daily by using a Durham gravitational trap placed on a roof at the Hokkaido Institute of Public Health during April–June from 1996–2015. The sampling device collected airborne pollen on petrolatum-coated glass slides. The collected samples were dyed with gentian violet and glycerin, and the daily pollen concentration were counted under a microscope; observed pollen data represent the amount of deposited pollen grains (grains cm^{-2} d^{-1}). Although I distinguished birch pollen from other species and other types of particles, all species within the genus *Betula* L. were considered a single group because of the difficulty of distinguishing different species based on morphological criteria.

The daily pollen concentration during the pollen season were summed to obtain the annual airborne pollen.

Meteorological data for the summer months (May–August) were obtained from a station within the Japanese Automated Meteorological Data Acquisition System (AMeDAS). The AMeDAS station observed the following five meteorological elements: air temperature (°C), sunshine duration (h), solar radiation (W m^{-2}), precipitation (mm), and relative humidity (%). Sunshine duration is defined as the time period during which direct solar irradiance exceeds 120 W m^{-2}. All data were collected at hourly intervals.

3.2.3 Principle of the forecasting model

The quantity of male flowers has a close relationship with the physiological growth conditions of the previous summer (Dahl and Strandhede, 1996) and the effects of masting behavior. Therefore, the quantity of male flowers in the current year i (F_i) can be formulated by considering the meteorological conditions of the previous year (M_{i-1}) and the quantity of male flowers in the previous year (F_{i-1}) as follows:

$$F_i = aM_{i-1} + bF_{i-1} + c, \tag{3-1}$$

where a, b, and c are constant coefficients. However, the difficulty in counting the number of male flowers creates low reliability for F_{i-1}, which results in uncertainty for the value of F_i in 3-1. Furthermore, meteorological conditions are more representative of a region than the number of male flowers. Therefore, I replaced F_{i-1} in the following equation:

$$F_{i-1} = aM_{i-2} + bF_{i-2} + c. \tag{3-2}$$

By substituting 3-2 into 3-1, I derived a recurrence formula by repeating further substitution:

$$F_i = aM_{i-1} + b(aM_{i-2} + bF_{i-2} + c) + c$$
$$= aM_{i-1} + baM_{i-2} + b^2F_{i-2} + bc + c$$
$$= aM_{i-1} + baM_{i-2} + b^2(aM_{i-3} + bF_{i-3} + c) + bc + c$$

$$F_i = a_1M_{i-1} + a_2M_{i-2} + a_3M_{i-3} + a_4M_{i-4} + \cdots\cdots, \qquad (3\text{-}3)$$

where a_i are the constant coefficients. This means I can obtain the quantity of male flowers by considering only the weather conditions of previous years.

3.2.4 Forecasting models: number of observation years (Model 1-6) and combinations of explanatory variables (A-F)

Previous research has indicated that the number of catkins is proportional to the total amount of dispersed pollen (Ranta et al., 2008; Yasaka et al., 2009). Thus, I used 3-3 to develop two principal types of forecasting model (basic and extended) for annual airborne pollen, using related meteorological factors as regular attributes and the annual airborne birch pollen as the label attribute. A flow chart of the data processing methods is shown in Fig. 3- 2.

Basic models

By extending the number of observation years (from one year to six years), I established six models (Model 1 to Model 6). Model 1 forecasts the annual airborne pollen by referencing the meteorological conditions of the previous year, Model 2 was based on the meteorological conditions of the previous two years, Model 3 on the previous three years, and so on. Each basic model consisted of two meteorological factors: temperature (T) and solar (S). These were used as explanatory variables because of their close relationship with male flower formation.

The regression equation for Models 1 to 6 is written as:

$$Y = \sum_{i=1}^{n} (a_i T_i + b_i S_i) + c \,, \tag{3-4}$$

where Y is the annual airborne pollen and n is the number of years of observation.

The temperature factor comprises three variables: daily mean air temperature (T_{mean}, °C), daily maximum air temperature (T_{max}, °C), and daily minimum air temperature (T_{min}, °C). The solar factor consists of two variables: daily solar radiation (SR, MJ m^{-2}) and sunshine duration (SD, h). Here, air temperature and solar radiation are monthly average values, and sunshine duration is the monthly summed value. Each variable of the temperature factor was combined with each variable of the solar factor to obtain six combinations of these factors in the basic models, named A, B, C, D, E, and F after the six combinations (Table 3- 1). These combinations were the foundations of the sub-models under each main model. Therefore, I had 6 (Models) × 6 (combinations) = 36 forecasting models.

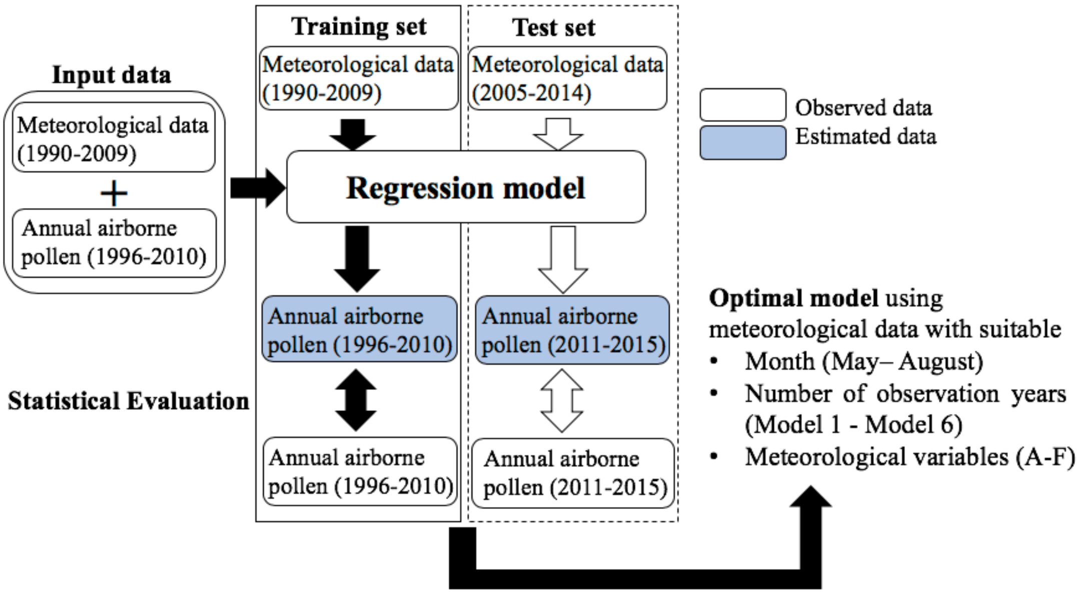


Fig. 3- 2 Flow chart of data processing methods used to estimate annual airborne pollen from meteorological data.

Extended models

To improve the predictive performance, I developed more complicated models by incorporating humidity factors into the basic models. These models were extended by integrating the humidity factors into Models 1–3. By adding the monthly total precipitation (P, mm), monthly average relative humidity (H, %), and both together (PH), I obtained three different models under each sub-model:

$$Y = \sum_{i=1}^{n}(a_i\, T_i + b_i S_i + c_i P_i) + d, \tag{3-5}$$

$$Y = \sum_{i=1}^{n}(a_i\, T_i + b_i S_i + c_i H_i) + d, \tag{3-6}$$

$$Y = \sum_{i=1}^{n}(a_i\, T_i + b_i S_i + c_i P_i + d_i H_i) + e, \tag{3-7}$$

where Y is the annual airborne pollen and n is the number of observation years.

3.2.5 Statistical indices for model performance

Having obtained the forecast results from the regression equation of each model, I undertook a statistical evaluation to assess model performance using root mean square error (RMSE) and coefficient of determination (r^2). The former shows the standard deviation of the difference between the observed and predicted values, while the latter provides a measure of how well the model replicates the observed outcomes (i.e., goodness of fit). Three types of evaluation were applied, according to the period considered: regression, forecast, and estimation. Regression evaluations were implemented to determine how well the observed pollen data of the training set (1996–2010) fitted the models by comparing the estimated results with the observed pollen data. Forecast evaluations were undertaken to assess the strength and utility of a predictive relationship by examining the difference between the forecast results and observed pollen data of the test data set (2011–2015). Estimation evaluations were performed to establish the overall (1996–2015) performance of each model. Hence, there were 3 (evaluation types) × 2 (statistical indices) = 6 evaluation indices. It should be noted that values of adjusted-r^2 were calculated only in the regression evaluations. Among the six evaluation indices, I especially focused on adjusted-r^2 in regression evaluation and RMSE in forecast evaluation for judging model performance, as they are applicable criteria for model selection (Phillips, 1995; Raftery et al., 1997).

Table 3- 1 The six combinations of temperature factor (T) and solar factor (S) in each basic model (SD: sunshine duration, SR: solar radiation).

Sub-models	Independent variables
A	T_{mean} and SD
B	T_{max} and SD
C	T_{min} and SD
D	T_{mean} and SR
E	T_{max} and SR
F	T_{min} and SR

3.3 Results

3.3.1 Temporal change in the annual airborne birch pollen

Figure 3-3 presents the temporal changes in annual airborne birch pollen from 1996–2015. The pollen concentration in this period were highly variable, fluctuating between 141 and 3299 grains cm^{-2}, with an annual mean value of 1146 grains cm^{-2}. In addition, high pollen dispersal in one year was always followed by lower levels of airborne pollen the following year, irrespective of the suitability of the meteorological conditions for male flower production in the previous year.

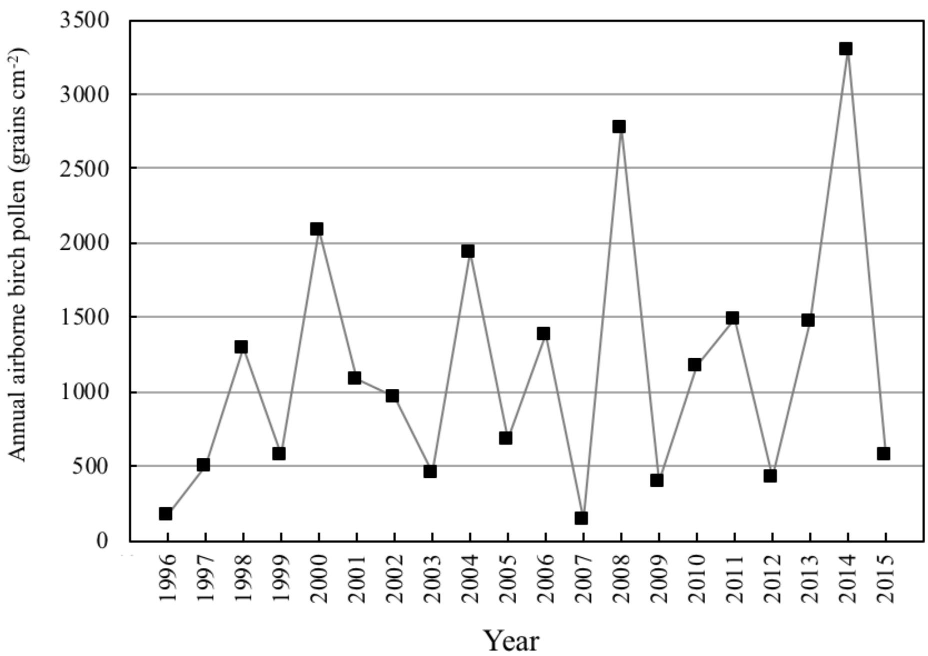


Fig. 3- 3 Temporal variation of annual airborne birch pollen from 1996–2015 at the study site.

3.3.2 Selection of the seasonal period used for meteorological data analysis

The explanatory variables adopted in the forecasting models were the monthly data of meteorological elements. In order to establish the most suitable period of meteorological data, I used r^2 as an index with which to examine the goodness of fit of each basic model. Figure 3-4 shows the average regression r^2 (adjusted-r^2) of each basic model built with meteorological data from the summer months (May–August). It is evident that the forecasting models achieved the highest accuracy when using meteorological data from June (values of r^2 of the analysis results in other months were <0.3). This result corresponds to the physiological fact that meteorological conditions in June are especially important for male flower production. Hence, all further results shown in this paper are based on models built using meteorological data from June.

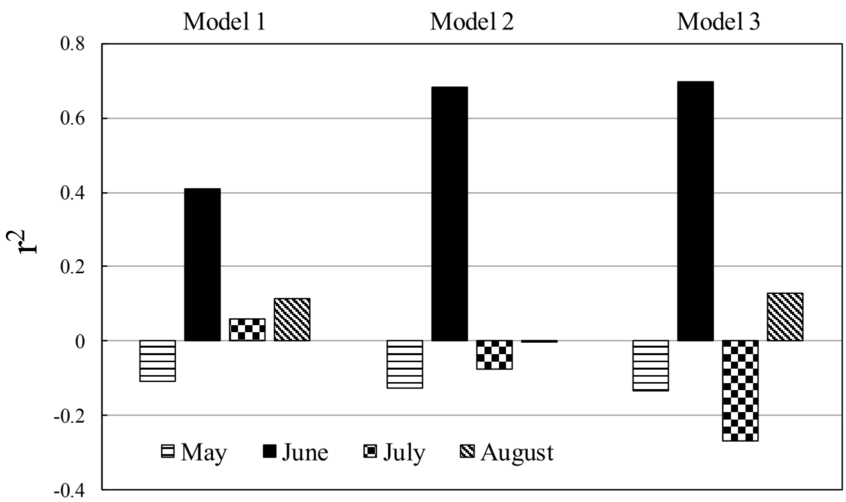


Fig. 3- 4 Average regression r^2 (adjusted-r^2) results for each basic model built with meteorological data for the summer months (May–August).

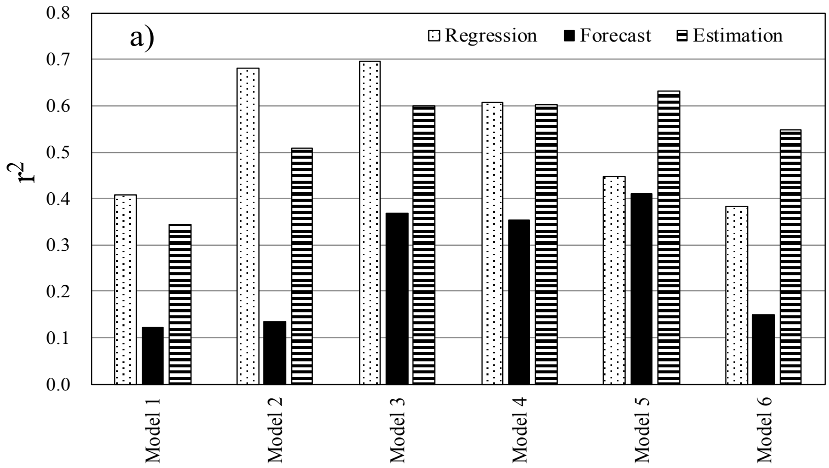
3.3.3 Performance evaluation of basic models

To select the most appropriate period of observation to predict the annual airborne pollen, I performed a statistical evaluation of each sub-model of each basic model. The results of the regression, forecast, and estimation evaluations are shown in Fig. 3- 5. The statistical values for each model were averaged over the six sub-models. The regression r^2 showed an increasing tendency from Model 1 to Model 3 and a decreasing tendency from Model 3 to Model 6 (Fig. 3- 5a). As the value of the adjusted-r^2 is not susceptible to the number of explanatory variables, this trend shows that the meteorological conditions of the previous three years are important for male flower formation. The trend of the estimation r^2 depicted a similar pattern of increase from Model 1 to Model 3. However, it remained at the same level for Model 4, reached its highest value for Model 5, and subsequently declined for Model 6. The forecast r^2 also presented a rising trend from Model 1 to Model 3, reached a peak for Model 5, and declined from Model 5 to Model 6.

The regression RMSE steadily declined from Model 1 to Model 6, a pattern attributable to the increase in the number of explanatory variables (Fig. 3- 5b). The forecast and estimation RMSEs showed the lowest values for Model 5. Model 1 did not fit the observed or forecast data well, but Models 3 and 4 showed better performances than most other models, with higher regression r^2 and lower forecast RMSE values. Model 5 had the lowest values for the forecast and estimation RMSEs, meaning that it had the best predictive performance of all the models.

The above results for the relationship between model performance and the number of observation years show that forecasting models built using the meteorological conditions in the previous three (Model 3), four (Model 4), and five (Model 5) years achieved the best performance. Analysis of the statistical data for these models revealed the optimum combination of meteorological elements for the prediction of the pollen amounts (Table 3- 2). For Model 3, Model 3E had the highest regression r^2 (0.78), forecast r^2 (0.95), and estimation r^2 (0.77) and the lowest regression RMSE (260 grains cm^{-2}), forecast RMSE (675 grains cm^{-2}), estimation RMSE (406 grains cm^{-2}). For Model 4, Model 4E had the highest regression r^2 (0.71) and estimation r^2 (0.77) and the lowest regression RMSE (256 grains cm^{-2}), forecast RMSE (686 grains cm^{-2}), and estimation RMSE (409 grains cm^{-2}), although Model 4D had the highest forecast r^2 (0.84). For Model 5, Model 5E had the highest regression r^2 (0.62), forecast r^2 (0.93), and estimation r^2 (0.83) and the lowest regression RMSE (241 grains cm^{-2}), forecast RMSE (592 grains cm^{-2}), and estimation RMSE (362 grains cm^{-2}). Thus, combination E (i.e., T_{max} and solar radiation) produced the best estimation results. Figure 3-6 shows the full forecast results for Model 5E.

As stated above, Model 3E had the best performance with six input variables and achieved the highest regression r^2 (0.78) and a low forecast RMSE (675 grains cm^{-2}) when built with T_{max} and solar radiation for the previous three years. Model 5E, which was built with 10 input variables and the same combination of meteorological elements, also showed accurate results with a regression r^2 of 0.62 and the lowest forecast RMSE (592 grains cm^{-2}); it produced similar temporal variation to that in Model 3E. These results clarify that a model with high predictability can be constructed using meteorological conditions from previous years.



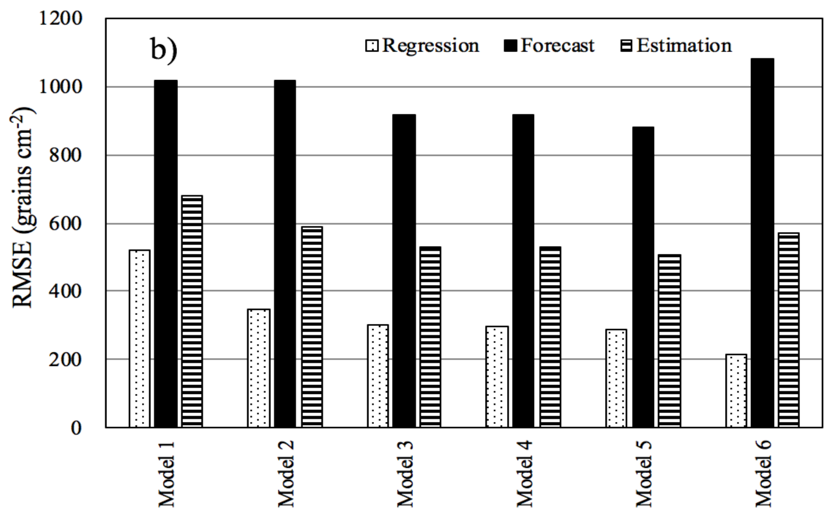


Fig. 3- 5 Performance evaluation of each model: (a) average value of regression r^2, forecast r^2, and estimation r^2; (b) average value of regression RMSE, forecast RMSE, and estimation RMSE.

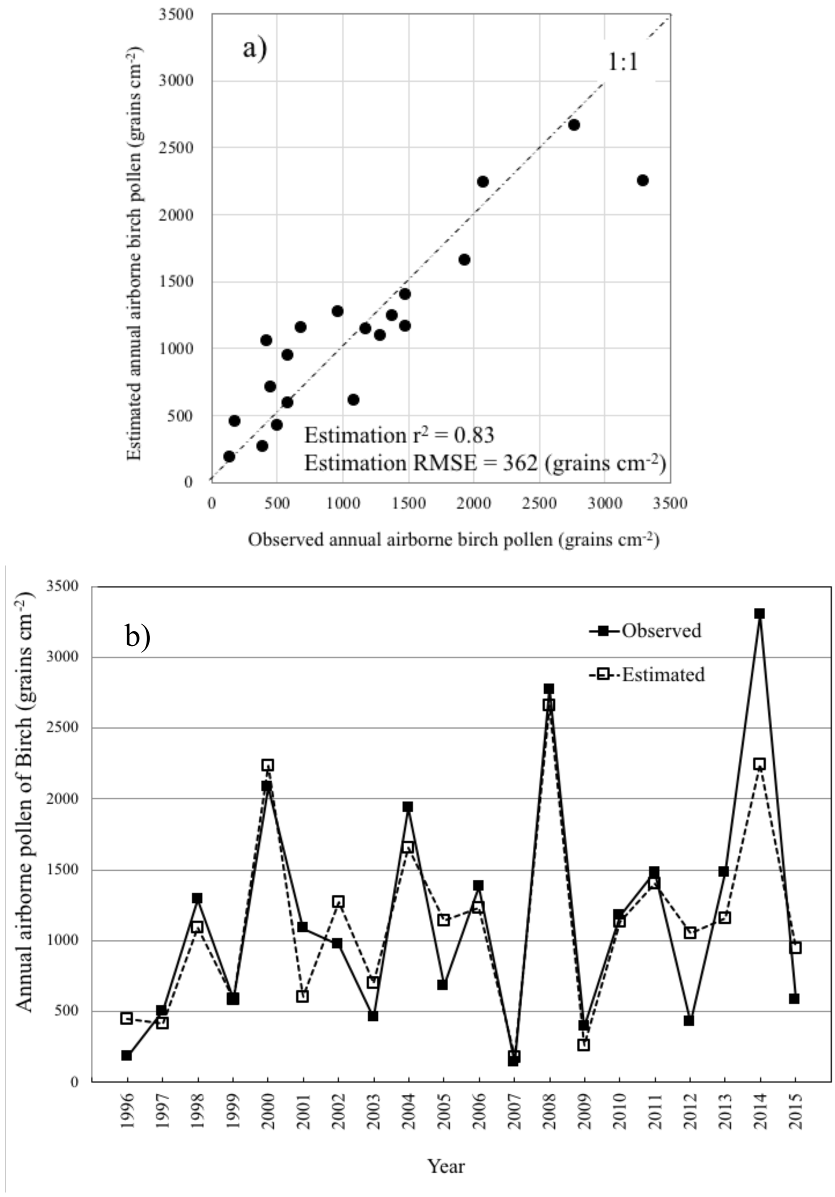


Fig. 3- 6 Forecast results for Model 5E (period of the training data set: 1996–2010; period of the test data set: 2011–2015): (a) scatter plot between observed data and forecast results; (b) annual changes in observed data and forecast results.

Table 3- 2 Model performance evaluation (r^2 and RMSE (grains cm^{-2})) for each basic model (regression, forecast, and estimation).

		Regression		Forecast		Estimation	
		Adjusted r^2	RMSE	r^2	RMSE	r^2	RMSE
Model 3	A	0.71	297	0.08	1058	0.52	588
	B	0.70	303	0.14	990	0.55	560
	C	0.75	275	0.13	1036	0.55	570
	D	0.67	320	0.76	793	0.67	484
	E	0.78	260	0.95	675	0.77	406
	F	0.57	361	0.16	963	0.53	574
Model 4	A	0.62	294	0.08	1038	0.53	578
	B	0.60	302	0.14	977	0.56	554
	C	0.68	269	0.11	1013	0.56	558
	D	0.57	315	0.84	797	0.68	483
	E	0.71	256	0.83	686	0.77	409
	F	0.46	353	0.12	1002	0.51	587
Model 5	A	0.47	285	0.15	994	0.57	555
	B	0.42	298	0.19	946	0.59	539
	C	0.56	261	0.19	985	0.60	542
	D	0.38	307	0.86	719	0.73	447
	E	0.62	241	0.93	592	0.83	362
	F	0.24	342	0.13	1060	0.49	607

3.4 Discussions

Dahl and Strandhede (1996) indicated that male flower development is regulated by the photosynthetic energy produced from May to late June, while Yli-Panula et al. (2009) reported a high correlation between annual airborne birch pollen and both monthly temperature and rainfall in June of the previous year. In this study, I aimed to determine which month has the greatest influence on male flower formation during the development period. Our forecasting models achieved the highest accuracy when using meteorological data from June, while meteorological data from other months produced much poorer results.

It might be considered inappropriate to predict the annual airborne pollen based only on the meteorological conditions of the previous year, given the low performance shown by Model 1 (all regression r^2 values <0.55). This deficiency in single-year models may be associated with the production fluctuations resulting from masting behavior, which also influences fluctuations in annual airborne pollen over different years.

Our results showed that model performance increased as the number of observation years increased from one to five, although performance decreased in the sixth year. These results illustrate the importance of multi-year meteorological conditions on plant growth. The optimal forecasting model was chosen by statistical analysis of the regression (r^2) and forecasting error (forecast RMSE) for each model. Model 3E had the highest regression r^2 and a low forecast RMSE with six variables, and Model 5E had a high regression r^2 value and the lowest forecast RMSE with 10 variables.

The principal idea behind this study (replacing the quantity of male flowers with data on meteorological conditions during previous years) was validated by forecast results from the models developed. As water stress has been identified as an important environmental factor regularizing masting behavior (Abrahamson and Layne, 2003; Espelta et al., 2008; Piovesan and Adams, 2001), precipitation and relative humidity were added to improve the models. Statistical values for each model were averaged over six sub-models; the results of the regression r^2 and forecast RMSE are shown in Fig. 3- 7 for precipitation, relative humidity, and both factors together. In both Model 2 and Model 3, the performances of both indices increase compared with the basic model when relative humidity is taken into consideration. A detailed performance evaluation (r^2 and RMSE) of each model incorporating humidity is shown in Table 3- 3. Considering the actual forecasting ability, the most important index affecting model performance is RMSE in forecast evaluation. The optimum performance (forecast RMSE = 593 grains cm^{-2}) occurred when the average daily maximum air temperature, solar radiation, and relative humidity of the previous three years were incorporated (Model 3EH, Fig. 3- 8). This finding is in agreement with previous research indicating that three years' weather data can provide a robust data set for the simulation of long-term mean yields for major cereal crops (Van Wart et al., 2015).

Although most observed and estimated data match up well in each year, clear mismatches occurred from 2011–2015; I suggest several potential reasons for this. First, errors may be caused by assumptions within the forecasting model. Although I rationally supposed that the amount of male flowers is proportional to the annual airborne pollen, the total dispersal amount is also affected by the temporal weather conditions during the peak period. Second, errors may be caused by the spatial representativeness of the observed pollen concentration in the subject area. Unusual peak or bottom values for pollen concentration might result from poor spatial representativeness. Third, the size of training data is relatively small for building the models, Model 5E and Model 3EH, which have the lowest forecast RMSE. These matters remain problematic and should be resolved in future research.

Table 3- 3 Model performance evaluation (r^2 and RMSE (grains cm^{-2})) for each model incorporating humidity (regression, forecast, and estimation). Humidity factors are precipitation (P), relative humidity (H), and both together (PH).

		Regression		**Forecast**		**Estimation**	
		Adjusted r^2	RMSE	r^2	RMSE	r^2	RMSE
Model 3 +Prec.	AP	0.59	278	0.07	1092	0.50	597
	BP	0.58	282	0.08	1081	0.51	593
	CP	0.61	272	0.12	1002	0.57	553
	DP	0.50	310	0.87	840	0.65	498
	EP	0.65	257	0.96	699	0.77	414
	FP	0.37	346	0.13	952	0.54	563
Model 3 +RH	AH	0.92	123	0.46	804	0.76	416
	BH	0.90	137	0.49	762	0.78	399
	CH	0.78	206	0.33	951	0.67	508
	DH	0.90	137	0.50	755	0.78	396
	EH	0.93	118	0.67	593	0.86	314
	FH	0.70	239	0.38	978	0.64	531
Model 3 +Prec. +RH	APH	1.00	15	0.52	1010	0.74	505
	BPH	0.98	38	0.51	1012	0.74	507
	CPH	0.95	60	0.40	1306	0.64	655
	DPH	0.88	96	0.52	730	0.80	374
	EPH	0.98	39	0.63	673	0.84	338
	FPH	0.80	122	0.55	891	0.74	458

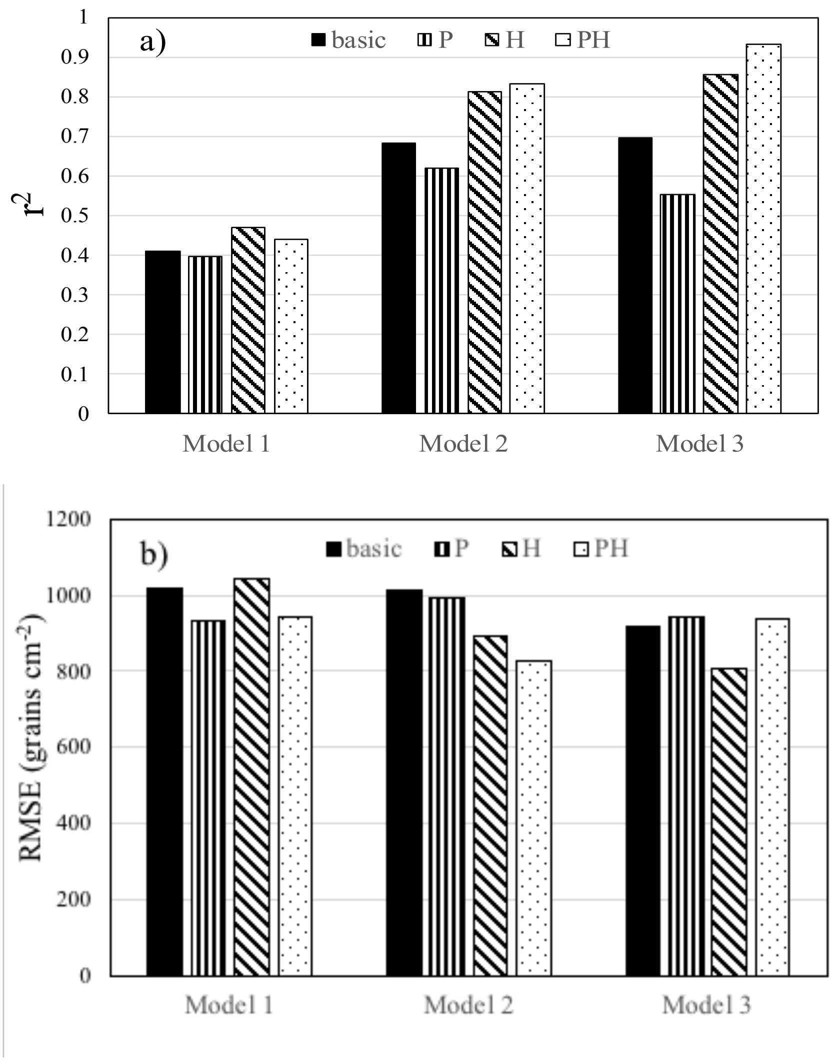


Fig. 3- 7 Performance evaluation of each model incorporating humidity: (a) average regression r^2 and (b) average forecast RMSE (P: precipitation, H: relative humidity, and PH: both together).

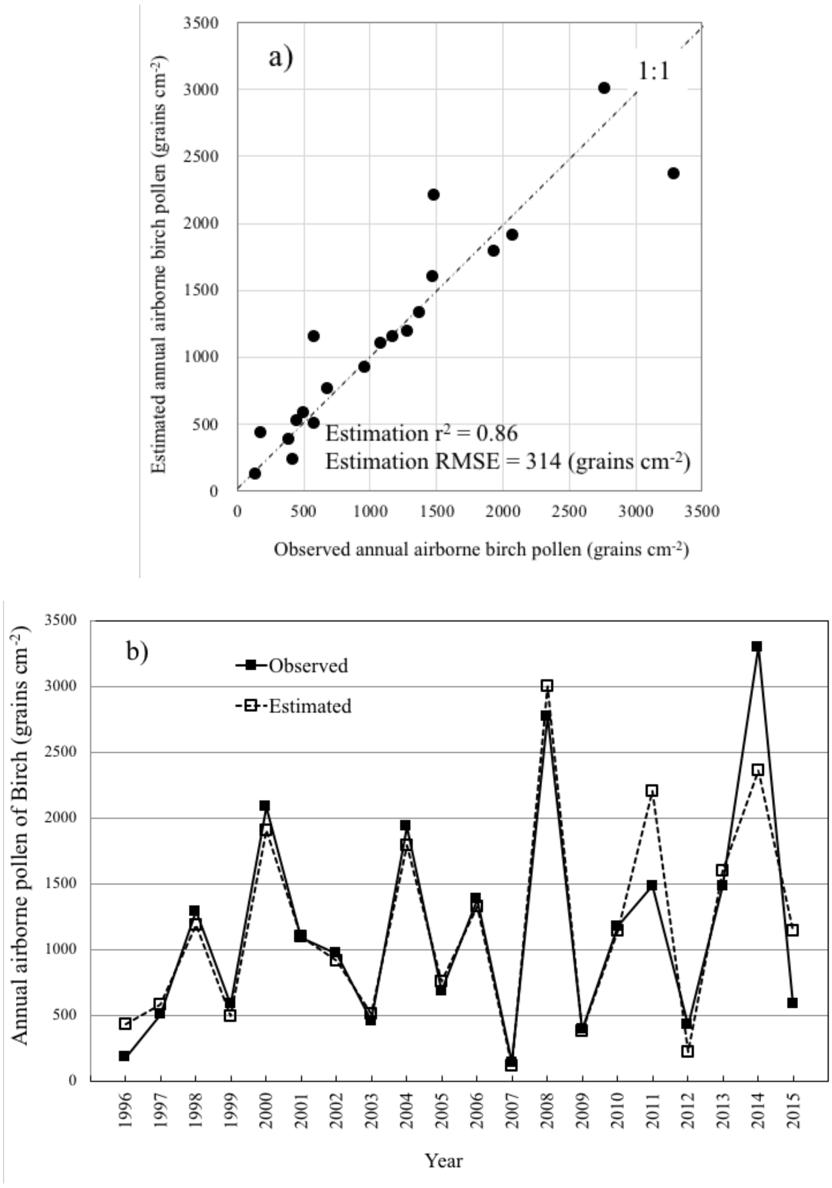


Fig. 3- 8 Forecast results for Model 3EH (period of the training data set: 1996–2010; period of the test data set: 2011–2015): (a) scatter plot between observed data and forecast results; (b) annual changes in observed data and forecast results.

3.5 Conclusions

Based on observational data and a series of forecasting tests, I confirmed that weather conditions in June are crucial for male birch flower production and that the meteorological conditions in June of the previous year are suitable for predicting the amount of birch pollen in the following year. However, I also found it insufficiently accurate to estimate the annual airborne pollen using only meteorological data from the previous year because high pollen dispersal in one year always causes a lower level of airborne pollen in the following year, irrespective of the suitability of meteorological conditions for male flower production in the previous year. Our results showed that the basic forecasting models obtained the highest accuracy when using meteorological data from June, and the best model performance was achieved by using the average daily maximum air temperature and solar radiation from the previous five years. Furthermore, the best forecast results for the extended model were achieved using the monthly average of daily maximum air temperature, solar radiation, and relative humidity from the previous three years.

In this paper, I verified the applicability of the principle explained in Section 3.2 to the forecasting of annual airborne pollen in a season. This approach can provide a useful foundation for other studies aimed at investigating the annual amount of airborne pollen in countries for which pollinosis is a problem. However, this algorithm was only proven to be valid for a single species in a specific region at present. Future research should examine the general applicability of the forecasting model in other regions; the model performance would also be improved with data sets spanning longer time frames.

Chapter 4
Influence of the topography on spatial representativeness of airborne pollen concentration

Tseng Y-T, Kawashima S. Applying a pollen forecast algorithm to the Swiss Alps clarifies the influence of topography on spatial representativeness of airborne pollen data. *Atmos. Environ.* 2019; 212: 153-162.

4.1 Introduction

In Chapter. 3, based on the proximate-level hypothesis that annual airborne pollen can be predicted by meteorological conditions and biological conditions in the previous year, I used the recursive characteristics of the hypothesis to develop an algorithm for the prediction of annual airborne birch pollen amount using only meteorological conditions from multiple previous years with multiple regression (Tseng et al., 2018); this was validated in a flat region of Hokkaido, Japan (25 m a.s.l.).

As models are often lacking of versatility with regard to applications in other research fields (Chuine et al., 1999), it is important to examine and clarify the applicability of this forecasting model to determine its overall feasibility and the causes of any limitations. In this study, the applicability of the forecasting model to monitoring sites in mountainous areas of Switzerland, where birch trees are also the major pollen-allergen-producing tree, was analyzed. Furthermore, the topographical features around the monitoring sites were suggested to influence on the spatial representativeness of airborne pollen data and associated model applicability.

4.2 Materials and Methods

4.2.1 Study area and data

Switzerland’s landscape encompasses a wide range of topography within three basic divisions: the Jura Mountains, the Swiss Plateau, and the Alps (covering 65% of the country’s surface). The region’s climate varies significantly by location, with observed average annual temperature ranging from -7.2–12.4 °C and observed average annual precipitation ranging from 545–2837 mm. The Jura mountains, Swiss Plateau, and the lower north slopes of the Alps are situated within the transition zone between temperate oceanic and humid continental climates; these areas are mainly influenced by the high Alps and the Atlantic Ocean. In comparison, the south side of the Alps is most affected by the Mediterranean Sea, resulting in a temperate oceanic climate. Higher-altitude regions experience polar or subarctic climates.

Airborne pollen sampling data for nine Swiss cities (Table 4- 1) were acquired from the Federal Office of Meteorology and Climatology (MeteoSwiss). The concentration of airborne birch pollen grains was measured using the volumetric method with a Burkard pollen trap (Hirst design) during birch pollen season (March—May) (Gehrig et al., 2018) from 1997–2016. Meteorological data for climatic variables that facilitate male flower formation were also acquired from MeteoSwiss, covering 1991–2015. Temperature (T, °C) factors included monthly average air temperature calculated from daily mean air temperature (T_{mean}), daily maximum air temperature (T_{max}), and daily minimum air temperature (T_{min}); solar (S) factors consisted of monthly total sunshine duration ($S_{duration}$, h) and monthly mean daily solar radiation ($S_{radiation}$, MJ m^{-2}).

Table 4- 1 Geographical information for nine observation sites in Switzerland (sorted by elevation).

	Basel	**Lugano**	**Locarno**	**Genève**	**Luzern**	**Neuchâtel**	**Zürich**	**Lausanne**	**Bern**
Longitude	7°58'E	8°96'E	8°79'E	6°15'E	8°3'E	6°95'E	8°57'E	6°64'E	7°42'E
Latitude	47°56'N	46°0'N	46°17'N	46°19'N	47°6' N	47°0''N	47°38' N	46°52' N	46°95'N
Elevation (a.s.l.)	273 m	273 m	366 m	380 m	460 m	490 m	556 m	570 m	570 m

4.2.2 Algorithm for forecasting annual airborne pollen using meteorological data

Masaka and Maguchi (2001) and Ranta et al. (2008) emphasized the importance of considering both the physiological growth conditions of the male flower and the effects of masting behavior for annual airborne pollen prediction of anemophilous trees to achieve robust results. The annual airborne pollen in year n can be therefore estimated based on meteorological conditions in summer in the year n-1 and the amount of male flowers in n-1.

Our model considered annual airborne pollen as the response variable (assuming that this was representative of the average airborne pollen concentration in a certain region) and meteorological conditions in the previous summer (M) as the explanatory variable. Temperature and sunlight were used since they have a direct influence on the amount of resources available for the induction of catkin development produced by photosynthesis (Kikuzawa, 1983; Koike, 1988). However, the number of explanatory variables was not always consistent, depending on the number of observation years. Input data information can be summarized as follows:

$$SPI = (SPI^{(1997)}, SPI^{(1998)}, SPI^{(1999)}, \dots, SPI^{(2016)})^T \tag{4-1}$$

$$M = (M^{(1997)}, M^{(1998)}, M^{(1999)}, \dots, M^{(2016)})^T \tag{4-2}$$

where $SPI^{(j)}$ is annual airborne pollen in year j, $M^{(j)}$ is $\sum_{i=1}^{m}(T_{j-i} + S_{j-i})$, and m is the number of observation years before year j.

By extending m from one year to five years, I established five models: Model 1 forecasts the annual airborne pollen by referencing the meteorological conditions of the previous year, Model 2 does the same based on the previous two years, and so on. Each temperature factor was combined with each solar factor to obtain six combinations in the basic models: T_{mean} and $S_{duration}$, T_{max} and $S_{duration}$, T_{min} and $S_{duration}$, T_{mean} and $S_{radiation}$, T_{max}

and $S_{radiation}$, T_{min} and $S_{radiation}$, renamed A, B, C, D, E, and F, respectively. These combinations were the foundations of the sub-models under each main model, producing a total of 5 models (Model 1 – Model 5) × 6 combinations (A, B, C, D, E, F) = 30 forecasting models.

4.2.3 Applicability indices for the forecasting models at observation sites

To achieve fair evaluations of the forecasting model in different regions, I followed the same principles as in our previous study (Tseng et al., 2018) by evaluating model applicability at the nine observation sites using the coefficient of determination (r^2) and root mean square error (RMSE). The former provides a measure of how well the model replicated the observed variation pattern while the latter shows the standard deviation of the concentration errors forecasted by the model. According to the period considered, three types of evaluation were applied: regression, estimation, and forecast evaluation (values of adjusted r^2 were calculated only in the regression evaluation). The regression r^2 and forecast RMSE were noted in the analysis process due to their ability to quantify explanatory power and predictive power directly. In this paper, r^2 and RMSE represent maximum regression r^2 and minimum forecast RMSE, respectively, among the 30 sub-models at an observation site if no specific annotation states otherwise.

4.3 Results

4.3.1 Annual variation in annual airborne pollen

The averages and standard deviations of observed annual airborne pollen from 1997–

2016 at the nine observation sites (Fig. 4- 1**)** show that annual airborne pollen values in Lugano, Locarno, Zürich, and Bern were highly variable, with average values from around 8000 grains m^{-3} to above 10000 grains m^{-3}. In contrast, annual airborne pollen values at other sites had lower variations with average values below 6000 grains m^{-3}.

4.3.2 Selection of optimum month of meteorological data for forecasting model

I used meteorological data from summer months (May–August) to predict annual airborne pollen for airborne birch pollen. In order to determine the optimum month for such predictions, I identified which month's meteorological data led to the highest r^2 and lowest RMSE at each site, then summed the results (Fig. 4- 2). Of the nine monitoring sites, seven recorded the highest r^2 and six recorded the lowest RMSE in June. Thus, meteorological data from June are a good fit for the forecasting model and have higher predictive power.

4.3.3 Examples of forecasting results

To eliminate factors causing uncertainty in the spatial representativeness of the observed airborne pollen data, I selected Basel and Locarno as examples of our forecasting results, as they had more positive regression r^2 in different models. Figure 4-3 shows the regression r^2 and the forecast RMSE of each model with meteorological data from different months in Basel and Locarno. In Basel (Fig. 4- 3a), meteorological data from June resulted in obviously higher r^2 in four of the five models (highest in model 4) and the lowest RMSE in model 3. In Locarno (Fig. 4- 3b), meteorological data from June resulted in obviously higher r^2 in two of the five models (two others were highest in August), while the highest r^2 for June occurred in model 5. RMSE was lowest in June for three of the five models, with the lowest overall value in model 5. Scatter plots and annual changes in observed and forecasted data from Basel (Model 3F) and Locarno (Model 5F) (Fig. 4- 4) clearly replicate the fluctuations and intermittent peaks of annual airborne pollen in forecast data at both sites despite their different annual airborne pollen ranges.

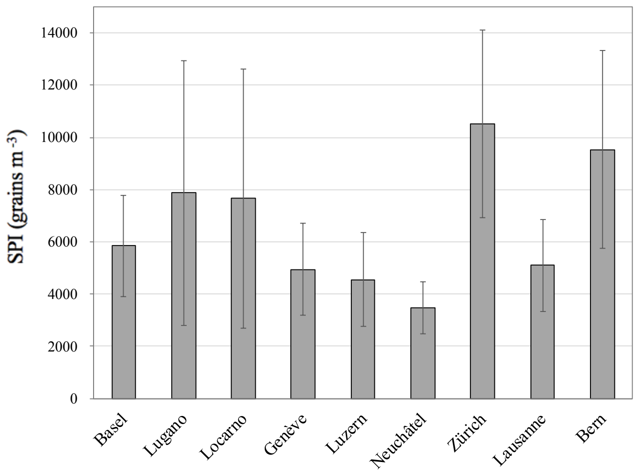


Fig. 4- 1 Averages and standard deviations of observed annual airborne pollen (SPI) from 1997–2016 at the nine observation sites.

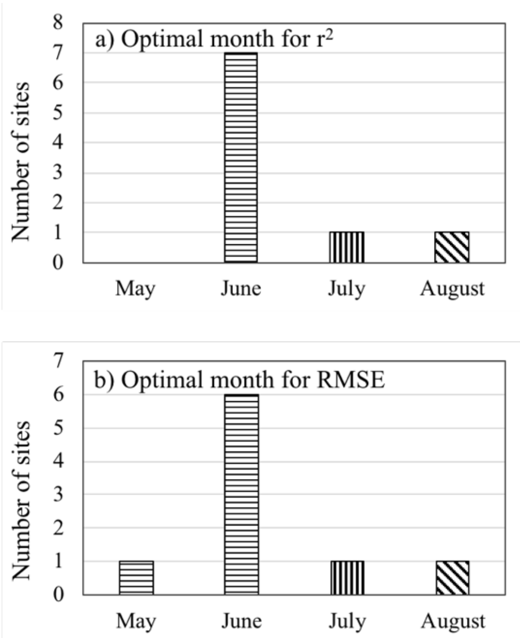


Fig. 4- 2 Number of sites with an optimal prediction result from meteorological data by month for: (a) r^2, (b) RMSE.

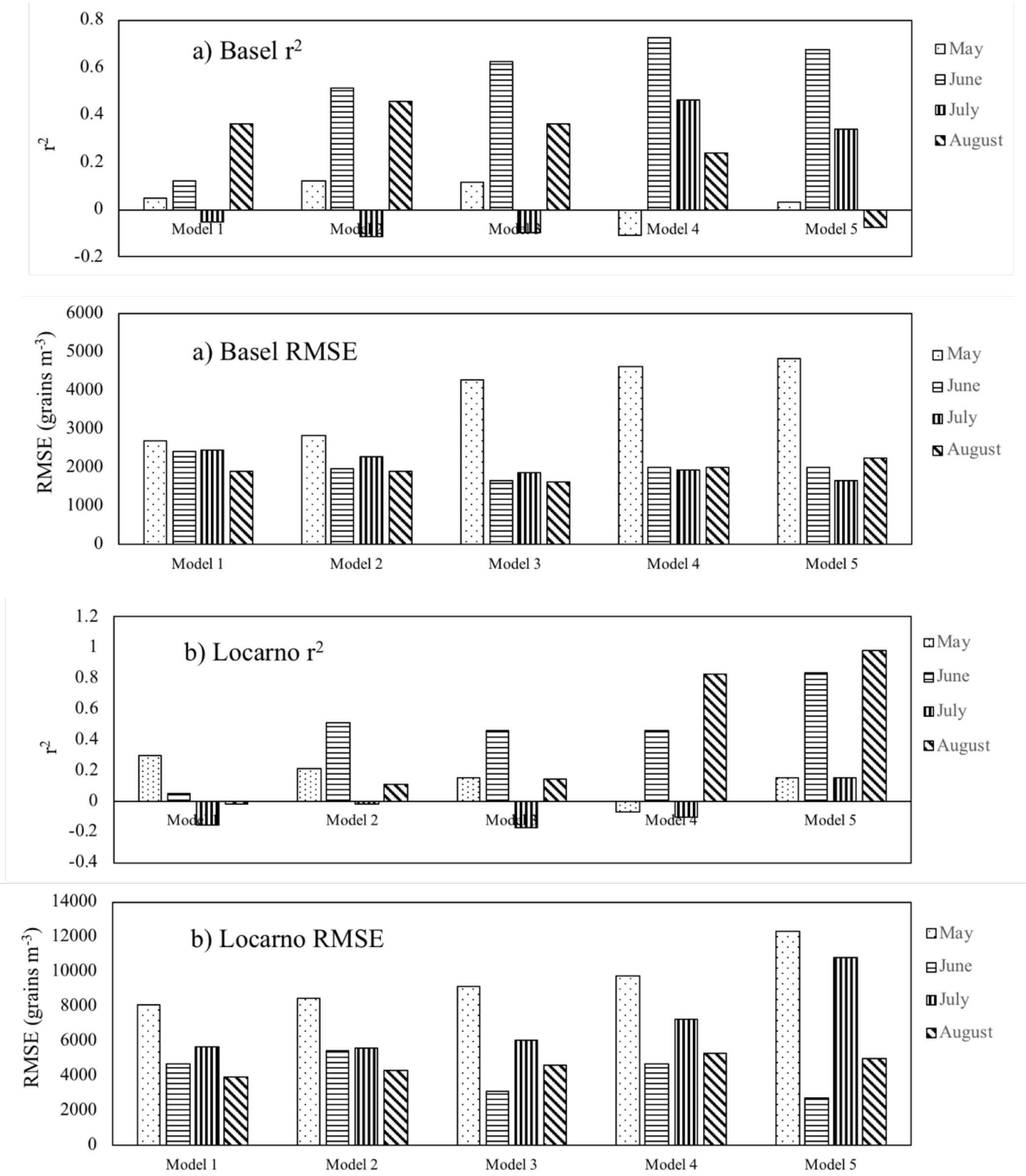
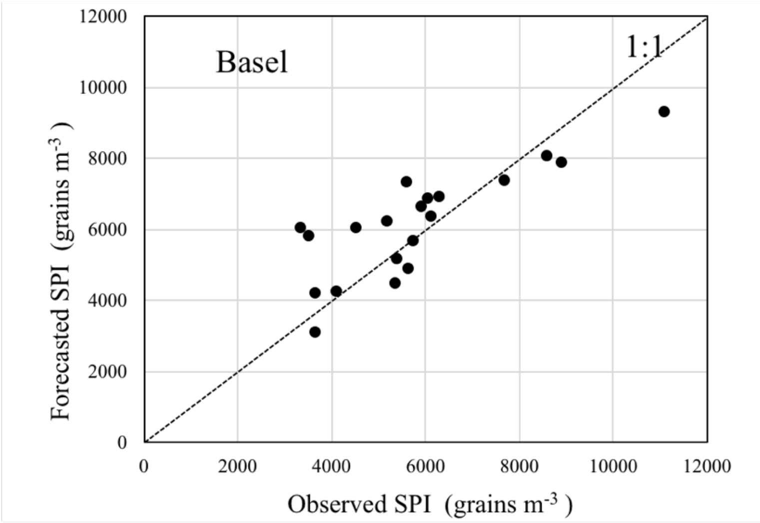


Fig. 4- 3 Regression r^2 (adjusted-r^2) and forecast RMSE for each model built with meteorological data for the summer months (May–August) for birch pollen in: (a) Basel, (b) Locarno.

(a)



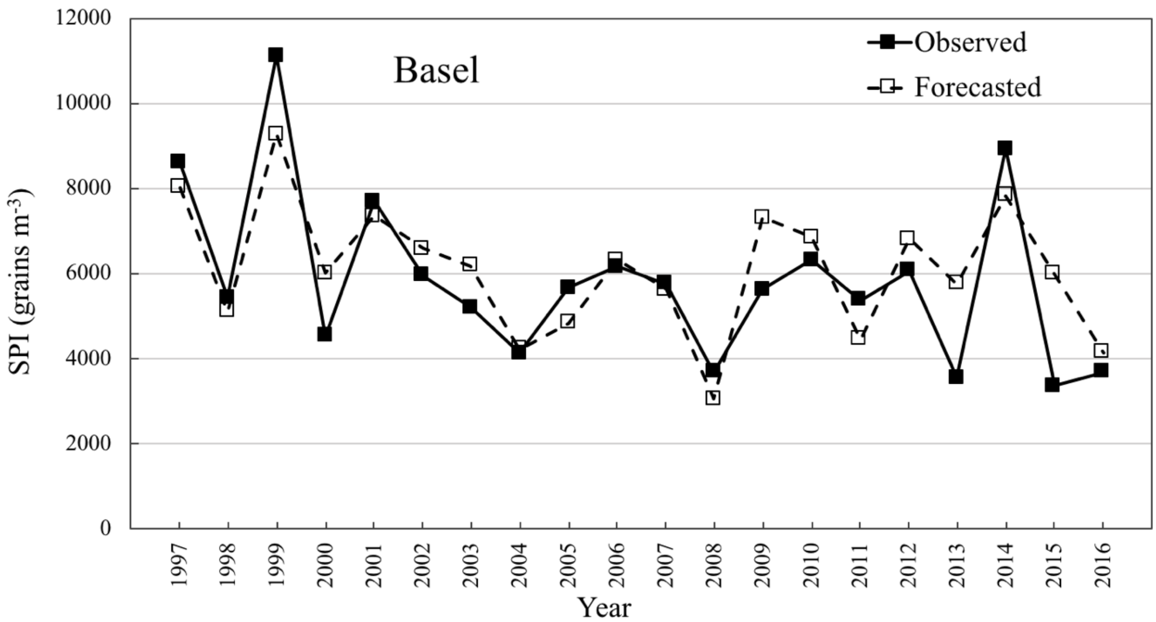
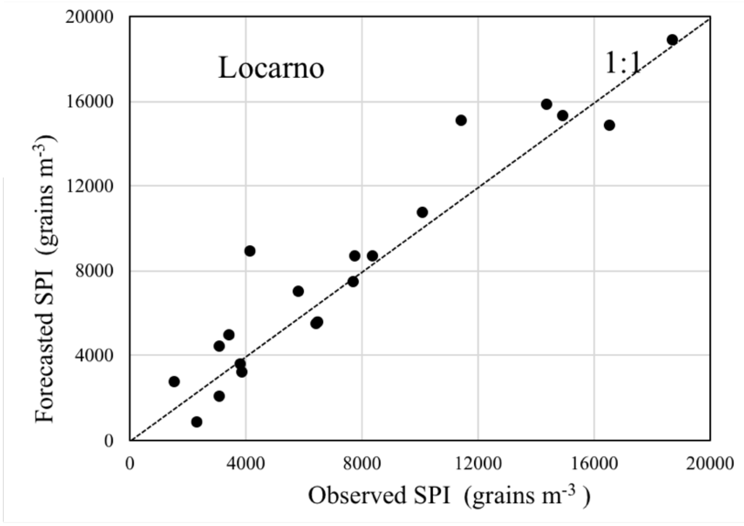


Fig. 4- 4 Scatter plots and annual changes in observed and forecasted data for (a) Basel (Model 3F), (b) Locarno (Model 5F).

(b)



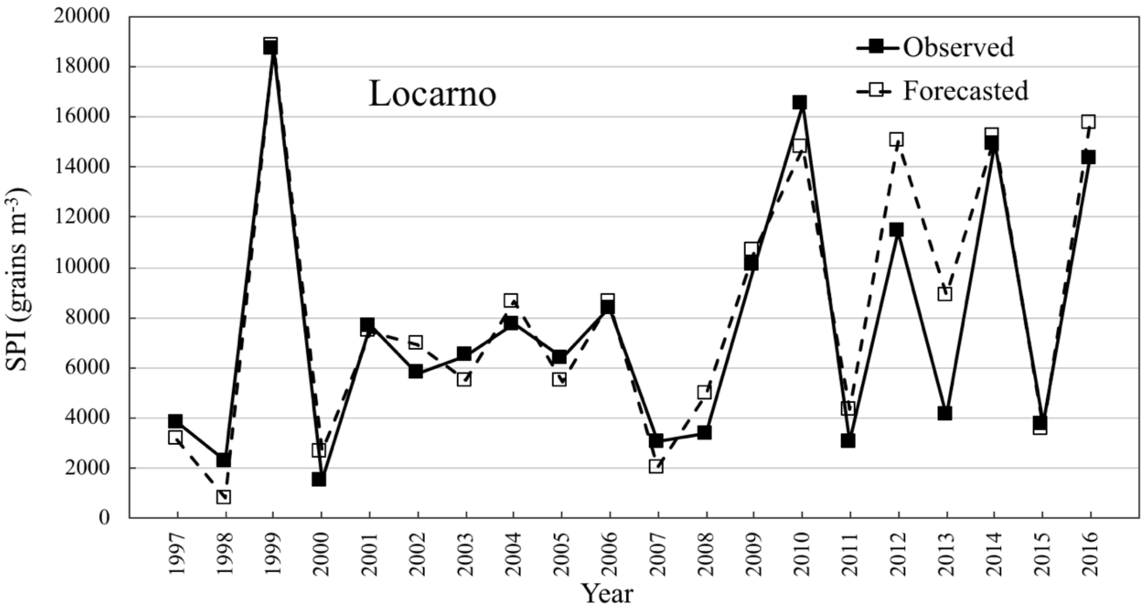


Fig. 4- 4 Scatter plots and annual changes in observed and forecasted data for (a) Basel (Model 3F), (b) Locarno (Model 5F).

4.3.4 Applicability of the forecasting models at the nine observation sites

The overall results showed that the forecast models in some regions had higher r^2 and lower RMSE, or vice versa. Further analysis showed a clear positive relationship between annual airborne pollen and RMSE for all sites except Lugano, but no clear relationship between annual airborne pollen and r^2 (Fig. 4- 5). The spatial distribution of RMSE in the study area (Fig. 4- 6) shows that sites in the west had relatively lower RMSE values.

A scatter plot of r^2 and RMSE for all nine observation sites (Fig. 4- 7) showed a clear negative trend correlating r^2 and RMSE for Zürich, Bern, Locarno, Basel, and Luzern. The forecasting models for these sites had either higher r^2 and smaller RMSE, or vice versa. Of the other four sites, Lugano plotted to the upper right with higher r^2 and RMSE values, while Luasanne, Genève, and Neuchâtel plotted to the lower left with lower r^2 and RMSE values.

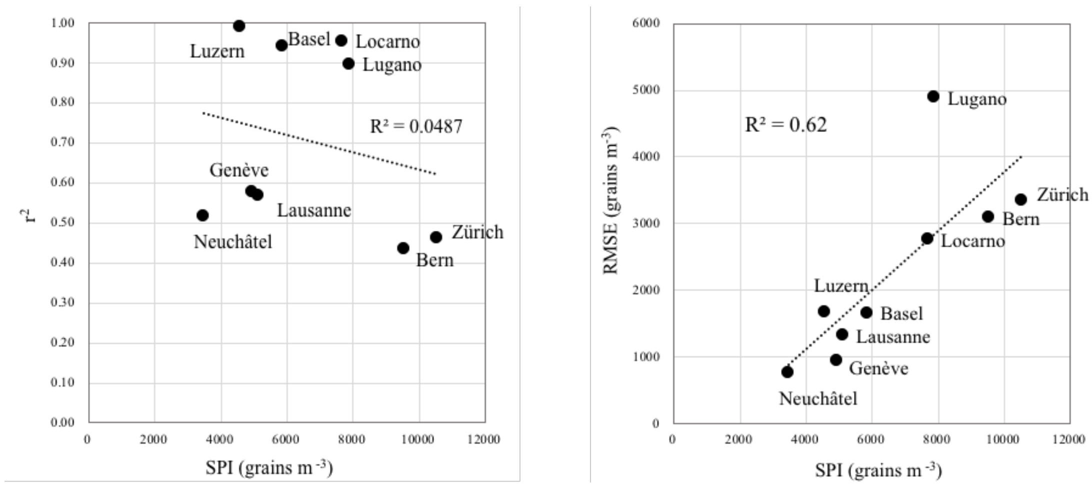


Fig. 4- 5 Scatter plot of annual airborne pollen and RMSE for the nine observation sites.

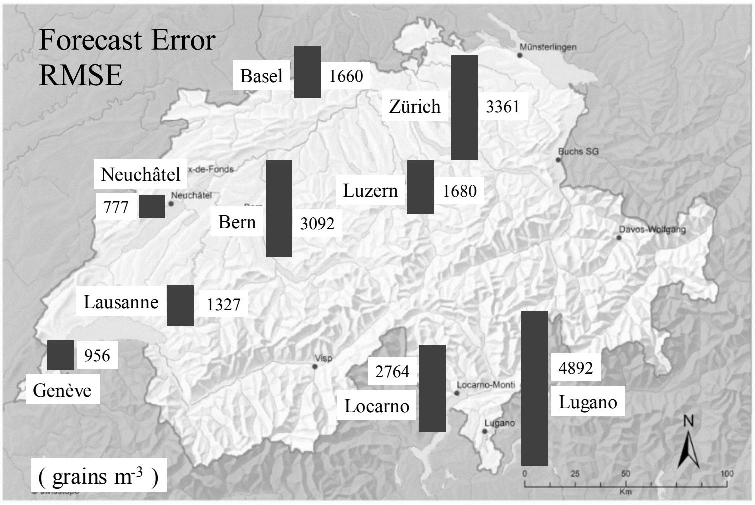


Fig. 4- 6 Spatial distribution of forecast RMSE (grains m^{-3}) for the nine observation sites.

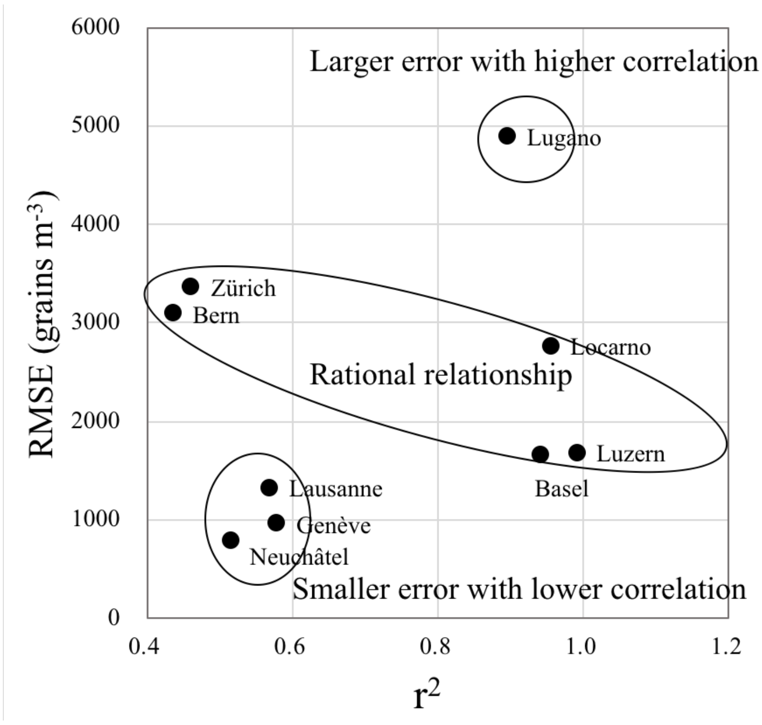


Fig. 4- 7 Scatter plot of r^2 and RMSE for the nine observation sites with relationship groupings.

4.4 Discussions

Our results showed that in some monitoring sites, the amount of airborne pollen can be predicted accurately using our previously developed algorithm with a suitable source month for meteorological data, sufficient observation years, and appropriate combinations of meteorological elements. Also, as in our previous study (Tseng et al., 2018), I found that prediction accuracy was highest when using meteorological data from June, when the initiation and growth of male catkins in *Betula spp.* occurs (Dahl and Strandhede, 1996).

The forecasting model assumes that meteorological and airborne pollen data are representative of the subject area. Based on the theory that the geospatial scale of meteorology is directly related to its timescale (Smagorinsky, 1974), I considered the meteorological data adopted in this study to be spatially representative of a wide region given their status as average or cumulative monthly data. On the contrary, the spatial representativeness of airborne pollen data remains ambiguous due to quantification difficulties. Therefore, I considered the spatial representatives of airborne pollen data to be a primary factor that can influence the applicability of the forecasting models.

By observing the prediction performance of the models at different sites, I found that sites at lower altitude (such as Basel and Locarno) had better performance than sites at higher altitude, suggesting that altitude and terrain complexity surrounding the monitoring sites may be factors influencing model performance. I therefore investigated the influence range of topography by calculating the average altitude and the standard deviation of altitude within a radius of 1–100 km around the monitoring sites in order to assess the relationship between topographical features and model performance (regression r^2 and forecast RMSE). In order to investigate the influence of topography on observed pollen concentration itself, I adopted forecast RMSE instead of normalized one. Also, indices of model performance were the average values of r^2 and RMSE for the 30 sub-models at each site to determine more general impacts of topographical features on model applicability.

4.4.1. Relationship between r^2 and topography

Variations in the correlation coefficient between mean altitude (AM) calculated from different radiuses and the average regression r^2 (Fig. 4- 8a) showed that a negative correlation between the two was especially apparent within 10 km of the observation sites; the highest negative correlation (-0.827) was found at 1 km (Fig. 4- 8b). Following reasons for this correlation between higher altitude and lower model accuracy were suggested:

(i) At higher altitudes, local topographical features tend to induce specific wind phenomena like mountain and valley winds (Oke, 2002).

(ii) Such local wind fields affect observed airborne pollen concentrations by introducing deviations from the area's average concentration.

(iii) These deviations lead to lower spatial representativeness at a local scale, so the forecasting model's accuracy decreases as the altitude increases.

This negative correlation held true except for one site (Genève) that plotted further from the regression line (Fig. 4- 8b), which may have been influenced by local environmental conditions such as urban land use. Other cases of inhomogeneous pollen distribution caused by local wind patterns have been reported (Maya-Manzano et al., 2017). In addition, a previous study in Switzerland showed that a considerable amount of pollen transport can occur between different altitudes (Gehrig et al., 2011).

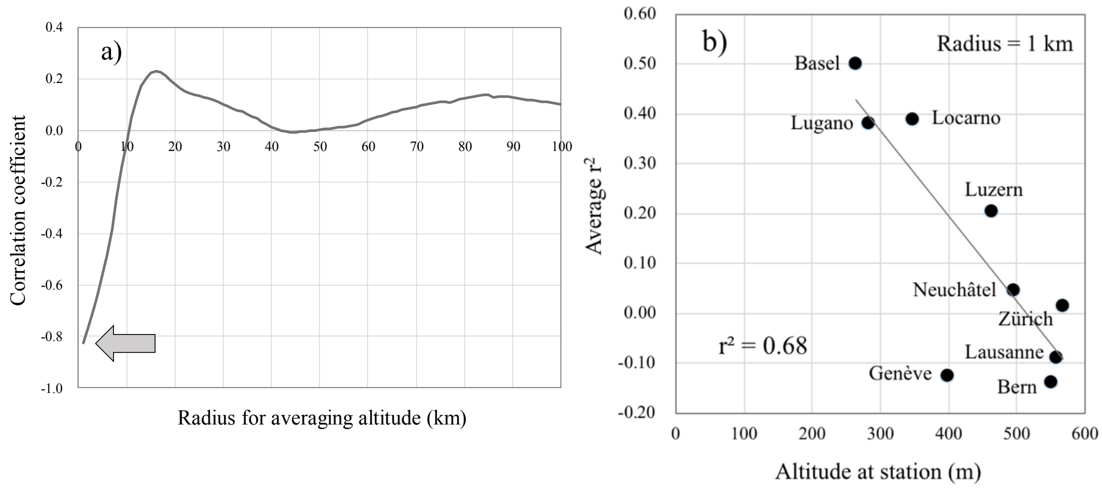


Fig. 4- 8 Relationship between r^2 and radius from site used for averaging altitude: (a) Change in correlation coefficient between the performance of the models (regression r^2) and topographical features (mean altitude), (b) Study sites plotted by r^2 and mean altitude at 1 km radius at the highest negative correlation coefficient (r = - 0.83) found in (a).

4.4.2. Relationship between RMSE and topography

Variations in the correlation coefficient between the standard deviation of altitude (SD) calculated from different radiuses and the average forecast RMSE (Fig. 4- 9a) showed a clear increasing trend of correlation from 30–100 km and a consistently high correlation value ($r >$ 0.65) from 70–100 km; the highest positive correlation (0.675) was found at 96 km (Fig. 4- 9b). I suggest the following reasons for this correlation between higher complexity of terrain and higher prediction errors:

(i) The strength of convective mixing in the Atmospheric Boundary Layer (ABL) tends to increase in meso-scale regions with higher terrain complexity (Oke, 2002).

(ii) Observed airborne pollen concentrations are affected by meso-scale pollen dispersal induced by active convective mixing, thus deviating from the average concentration in the meso-scale region.

(iii) This deviation leads to lower meso-scale spatial representativeness, such that the forecasting model's accuracy decreases as the terrain complexity increases.

This positive correlation can be further decreased by the effect of surrounding environmental conditions. For instance, the large lake adjacent to Neuchâtel and Lausanne can help explain their location below the regression line (Fig. 4- 9b).

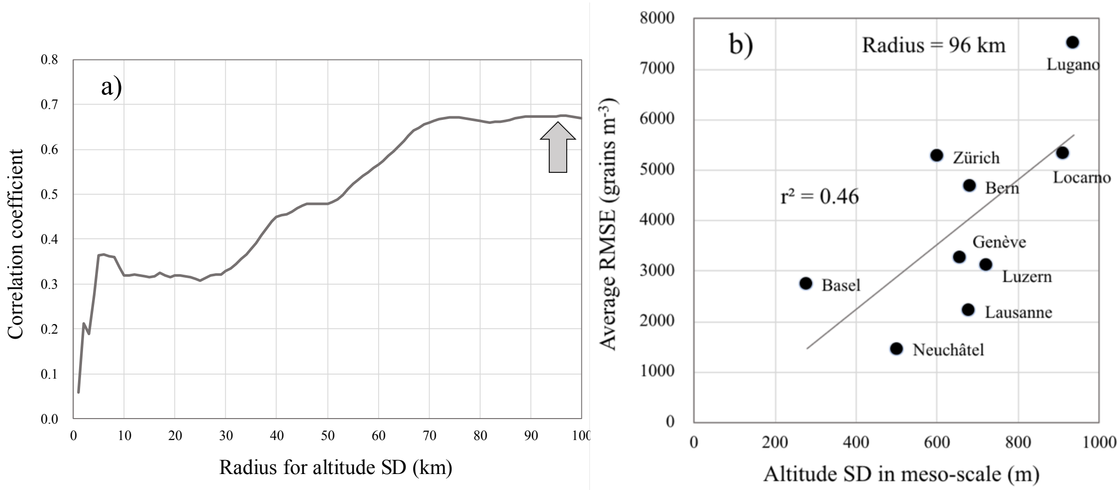


Fig. 4- 9 Relationship between RMSE and radius from site used for determining altitude standard deviation (SD): (a) Change in correlation coefficient between the performance of the models (forecast RMSE) and topographical feature (SD), (b) Study sites plotted by RMSE and SD in altitude at 96 km radius at the highest positive correlation coefficient (r = 0.68) found in (a).

Based on the assumptions discussed above, I assumed that the applicability of the forecasting models as only influenced by the spatial representatives of the airborne pollen data. The varying applicability of the models at different monitoring sites can be explained by variable topographical conditions at each site. Other studies have considered the transportation of air pollutants by local wind field or convective mixing, which can also cause heterogeneous concentrations in the air (He et al., 2014; Jazcilevich et al., 2005; Liao et al., 2017; Miao et al., 2015; Palau et al., 2005; Triantafyllou and Kassomenos, 2002). I therefore suggest that r^2 and RMSE reflect the deviations in airborne pollen concentration caused by the effect of transportation in complex terrain at local and meso-scales, respectively.

4.5 Conclusions

Considering the common lack of versatility in annual airborne pollen forecasting models, I validated the applicability of the pollen prediction algorithm developed in our previous study by utilizing pollen data from nine observation sites with different topographical characteristics in Switzerland. I found a great variation in model applicability among the monitoring sites: accurate predictions could be made at some monitoring sites by using a suitable source month for meteorological data, sufficient years of observations, and a combination of meteorological elements with representative data, but this was not the case at other sites. I suggest that the mismatches in observed and estimated data at some monitoring sites may have been caused by the influence of topographical features. At higher-altitude monitoring sites, the local spatial representativeness of airborne pollen declines due to the influence of local climate factors. On the other hand, dispersal conditions in the atmospheric boundary layer may undermine the meso-scale representativeness of airborne pollen at monitoring sites with higher terrain complexity (Fig. 4- 10). In addition to topographical features, surrounding surface types may also cause further influence on the spatial representativeness of pollen data. Our study provides a reference approach for determining the spatial representativeness of airborne pollen data by applying a pollen forecast algorithm.

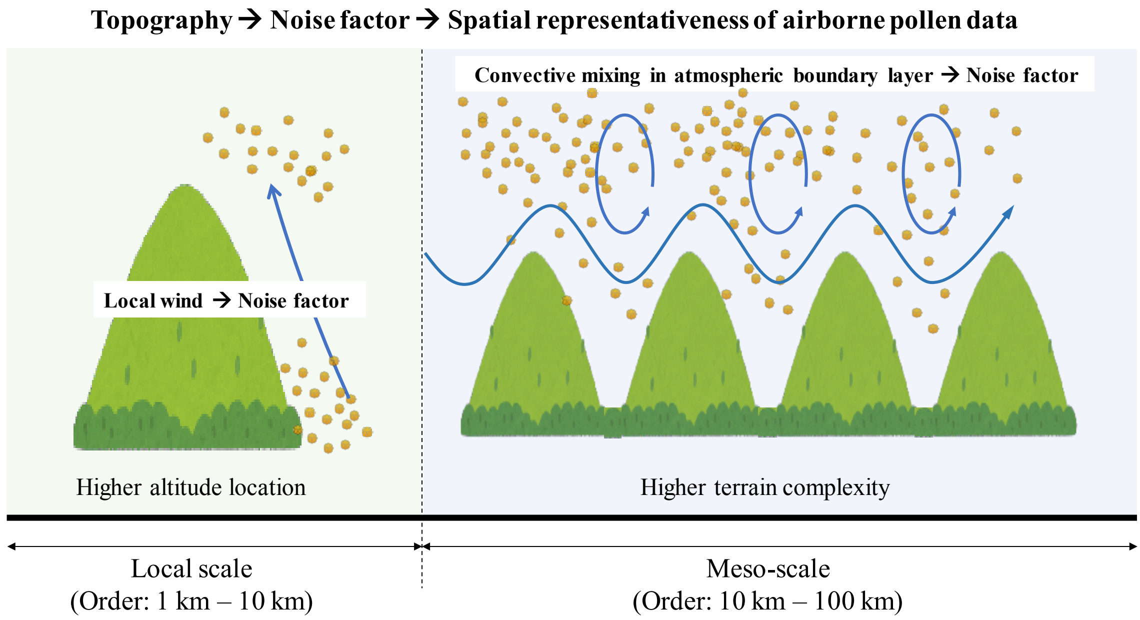


Fig. 4- 10 Schematic diagram for concluding remarks on spatial representativeness of airborne pollen data.

Chapter 5
A stochastic model for predicting annual airborne pollen – Hidden Markov model

Tseng Y-T, Kawashima S, Kobayashi S, Takeuchi S, Nakamura K. Forecasting the seasonal pollen index by using a hidden Markov model combining meteorological and biological factors. *Sci. Total Environ.* 2019: 134246.

5.1 Introduction

The motivations for using the HMM are as follows: (i) Mast flowering is inherently stochastic; thus, using the Markov process simplifies its prediction. (ii) The HMM easily integrates other factors correlated with the stochastic (Markov) process for predicting the next state. In addition, the probability of each outcome in the next state can be estimated from parameters. Therefore, based on the physiological mechanism behind the masting, I propose an alternative stochastic approach, HMM, to predict the annual airborne pollen. In the model, I assume the variation of annual airborne pollen to be a hidden Markov process considering the intrinsic stochastic properties in the masting, designating meteorological conditions in the previous summer as observable variables to recover the intra-annual variation in the annual airborne pollen. This study was conducted with the following objectives: (i) To parameterize the masting phenomenon and the relationship between male flower and the meteorological conditions in the previous summer using transition probability and emission distribution, respectively. (ii) To validate the performance of the stochastic model in annual airborne pollen prediction.

5.2 Materials and Methods

5.2.1 Study area

The monitoring work was conducted in Sapporo, the largest city on the northern Japanese island of Hokkaido. Sapporo is located in the transitional zone between the northern temperate zone and subarctic zone, which represents the floristic composition of forests similar to that of Europe and eastern North America at the genus level (Maekawa, 1974). *Quercus*, *Acer*, *Tilia*, *Fagus,* and *Betula*, for example, are common woody flora in such areas. Three forest types appear across the area: the cool-temperate forests, boreal forests, and subalpine forests (Ishizuka, 1974). Sapporo has a humid continental climate with an average annual precipitation of approximately 1100 mm and a mean annual temperature of approximately 8.5 °C due to its coastal location. The major source of allergenic pollen is the birch tree, which flowers from April to June, during which period widespread pollinosis occurs (Takeuchi et al., 2013).

5.2.2 Input data

A Durham gravimetric pollen sampler (Durham, 1946) was placed on a roof at the Hokkaido Institute of Public Health (43°4′55″ N, 141°20′1″ E; elevation: 25 m above sea level) during the months April–June from 1996 to 2017 to measure the amount of deposited birch pollen grains (grains cm^{-2} d^{-1}). In the counting process, based on morphological criteria, I distinguished birch pollen from others and considered all species within the genus *Betula* L. as the measuring object. The daily pollen concentration during the pollen season was summed to obtain the total annual pollen concentration. The pollen data obtained from the Durham gravimetric sampler have been proven to exhibit a similar trend to data from the Hirst volumetric sampler (Bricchi et al., 2000; Kishikawa et al., 2009).

Meteorological data were obtained from a station within the Japanese Automated Meteorological Data Acquisition System (AMeDAS). The AMeDAS station observed the following daily meteorological elements: mean air temperature (t_{avg}, °C), maximum air temperature (t_{max}, °C), minimum air temperature (t_{min}, °C), sunshine duration (sd, h), solar radiation (sr, MJ m^{-2}), precipitation (prec, mm), vapor pressure (vaporp, hPa), and relative humidity (rhum, %). All data were collected at daily intervals.

5.2.3 Feature selection

Feature selection was implemented before the model training process to determine the most relevant meteorological variables for male flower formation before the flowering year. The process to determine features that were highly correlated with the annual airborne pollen but exhibited low correlation to each other involved two steps: (i) Selection of the optimized agglomeration period, as described in Oteros et al. (2019). Instead of using simple monthly meteorological data, for each variable, I used the average over an optimized agglomeration period before the flowering year as the candidate input data. By moving the start date of the time-window from the 1st to the 30th of May, June, and July with variable length (number of days) from 14 to 30 days, I calculated the Pearson correlation coefficient of the annual data of annual airborne pollen and the average of the variable over each time-window. The optimal agglomeration period was then specified as the periods of occurrence of the maximum positive and minimum negative correlation coefficients, which can explain most effectively the year to year variation of pollen abundance. (ii) Elimination of one of two features having a high correlation: Subsequently, the correlations of each pair of features selected in (i) were calculated. One of the two features that had a correlation larger than 0.6 was removed, and the feature having a higher correlation with the pollen level was retained. Only training data were used in the feature selection process.

5.3 Hidden Markov models (HMMs)

5.3.1 Outline of HMMs

HMMs can be viewed as an extension of the state-space model (SSM) with discrete (i.e., discontinuous) hidden states. It is a doubly stochastic process with a hidden Markov process that is not observable, but can only be inferred through an observation sequence (Rabiner and Juang, 1986).

Figure 5-1 shows how transition occurs from one state to another. The HMM model can be described by the specification of parameter N and $\lambda = (\pi, A, B)$:

π: initial state probability distribution defined as $\pi = \{\pi(i)\}$, where $\pi(i) = P[q_1 = s_i]$, and $1 \le i \le N$.

A: state transition probability matrix defined as $A = \{a_{ij}\}$, where $a_{ij} = P\big[q_{t+1} = s_j \big| \, q_t = s_i]$, and $1 \le i, j \le N$.

B: emission probability distribution in state Sj defined as $B = \{b_{jk}\}$, where $b_{jk} = P[v_k | q_t = s_j]$; $1 \le j \le N, 1 \le k \le M$.

N: number of states of the Markov chain. Individual states are denoted as $S = \{s_1, s_2, \ldots s_N\}$, and the state at time t is denoted as q_t.

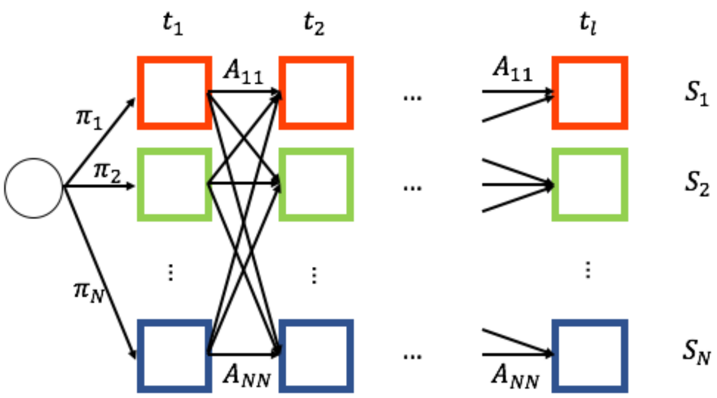


Fig. 5- 1 Transition between different hidden states.

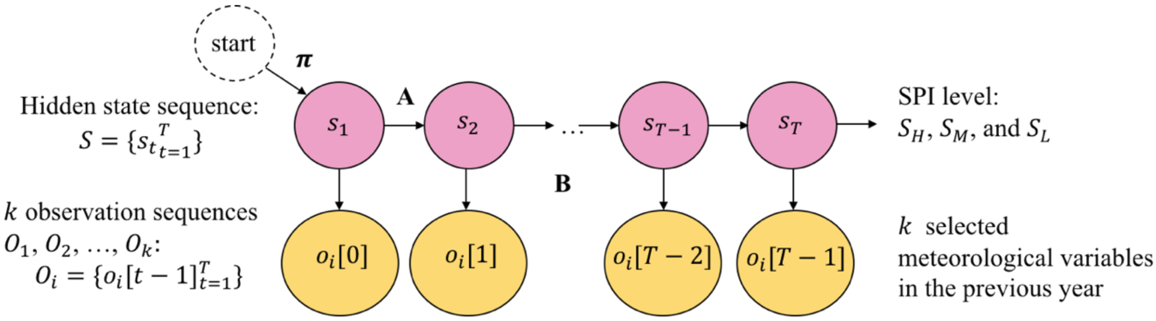


Fig. 5- 2 Illustration of hidden Markov model (HMM) for pollen prediction

5.3.2 Supervised learning process of HMM

Instead of assuming that the annual airborne pollen is governed by invisible hidden states, and to predict the temporal change in annual airborne pollen without knowledge of what the hidden states represent, I adopted a supervised training method to describe the hidden states practically as the annual airborne pollen levels and the observations as the meteorological conditions in the previous year. A supervised learning process of the proposed prediction model was implemented in Python 3 following the four steps described below:

(1) Preparation of input data. As shown in Fig. 5- 2, the annual airborne pollen level was designated as the hidden state sequence S, and k multiple sequences of different meteorological factors in the previous year $\{O_1, O_2, \ldots, O_k\}$ were designated as observation sequences. The meteorological variable and the period of meteorological data for each were determined through the approach noted in Section 2.3.

(2) Presetting of hidden states. Because increasing the number of hidden states leads to fewer observations for each hidden state, despite the realization of better description for the actual annual airborne pollen with narrowed ranges, in this study, I specified the number of hidden states to be three $(N = 3)$. The annual airborne pollen ranges corresponding to the hidden states (S_L, S_M, S_H) and the summary statistics of the training dataset are listed in Table 5- 1. As most of the mast years in Hokkaido occurred once every two years, non-mast years were determined as S_L, which accounts for 50% of the training data. Middle mast years and high mast years were then determined as S_M and S_H, where S_M and S_H are below and above approximately one standard deviation from the average of the overall data.

(3) Labeling of training dataset with S_L, S_M, and S_H. The annual airborne pollen data were labeled with the hidden states for determining the HMM parameters.

$$S = \{s_1,\ s_2,\ s_3, \ldots ,\ s_T\} \tag{5-1}$$

(4) Separation of complete dataset into training and test datasets. I divided the complete dataset (1996–2017) into training dataset (1996–2013) and test dataset (2014–2017) to validate the predictive performance of the model.

(5) Determination of HMM parameter $\lambda=(A, B, \pi)$. I trained the HMM parameters using the training dataset obtained from (4).

The initial probability (π_i) of state i was calculated as the percentage of appearance of state i in the training period:

$$\pi_i = S_i / \sum_{j=1}^{N} S_j \tag{5-2}$$

where S_i is the frequency of state i, and N is the number of hidden states.

The transition matrix (A) is computed as the percentage of transition state i to state j in the training period. Considering the Markov property, only the transition from the previous and the current states is considered.

$$A(i,j) = a(i,j) / \sum_{k=1}^{N} a(i,k) \tag{5-3}$$

where $a(i,j)$ is the frequency of state i to state j.

As all the observed meteorological data were continuous, the emission distributions were probability density functions instead of a matrix. I assumed that the hidden state output (B) followed a Gaussian distribution. Hence, for the determination of B, I simply computed the average and variance of each observation from each hidden state (S_L, S_M, S_H).

$$B(i,o)_{mean} = b(i,o)/S_i \tag{5-4}$$

$$B(i,o)_{var} = (b(i,o) - B(i,o)_{mean})^2 / S_i \tag{5-5}$$

where $b(i,o)$ is the emission of observation at state i.

(6) Defining the most likely hidden state sequence. The Viterbi algorithm (VA) was then used to find the most likely hidden state sequence $Q = \{q_1, q_2, .., q_T\}$ for the given

observations $O = \{O_1, O_2, .., O_T\}$ based on trained parameters λ. For details of the Viterbi algorithm, please refer to the work of Forney (Forney, 1973).

Table 5- 1 Annual airborne pollen ranges corresponding to hidden states (S_L, S_M, S_H) and summary statistics of the training dataset (SPI: annual airborne pollen; STD: standard deviation).

Hidden state	SPI (grains m^{-2})	Mean	STD	Min	Max	Percentage
S_L	0 < SPI < 1000	482.62	252.23	141	971	50.00%
S_M	1000 < SPI < 1700	1317.53	162.52	1088	1485	33.33%
S_H	1700 < SPI	2263.33	447.46	1935	2773	16.67%

5.4 Results

5.4.1 Feature selection

The results of two steps in the feature selection process are presented below:

(i) Optimized agglomeration periods of meteorological variables

In this step, I selected the period that can explain most effectively the year to year variation of pollen abundance for each variable. From the value of the correlation coefficient, I inferred the suitable meteorological conditions in the optimal period of the initiation year for male flower formation. The results of the optimized agglomeration periods of each variable are shown in Fig. 5- 3. It can be seen that highly correlated variables—among each group of temperature, solar, and water—demonstrate a similar pattern of correlation variation corresponding to different moving time-windows. In mid-May, the blue span in sr (r = -0.46) and sd (r = -0.45) as well as red span in rhum (r = 0.30) and prec (r = 0.50) show that less sunshine and higher humidity are more preferable. From the end of May until late June, higher temperature and solar exposure are especially desirable. Red spans in four parameters—vaporp, t_{min}, t_{max}, and t_{avg}—occur with the highest positive correlations of 0.68, 0.64, 0.68, and 0.68, respectively. Furthermore, the red span in solar factor sr (r = 0.70) and sd (r = 0.72) also occur around the same period. However, the appropriate conditions change in early and mid-July, the blue spans in the four parameters (vaporp, t_{min}, t_{max}, and t_{avg}) occur with the highest negative correlations of -0.46, -0.43, -0.31, and -0.38. In addition, the blue spans in rhum (r = -0.55) and prec (r = -0.56) also occur.

(ii) Elimination of one of two features having a high correlation

Figure 5-4 shows the correlation coefficients for each pair of variables. Significant positive correlation is noted among each group of temperature, solar, and humidity variables. Furthermore, because of the opposite trend of rhum and sr, as shown in Fig. 5- 3, a negative correlation is noted to exist between rhum and sr. Finally, the meteorological variables pos_vaporp (hPa), pos_sd (h), pos_rhum (%), pos_prec (mm), neg_sr (MJ m^{-2}), neg_rhum (%), and neg_prec (mm) were selected, with the correlation of one of any two features not exceeding 0.6.

[a]

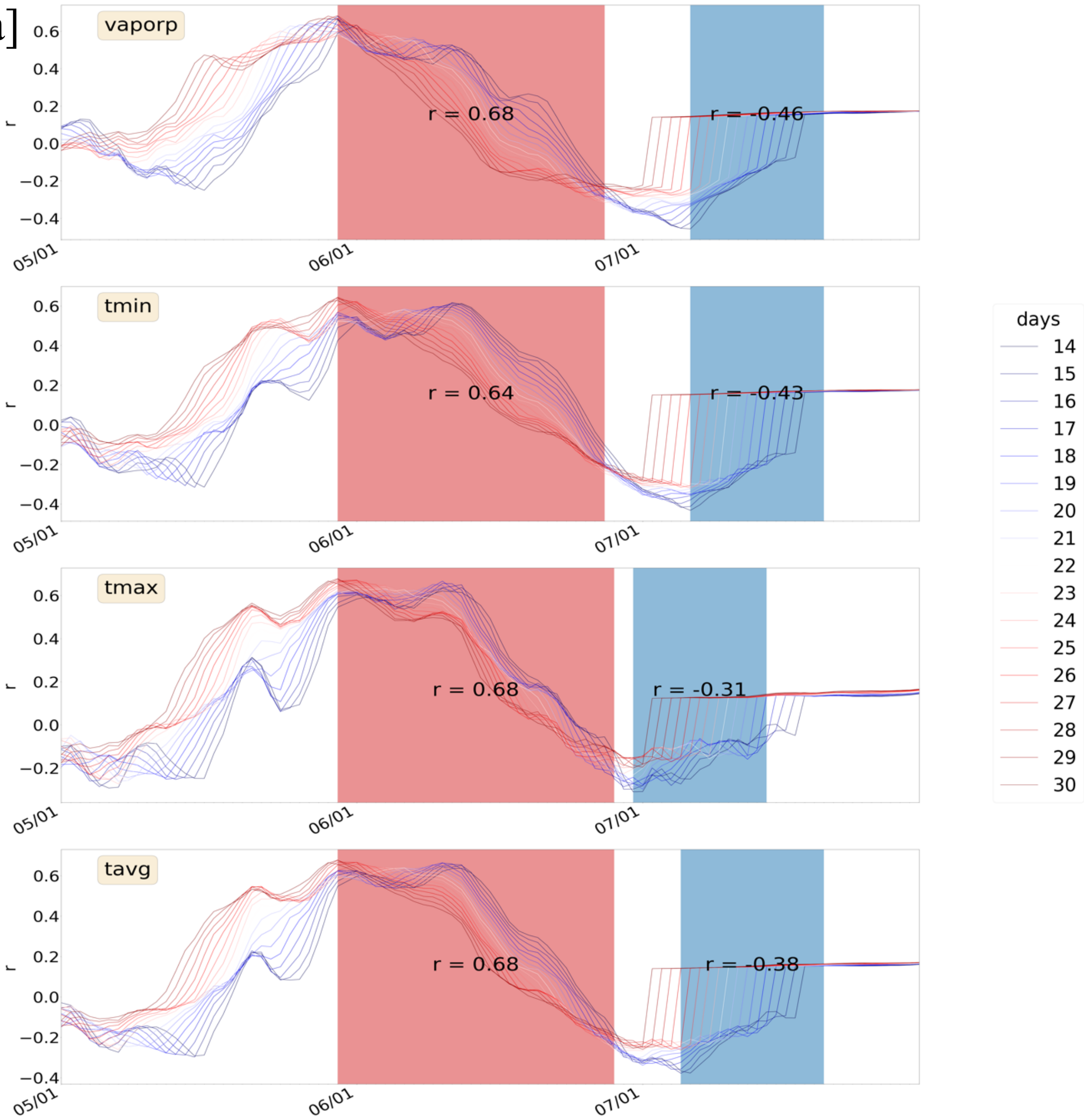


Fig. 5- 3 Pearson correlation variation corresponding to moving time-window (May–July) with variable length (14–30 days). Optimized agglomeration periods of each variable are noted, with the red and blue spans respectively corresponding to the periods when the maximum and minimum correlations with annual airborne pollen occur. [a] vaporp: vapor pressure (hPa); tmin: minimum air temperature (°C); tmax: maximum air temperature (°C); tavg: mean air temperature (°C). [b] sr: solar radiation (MJ m^{-2}); sd: sunshine duration (h); rhum: relative humidity (%); prec: precipitation (mm).

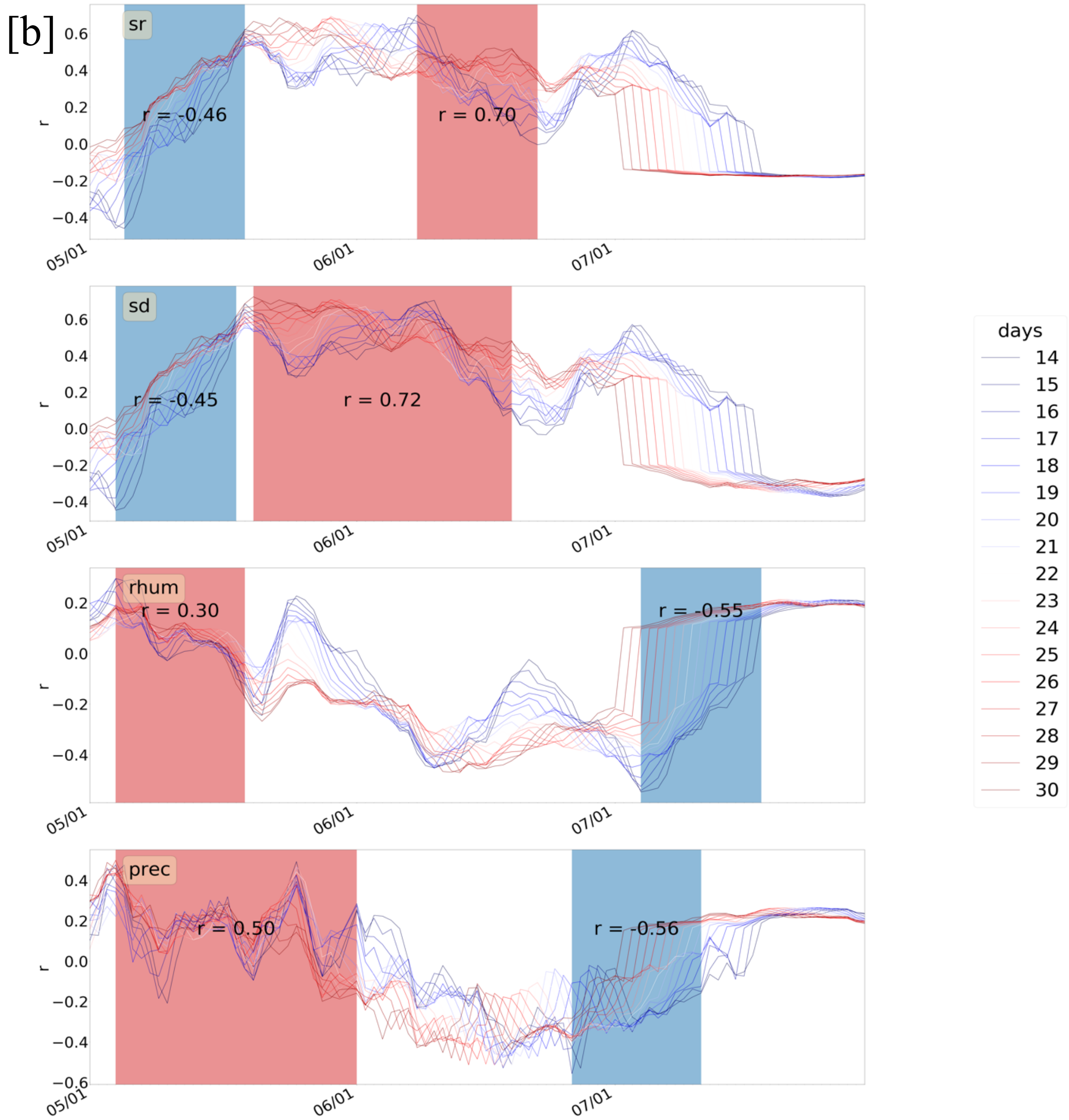


Fig. 5- 3 Pearson correlation variation corresponding to moving time-window (May–July) with variable length (14–30 days). Optimized agglomeration periods of each variable are noted, with the red and blue spans respectively corresponding to the periods when the maximum and minimum correlations with annual airborne pollen occur. [a] vaporp: vapor pressure (hPa); tmin: minimum air temperature (°C); tmax: maximum air temperature (°C); tavg: mean air temperature (°C). [b] sr: solar radiation (MJ m^{-2}); sd: sunshine duration (h); rhum: relative humidity (%); prec: precipitation (mm).

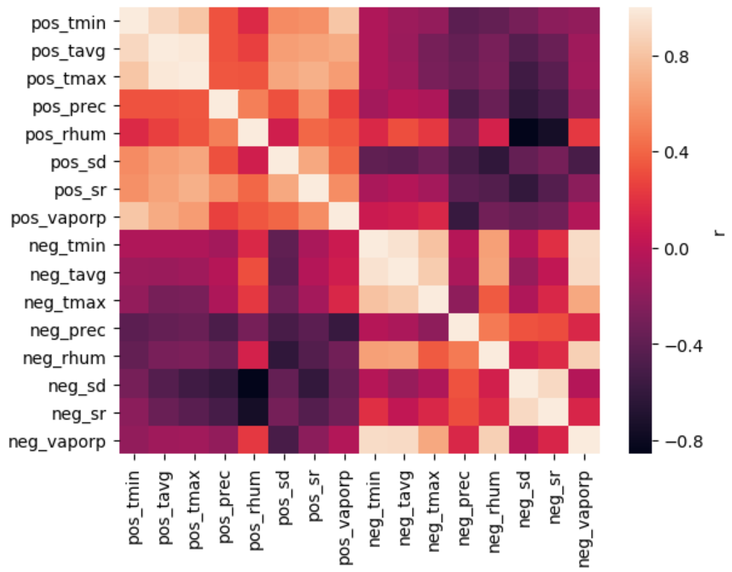


Fig. 5- 4 Matrix plot of correlation coefficient of each pair of variables among the features selected for the optimized agglomeration period (pos_ and neg_ in the variable name respectively refer to the features that have positive and negative correlations with the male flower formation, corresponding to the red and blue spans in Fig. 5- 3).

5.4.2 Trained HMM parameters

After the training process described in Section 3.2 was conducted, the parameters N and $\lambda = (\pi, A, B)$, which are necessary for the specification of an HMM, could be obtained. Among these parameters, the initial probabilities (π) of each hidden state were 50.00% (S_L), 33.33% (S_M), and 16.67% (S_H), demonstrating the majority of S_L followed by S_M and S_H in turn. The transition matrix (A) is shown in Fig. 5- 5, in which I can observe that S_L tends to switch to S_M with high probability (44%), albeit with relatively low tendency, to move to S_H (33%) or stay within the current annual airborne pollen range (22%). Moreover, S_H only occurs after S_L, and it turns out that there is only a 16.67% presence of S_H in the training period. Finally, S_M and S_H sparsely shift to S_M (20% and 33%), but it moves to S_L (80% and 67%).

Figure 5-6 shows the emission distributions (B) of the selected features for the three hidden states. The expected result was observed: as annual airborne pollen increased, pos_vaporp (hPa), pos_sd (h), pos_rhum (%), pos_prec (mm) increased, whereas neg_sr (MJ m^{-2}), neg_rhum (%), and neg_prec (mm) decreased.

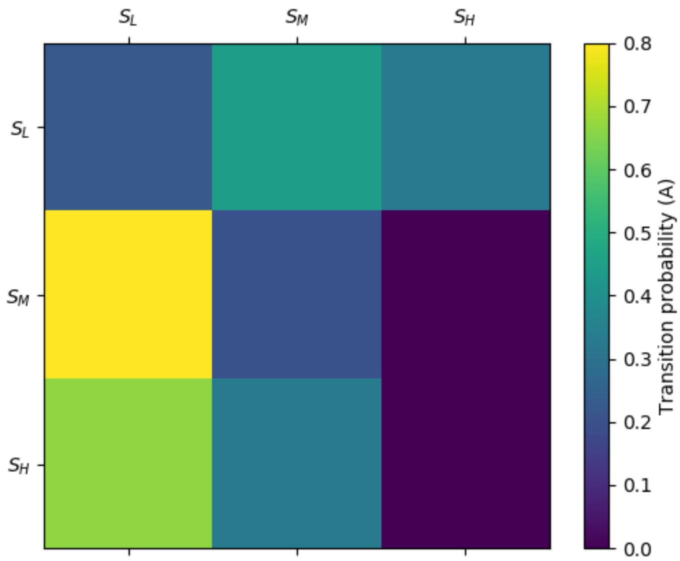


Fig. 5- 5 Matrix plot of transition matrix of supervised HMM using training data from 1996–2013.

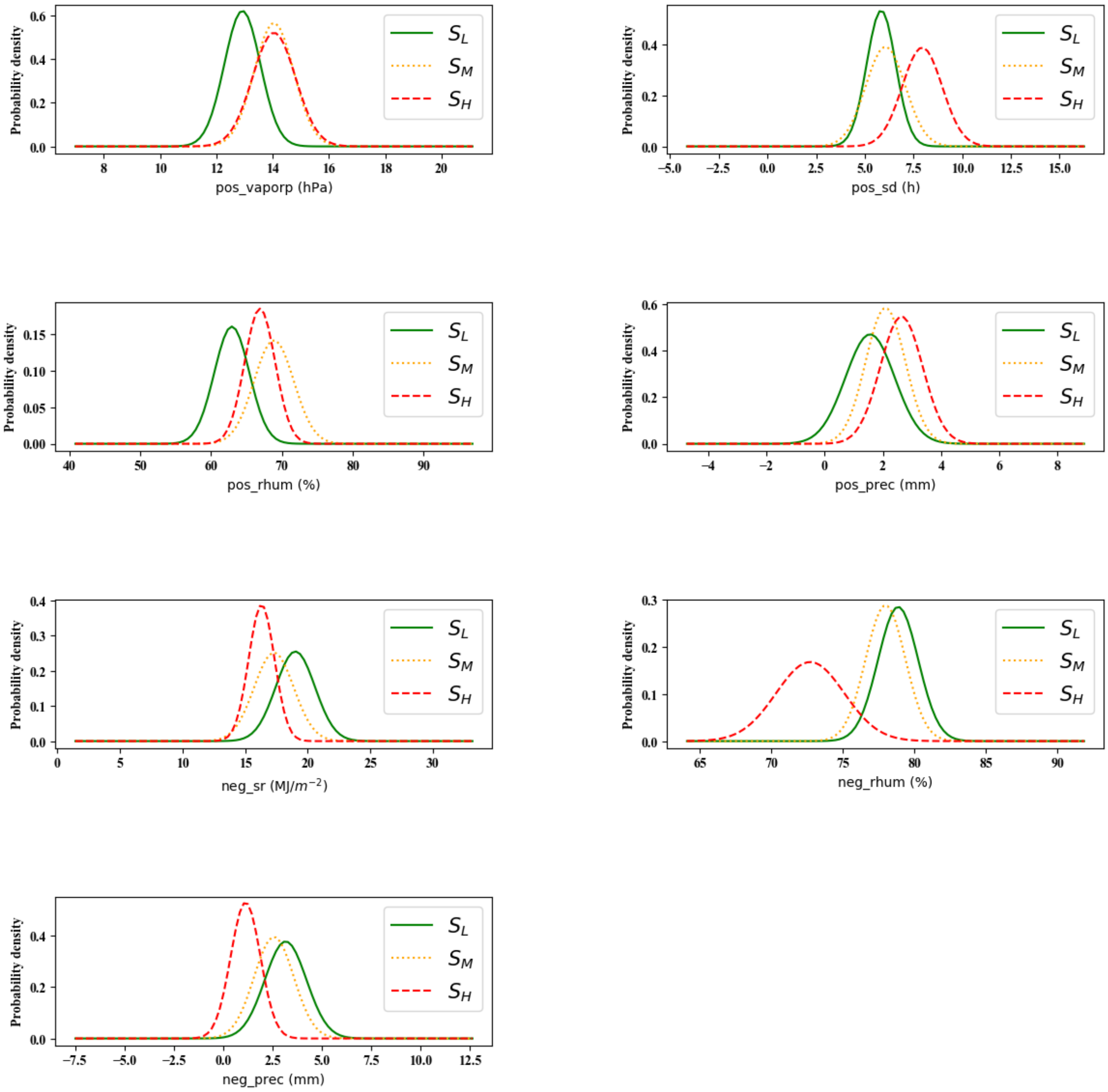


Fig. 5- 6 Emission distributions of selected features, specifically, pos_vaporp (hPa), pos_sd (h), pos_rhum (%), pos_prec (mm), neg_sr (MJ m^{-2}), neg_rhum (%), and neg_prec (mm), for three hidden states S_L, S_M, and S_H (pos_ and neg_ in the variable name respectively correspond to the red and blue span of the meteorological variable in Fig. 5- 3).

5.4.3 Prediction Results

Figure 5-7 shows the forecast results, the most likely hidden state sequence, obtained by VA based on the given observation (Table 5- 2) and trained parameters (Fig. 5- 5 and Fig. 5- 6). From the observed annual airborne pollen, the biennial cycle—in which most of the even years are mast years—was found, except for three interruptions (1996, 2002, and 2012). The recurrence in the observed data can be reproduced in most of the predicted data. However, the remaining mismatches between these two datasets led to the final prediction accuracies being 83.3% (three mismatches out of 18) in the training period and 75.0% (one mismatch out of four) in the test period. The model demonstrated two mismatches in S_H years and one mismatch in each of the S_L and S_M dispersal years.

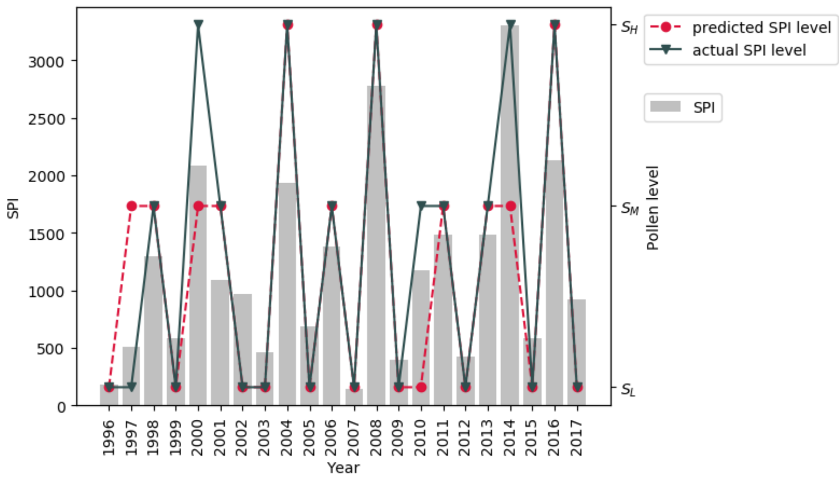


Fig. 5- 7 Annual airborne pollen (SPI) prediction results (1996–2013: training dataset, 2014–2017: test dataset).

5.5 Discussions

Predicting the annual airborne pollen is important for providing a reference for allergy stakeholders. Previous studies have primarily used deterministic models with the forecast results of the model being fully determined by the parameters. Most existing models derive the numerical value of annual airborne pollen by employing either regression (Garcia-Mozo et al., 2014; Latałowa et al., 2002; Ranta et al., 2008; Subiza et al., 1992), neural networks (Puc, 2012), or time-series models (Garcia-Mozo et al., 2014). The remaining models attempt to predict the level of the annual airborne pollen (Aznarte et al., 2007; Mesa et al., 2005; Sánchez-Mesa et al., 2002) or combine the classification and regression prediction (Oteros et al., 2013).

In contrast to previous approaches, this study employed a stochastic model, utilizing the probability properties of the innate flowering rhythm with a transition matrix (A) and the correlation of meteorological conditions and the number of male flowers with the emission distribution (B), to determine the most likely hidden state sequence. The stochastic model has the advantage that from the current state, it can predict each outcome of the next state with probabilities; the importance of providing the probability of each outcome rather than a strict numerical value has been demonstrated in the case of weather forecasting. Therefore, the classification results of annual airborne pollen level obtained using the stochastic model are beneficial for determining the probability of each outcome in the next season as a useful reference for allergy stakeholders.

The optimal period of each meteorological variable is found by moving time-windows from the start of May to the end of July with variable length, as described in Section 2.3. The peak correlation of the temperature (t_{min}, t_{avg}, and t_{max}) and the solar exposure (sd and sr) with annual airborne pollen occurs in the period from the end of May to middle or late June. This result coincides with the plant physiological fact that birch tree male flower development is regulated by the photosynthetic energy produced from May to late June (Dahl and Strandhede, 1996). The lower temperature before and after this period has been proved to be correlated with higher annual airborne pollen in the flowering year. The significant negative correlations for relative humidity and precipitation noted for the period around July, which is similar to previous results (Houle, 1999), may be related to proper environmental conditions for microspore initiation or a result of the water stress as an important regulating factor for masting behavior.

Depending on the different pollen types, the influential meteorological factors for pollen abundance may differ. Significant and positive correlation between both mean and maximum temperature and annual airborne pollen was found in arboreal taxa. On the other hand, significant and negative correlation between minimum air temperature and annual airborne pollen was found in some herbaceous taxa (Ruiz-Valenzuela and Aguilera, 2018). For birch pollen, previous studies reported the mean air temperature, maximum air temperature, and relative humidity in previous years to be relevant factors affecting the annual airborne pollen (Masaka and Maguchi, 2001; Tseng et al., 2018). Another regional model for the birch annual airborne pollen (Ritenberga et al., 2018) incorporated CO_2 as a significant factor, as the concentration of CO_2 is a primary factor affecting phenology and pollen production (Cecchi et al., 2010; D'Amato, 2011). In contrast, in our case, incorporating CO_2 does not result in significant improvement because the short annual airborne pollen time-series does not demonstrate a stable increase with CO_2.

The results of trained parameters A and B, show that the model can capture the masting behavior of birch trees and the effect of meteorological conditions in the previous year on the annual airborne pollen. Birch trees are known to have a biennial cycle of pollen production, but with some special years not following the cycle (Grewling et al., 2012; Jato et al., 2007; Latałowa et al., 2002; Malkiewicz et al., 2016; Skjøth et al., 2015). The high probability from S_L to a higher level (78%), as well as S_M and S_H to a lower level (80% or 100%) in trained parameter A could be an explanation for this. Because the masting behavior is an intrinsic process that cannot be easily explained using only one simple factor (Koenig and Knops, 2000), transition matrix A can act as a means to parameterize the complex phenomenon. The results of B appear to be in agreement with those of previous studies that stated that temperature and sunlight positively influenced male flower formation (Matthews, 1955; Norton and Kelly, 1988). It was also noted that the relative humidity, which causes water stress in plants, negatively influenced the flowering intensity.

The forecast result is the most probable path of annual airborne pollen level obtained from the trained parameters ($\lambda = (\pi, A, B)$) and the observation. Overall, the higher frequency of S_L and S_M in the actual sequence is thought to be the reason for the proposed model's superiority in predicting the occurrence of either a low or medium dispersal year. The mismatches in 1997 and 2000 could be attributable to (i) high transition probability of S_L to S_M rather than staying in S_L or moving up to S_H, and (ii) higher relative humidity in mid-May (pos_rhum) in 1996 and higher relative humidity in early July (neg_rhum) in 1999 leading to higher probability in predicting S_M than S_L and S_H, respectively (Fig. 5- 6 and Table 5- 2). The mismatch in 2010 could be attributable to lower relative humidity in mid-May (pos_rhum) and precipitation in May (pos_prec) as well as higher solar radiation in mid-May (neg_sr) in 2009, making it prone to predict S_L over S_M in 2010. Another error in 2014 is attributable to the transition probability of S_M to S_H being zero, which I trained from the training dataset. These results show how the tendency in the training data reflect on future prediction, showing that limited information in the dataset will restrict the predictive performance of the model. In addition, mismatches between the predicted and observed annual airborne pollen can result from the assumption of high correlation between flower intensity and pollen abundance in our model. The correlation fails when long-distance transport of pollen (Hjelmroos, 1991; Skjøth et al., 2007; Sofiev et al., 2006) or other factors diminish pollen in the air in the emission, transportation, or deposition process cannot be neglected (Ranta et al., 2008; Tseng and Kawashima, 2019).

Altogether, this study validated the applicability of the alternative approach, HMM, over the deterministic model used in previous studies for annual airborne pollen prediction. However, HMM has its limitation based on the Markov properties that the current state only depends on the previous state. The limitation makes the consideration of cycle interruption in the masting difficult. Further improvement can be implemented in terms of augmentation of the dataset and integration of other potential variables, including pollen production per male flower. In addition, HMM can also be applied to the prediction of airborne pollen concentration in different time-scales by considering the pollen level transition in time and meteorological conditions from the atmospheric environmental perspective.

Table 5- 2 Meteorological data of selected variables used as observation sequence for annual airborne pollen prediction from 1995 to 2016 (pos_ and neg_ in the variable name respectively correspond to the red and blue span of the meteorological variable in Fig. 5- 3; values with bold and underline are for indicating possible causes for mismatches in 1997, 2000, and 2010).

Year	pos_ vaporp (hPa)	pos_ sd (h)	pos_ rhum (%)	pos_ prec (mm)	neg_ sr (MJ m^{-2})	neg_ rhum (%)	neg_ prec (mm)
1995	13.01	5.67	65.00	1.57	17.57	76.07	1.57
1996	12.34	5.41	**<u>72.06</u>**	2.50	15.29	77.33	2.50
1997	13.37	5.34	65.50	2.62	15.31	76.13	2.62
1998	12.73	5.40	62.13	0.66	19.57	80.47	0.66
1999	13.81	**<u>6.50</u>**	72.50	**<u>3.21</u>**	15.23	**<u>80.67</u>**	**<u>3.21</u>**
2000	14.32	5.38	82.75	2.29	15.30	81.73	2.29
2001	12.93	6.64	63.00	0.60	17.85	80.40	0.60
2002	12.29	5.96	55.75	1.41	22.94	81.47	1.41
2003	13.48	8.35	61.19	1.93	17.75	69.73	1.93
2004	12.94	6.98	63.19	2.47	16.99	78.80	2.47
2005	14.24	6.44	69.81	2.03	14.88	76.07	2.03
2006	13.27	5.31	50.38	1.43	22.03	77.60	1.43
2007	14.85	8.98	67.00	2.79	15.95	67.73	2.79
2008	13.11	5.65	66.75	2.47	21.33	76.40	2.47
2009	13.81	4.61	**<u>57.69</u>**	**<u>1.47</u>**	**<u>21.91</u>**	77.33	**<u>1.47</u>**
2010	14.88	7.06	64.25	1.53	19.00	78.80	1.53
2011	13.69	5.54	68.13	0.97	17.47	81.20	0.97
2012	13.64	7.56	73.25	2.62	17.37	77.80	2.62
2013	15.11	7.45	72.25	1.69	15.88	76.20	1.69
2014	14.91	5.50	62.50	2.03	16.54	70.80	2.03
2015	13.64	7.56	56.13	1.28	19.97	66.60	1.28
2016	13.27	6.00	57.38	1.76	17.50	72.87	1.76
Average	13.62	6.33	64.93	1.88	17.89	76.46	1.88

5.6 Conclusion

This study represents the first application of a stochastic model to pollen concentration prediction. It combined the biological and meteorological factors affecting the flowering intensity as a hidden Markov process and observation sequence in an HMM—parameterizing the intrinsically stochastic process of masting behavior with a transition probability matrix and the causal relationship of meteorological conditions in the initiation year and annual airborne pollen in the flowering year with emission probability distribution. From the results, it was found that using a transition matrix to explain the biennial cycle in birch tree is valid, as it showed that a low annual airborne pollen level tended to switch to a higher one with higher probability than it staying in the same level, and vice versa. Furthermore, the exact periods of meteorological variables were determined through the process of feature selection to attain better results, and higher temperature and solar exposure in June, as well as less rainfall in July, were examined and found to be more preferable for male catkin formation. The trained parameters and the Viterbi algorithm gave the forecast results with an accuracy of 83.3% in the training period and 75.0% in the test period.

Based on the current meteorological information and trained parameters in HMM, the probability of each annual airborne pollen level in the next pollen season can also be estimated, which would be useful for clinical strategies and anti-allergy medicine production decision-making. However, unresolved problems remain, such as insufficient information in limited training data for future prediction and difficulty dealing with the interruption in the biennial cycle for the Markov properties (where the current state only depends on the previous state). Future studies should focus on improvement of the performance by augmentation of the dataset, integration of other potential variables, and replacement for other probability distributions rather than normal distribution in emission probability. Finally, I also expect to apply the presented approach to pollen concentration prediction in hourly and daily time-scales.

Chapter 6
General Discussion

6.1 The foundation of the forecasting models

The two proposed models were based on the following recurrence relation:

$$P_i = f(M_{i-1},\ P_{i-1}) \tag{6-1}$$

where P_i and P_{i-1} are the annual airborne pollen in year i and year $i-1$; M_{i-1} is the weather condition in the previous year (M_{i-1}).

This equation is derived from (i) the mechanism behind mast flowering, in which flowering intensity (F_i) including male flower formation as well as synchronization among individuals that are influenced by the weather condition from the previous year (M_{i-1}) and the flowering intensity in the previous year (F_{i-1}), as written in (6-2), and (ii) the premise of the model that the annual pollen abundance (P_i) is proportional to the annual flowering intensity (F_i) (6-3).

$$F_i = f(M_{i-1},\ F_{i-1}) \tag{6-2}$$

$$F_i = kP_i \tag{6-3}$$

where k is constant.

6.2 The deterministic solution and stochastic solution

In Chapter 3, on the basis of the recurrence relation in (6-1), the variable P_{i-1} was substituted repeatedly to obtain the multiple regression equation (6-4) with only the meteorological conditions in the previous years M_{i-1}, M_{i-2}, …, $M_{i-\mathrm{n}}$.

$$P_i = \sum_{i=1}^{n} a_i M_{i-1} + b \tag{6-4}$$

The results of the basic models (models built with temperature (T) and solar (S)) showed an increasing trend in performance as the number of years observed increased from one to three, representing that at least three years' meteorological conditions are needed to replace the biological factor P_{i-1}.

In Chapter 5, HMM was applied to pollen level prediction using the relation (6-1) found in the hidden Markov model (HMM): the most probable hidden state at time t (s_t) is determined by not only previous hidden state (s_{t-1}) but also the corresponding observation (o_{t-1}). In HMM, the series of annual airborne pollen is designated as the hidden Markov process and the series of annual meteorological conditions are specified as observation sequences. The trained parameters showed that the proposed model enables (i) using Markov process to simplify and predict the mast flowering and (ii) integrating related factors of annual airborne pollen, meteorological conditions in the previous year, for predicting the next state.

Both models can predict the annual airborne pollen using only meteorological data and are characterized differently. In deterministic model, the prediction of numerical concentration is entirely determined by the trained parameters. On the other hand, the stochastic model can provide the most probable sequence of annual airborne pollen level and the probability of each pollen level of the next season.

The two models are suggested to be applied according to different needs. For instance, since deterministic model can predict the value of the annual airborne pollen, it is useful for prediction of pollinosis prevalence and crop yields. In comparison, stochastic model can provide the probability of each annual airborne pollen level, which is therefore helpful in the process of decision making by medical doctors and allergy patients (Tseng et al., 2019).

6.3 Common restrictions shared by the two models

Although the premise of the models (6-3) is supported by previous studies (Ranta et al., 2008; Yasaka et al., 2009), this proportional relation fails when the pollen data does not represent the pollen concentration in the observed region in the following cases: (i) the long-distance transport of pollen (Hjelmroos, 1991; Skjøth et al., 2007; Sofiev et al., 2006) and (ii) other factors that diminish airborne pollen in the process of emission, transportation, and deposition (Ranta et al., 2008; Tseng and Kawashima, 2019).

In chapter 4, the influence of the topographical conditions to spatial representativeness of pollen data is discussed. The applicability of the forecasting algorithm suggested in Chapter 3 was investigated in nine monitoring stations under different topographical conditions. The findings concluded that, the difference in spatial representatives of airborne pollen data is a primary factor for the difference in the applicability of the forecasting models.

Chapter 7
Conclusions

7.1 Concluding remarks

Annual airborne pollen is a continuing concern within the fields of aerobiology, ecology, botany, and epidemiology. However, understanding of the fluctuation of annual airborne pollen is still inadequate since there are a limited number of areas where long-term airborne pollen observation is available.

As shown in Fig. 7- 1, the two prediction models of annual airborne pollen proposed in this thesis are based on (6-1). The deterministic algorithm in Chapter 3 was derived by substituting the explanatory variable P_{i-1} repeatedly. the algorithm predicted the annual airborne pollen using only meteorological data from previous years, and the forecast result is entirely determined by the trained coefficient in the algorithm. The stochastic model HMM in Chapter 5 assumed the variation of annual airborne pollen to be a hidden Markov process considering the intrinsic stochastic properties in masting, designating meteorological conditions, in the previous summer as observable variables to recover the inter-annual variation in annual airborne pollen. In contrast to the deterministic model, this model provides the probability of each pollen level. Each model accurately predicted pollen concentration and pollen level.

The proposed models provided an important reference for decision making on clinical strategies for allergic disease and management of anti-allergy medicine production. In addition, since the input data of both models only contain meteorological data, the annual airborne pollen can be predicted even if the data of the current season is missing or unavailable.

7.2 Future work

As the results showed in the two proposed models, seven or more explanatory variables are necessary to develop the forecasting models from the limited sample size. To make a model with lower generalized error, future studies should focus on collecting or augmenting longer dataset for model training.

The proposed models are expected to be applicable to (i) longer time scale: with predicted meteorological data, the simulation of the variation in annual airborne pollen under climate change can be realized and (ii) larger space scale: by clustering the regions with high similarity in topography and climate, the proposed model can be used in regions where there is no pollen data because only meteorological data is needed. Finally, responding to the needs of modern society, the proposed models can be applied on a mobile alert system, by combining it with epidemiological data and allergenicity information such as existence of co-allergen and air pollutants.

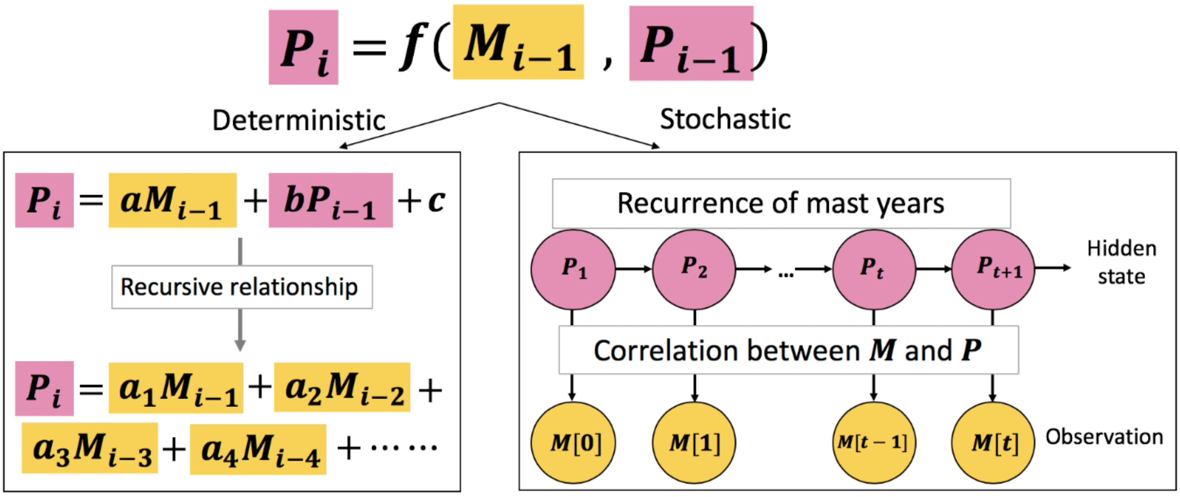


Fig. 7- 1 Basic concept of the thesis

Acknowledgements

First and foremost, I would like to thank my supervisor Honorary Professor Dr. Shigeto Kawashima. Through his supervising, I could build the foundation of my academic skills, including the construction of the research design, development of appropriate models, and conceptualization of the research problems. I also appreciate Professor Dr. Kimihito Nakamura for helpful advices for my study. His academic attitude was motivational for me. I appreciate all their time, ideas, and warmness to make my PhD experience, or these five years in Japan, so fruitful.

I would also like to thank the two reviewers: Professor Dr. Satoshi Hoshino and Professor Dr. Masayuki Fujihara, Graduate School of Agriculture Kyoto University, for their helpful advices for improving the quality of this thesis.

I would also like to thank Professor Dr. Jeroen Buters at the Center of Allergy and Environment (ZAUM) for providing the valuable opportunity for three-month collaborative research in Munich. Through having a European perspective on the pollinosis issue, I become to further realize the significance of my doctoral work. Every time passing by the only pollen indicator in the world besides research institute ZUAM, I got deeply motivated and kept asking myself, “how to make a better prediction”. Special thanks also go to Professor Dr. Jose Oteros for giving me considerable useful technical suggestions for improving the prediction models. During this stay, I got kind support from them and the team members not only in research, but also living and experiencing different cultures.

I would also like to thank Dr. Athanasios Damialis at Institute of Environmental Medicine (UNIKA-T) for giving me substantial introduction of his studies and Dr. David Topping at

Manchester University for giving me advice for my research.

For proposing the concept of applying hidden Markov model in annual airborne pollen prediction. I would like to thank our colleague, Jiuding Duan, Graduate School of Informatics Kyoto University. For a mentoring in data science, I also thank Kazuki Otsuka for impressive data analysis skill and lot of inspirational discussions.

I shared time, personally and professionally, with the members in hydrological environment engineering lab. As a foreign student, I could not further enjoy the Japanese culture without them. I am especially thankful for Kenji Miki. We shared lot of time working and exploring in the aerobiology field together. His enthusiasm for research was contagious for me.

Another special thanks to Haruka Ueda, for his time and warm support during PhD life.

I gratefully thank Think Global Education Trust Foundation and Japan-Taiwan Exchange Association for their support during master’s and doctoral studies. I also thank to European Academy of Allergy and Clinical Immunology (EAACI). They provided me the opportunity to implement the study in international research environment.

Finally, I would like to thank my dearest family. Their love and encouragement are always the best mental support for me to keep going forward despite all the challenges.

References

Abrahamson WG, Layne JN. Long-term patterns of acorn production for five oak species in xeric Florida uplands. Ecology 2003; 84: 2476-2492.

Ashman T-L, Knight TM, Steets JA, Amarasekare P, Burd M, Campbell DR, et al. Pollen limitation of plant reproduction: ecological and evolutionary causes and consequences. Ecology 2004; 85: 2408-2421.

Aznarte JL, Sanchez JMB, Lugilde DN, Fernandez CDL, de la Guardia CD, Sanches FA. Forecasting airborne pollen concentration time series with neural and neuro-fuzzy models. Expert Systems with Applications 2007; 32: 1218-1225.

Bailey NT, Bailey NTJ. The elements of stochastic processes with applications to the natural sciences. Vol 25: John Wiley & Sons, 1990.

Beggs PJ. Impacts of climate change on aeroallergens: past and future. Clinical & Experimental Allergy 2004; 34: 1507-1513.

Behrendt H, Becker WM, Friedrichs KH, Darsow U, Tomingas R. Interaction between aeroallergerns and airborne particulate matter. International Archives of Allergy and Immunology 1992; 99: 425-428.

Binkley CS. Is succession in hardwood forests a stationary Markov process? Forest Science 1980; 26: 566-570.

Blackley CH. Experimental Researches on the Causes and Nature of Catarrhus Aestivus (hay-fever Or Hay-asthma): Baillière, Tindall & Cox, 1873.

Bogdziewicz M, Fernández-Martínez M, Bonal R, Belmonte J, Espelta JM. The Moran effect and environmental vetoes: phenological synchrony and drought drive seed production in a Mediterranean oak. Proceedings of the Royal Society B: Biological Sciences 2017a; 284: 20171784.

Bogdziewicz M, Steele MA, Marino S, Crone EE. Correlated seed failure as an environmental veto to synchronize reproduction of masting plants. New Phytologist 2018; 219: 98-108.

Bogdziewicz M, Szymkowiak J, Fernández-Martínez M, Peñuelas J, Espelta JM. The effects of local climate on the correlation between weather and seed production differ in two species with contrasting masting habit. Agricultural and forest meteorology 2019; 268: 109-115.

Bogdziewicz M, Szymkowiak J, Kasprzyk I, Grewling Ł, Borowski Z, Borycka K, et al.

Masting in wind-pollinated trees: system-specific roles of weather and pollination dynamics in driving seed production. Ecology 2017b; 98: 2615-2625.

Bostock J. Case of a periodical affection of the eyes and chest. Medico-chirurgical transactions 1819; 10: 161.

Bostwick D. Yield probabilities as a Markov process: Marketing Economics Division, Economic Research Service, US Department of Agriculture, 1962.

Bricchi E, Frenguelli G, Mincigrucci G. Experimental results about Platanus pollen deposition. Aerobiologia 2000; 16: 347.

Buters JTM, Antunes C, Galveias A, Bergmann KC, Thibaudon M, Galan C, et al. Pollen and spore monitoring in the world. Clinical and Translational Allergy 2018; 8: 5.

Caesar JC, Macdonald AD. Shoot development in Betula papyrifera. II. Comparison of vegetative and reproductive short-shoot growth. Canadian Journal of Botany 1983; 61: 3066-3071.

Caesar JC, Macdonald AD. Shoot development in Betula papyrifera. IV. Comparisons between growth characteristics and expression of vegetative long and short shoots. Canadian journal of botany 1984; 62: 446-453.

Cecchi L, D'Amato G, Ayres JG, Galan C, Forastiere F, Forsberg B, et al. Projections of the effects of climate change on allergic asthma: the contribution of aerobiology. Allergy 2010; 65: 1073-1081.

Chuine I, Cour P, Rousseau DD. Selecting models to predict the timing of flowering of temperate trees: implications for tree phenology modelling. Plant, Cell & Environment 1999; 22: 1-13.

Churchill GA. Stochastic models for heterogeneous DNA sequences. Bulletin of Mathematical Biology 1989; 51: 79-94.

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

Cristofolini F, Gottardini E. Concentration of airborne pollen of Vitisvinifera L. and yield forecast: a case study at S. Michele all'Adige, Trento, Italy. Aerobiologia 2000; 16: 125-129.

Crone EE, Miller E, Sala A. How do plants know when other plants are flowering? Resource depletion, pollen limitation and mast-seeding in a perennial wildflower. Ecology Letters 2009; 12: 1119-1126.

Cunha M, Ribeiro H, Abreu I. Pollen-based predictive modelling of wine production:

application to an arid region. European journal of agronomy 2016; 73: 42-54.
D'Amato G. Effects of climatic changes and urban air pollution on the rising trends of respiratory allergy and asthma. Multidisciplinary Respiratory Medicine 2011; 6: 28-37.
D'Amato G, Pawankar R, Vitale C, Lanza M, Molino A, Stanziola A, et al. Climate Change and Air Pollution: Effects on Respiratory Allergy. Allergy Asthma & Immunology Research 2016; 8: 391-395.
Dahl Å, Strandhede S-O. Predicting the intensity of the birch pollen season. Aerobiologia 1996; 12: 97-106.
Damialis A, Halley JM, Gioulekas D, Vokou D. Long-term trends in atmospheric pollen levels in the city of Thessaloniki, Greece. Atmospheric Environment 2007; 41: 7011-7021.
Durham OC. A proposed standard method of gravity sampling, counting and volumetric interpolation of results. Journal of Allergy and Clinical Immunology 1946; 17: 79-86.
D’Amato G, Cecchi L, Bonini S, Nunes C, Annesi-Maesano I, Behrendt H, et al. Allergenic pollen and pollen allergy in Europe. Allergy 2007; 62: 976–990.
D’Amato G, Holgate ST, Pawankar R, Ledford DK, Cecchi L, Al-Ahmad M, et al. Meteorological conditions, climate change, new emerging factors, and asthma and related allergic disorders. A statement of the World Allergy Organization. World Allergy Organization Journal 2015; 8: 1-52.
Emberlin J, Savage M, Woodman R. Annual variations in the concentrations of Betula pollen in the London area, 1961–1990. Grana 1993; 32: 359-363.
Espelta JM, Cortés P, Molowny-Horas R, Sánchez-Humanes B, Retana J. Masting mediated by summer drought reduces acorn predation in Mediterranean oak forests. Ecology 2008; 89: 805-817.
Fernández-Martínez M, Belmonte J, Espelta JM. Masting in oaks: disentangling the effect of flowering phenology, airborne pollen load and drought. Acta Oecologica 2012; 43: 51-59.
Forney GD. The viterbi algorithm. Proceedings of the IEEE 1973; 61: 268-278.
Franke A, Caelli T, Hudson RJ. Analysis of movements and behavior of caribou (Rangifer tarandus) using hidden Markov models. Ecological Modelling 2004; 173: 259-270.
Franke A, Caelli T, Kuzyk G, Hudson RJ. Prediction of wolf (Canis lupus) kill-sites using hidden Markov models. Ecological modelling 2006; 197: 237-246.
Galán C, García-Mozo H, Cariñanos P, Alcázar P, Domínguez-Vilches E. The role of temperature in the onset of the Olea europaea L. pollen season in southwestern Spain.

International Journal of Biometeorology 2001; 45: 8-12.

Galán C, García-Mozo H, Vázquez L, Ruiz L, Díaz De La Guardia C, Domínguez-Vilches E. Modeling olive crop yield in Andalusia, Spain. Agronomy Journal 2008; 100: 98-104.

Garcia-Mozo H, Galan C, Jato V, Belmonte J, de la Guardia CD, Fernandez D, et al. Quercus pollen season dynamics in the Iberian Peninsula: Response to meteorological parameters and possible consequences of climate change. Annals of Agricultural and Environmental Medicine 2006; 13: 209-224.

Garcia-Mozo H, Yaezel L, Oteros J, Galan C. Statistical approach to the analysis of olive long-term pollen season trends in southern Spain. Science of the Total Environment 2014; 473: 103-109.

García-Mozo H, Gómez-Casero MT, Domínguez E, Galán C. Influence of pollen emission and weather-related factors on variations in holm-oak (Quercus ilex subsp. ballota) acorn production. Environmental and Experimental Botany 2007; 61: 35-40.

Gehrig R, Jud S, Schuepbach E, Clot B. Pollen measurements in an alpine environment – altitudinal gradients and transport. Vol 1: MeteoSwiss (CH) and University of Montreal (CA), Montreal, Canada, 2011.

Gehrig R, Maurer F, Schwierz C. Designing new automatically generated pollen calendars for the public in Switzerland. Aerobiologia 2018; 34: 349-362.

Gotoda H, Maguchi S, Kawahara H, Terayama Y, Fukuda S. Springtime pollinosis and oral allergy syndrome in Sapporo. Auris Nasus Larynx 2001; 28: S49–S52.

Graham-Rowe D. Lifestyle: When allergies go west. Nature 2011; 479: S2-S4.

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

Hatfield JL, Prueger JH. Temperature extremes: Effect on plant growth and development. Weather and climate extremes 2015; 10: 4-10.

He H, Wang Y, Ma Q, Ma J, Chu B, Ji D, et al. Mineral dust and NOx promote the conversion of SO2 to sulfate in heavy pollution days, Sci. Rep., 4, 4172, 2014.

Heinrich J, Wichmann H-E. Traffic related pollutants in Europe and their effect on allergic disease. Current opinion in allergy and clinical immunology 2004; 4: 341-348.

Hirst JM. An automatic volumetric spore trap. Annals of applied Biology 1952; 39: 257-265.

Hjelmroos M. Evidence of long-distance transport of Betula pollen. Grana 1991; 30: 215-228.

Houle G. Mast seeding in Abies balsamea, Acer saccharum and Betula alleghaniensis in an old growth, cold temperate forest of north-eastern North America. Journal of Ecology

1999; 87: 413-422.

Isagi Y, Sugimura K, Sumida A, Ito H. How does masting happen and synchronize? Journal of Theoretical Biology 1997; 187: 231-239.

Ishizuka K. Mountain vegetation. The flora and vegetation of Japan 1974.

Janzen DH. Seed predation by animals. Annual review of ecology and systematics 1971; 2: 465-492.

Jato V, Rodríguez-Rajo FJ, Aira MJ. Use of phenological and pollen-production data for interpreting atmospheric birch pollen curves. Annals of Agricultural and Environmental Medicine 2007; 14.

Jazcilevich AD, García AR, Caetano E. Locally induced surface air confluence by complex terrain and its effects on air pollution in the valley of Mexico. Atmospheric Environment 2005; 39: 5481-5489.

Jochner S, Höfler J, Beck I, Göttlein A, Ankerst DP, Traidl-Hoffmann C, et al. Nutrient status: a missing factor in phenological and pollen research? Journal of experimental botany 2013; 64: 2081-2092.

Jonsen ID, Basson M, Bestley S, Bravington MV, Patterson TA, Pedersen MW, et al. State-space models for bio-loggers: A methodological road map. Deep Sea Research Part II: Topical Studies in Oceanography 2013; 88: 34-46.

Kasprzyk I, Ortyl B, Dulska-Jeż A. Relationships among weather parameters, airborne pollen and seed crops of Fagus and Quercus in Poland. Agricultural and Forest Meteorology 2014; 197: 111-122.

Kawashima S, Clot B, Fujita T, Takahashi Y, Nakamura K. An algorithm and a device for counting airborne pollen automatically using laser optics. Atmospheric Environment 2007; 41: 7987-7993.

Kelly D. The evolutionary ecology of mast seeding. Trends in ecology & evolution 1994; 9: 465-470.

Kelly D, Sork VL. Mast seeding in perennial plants: why, how, where? Annual Review of Ecology and Systematics 2002; 33: 427-447.

Kikuzawa K. Leaf survival of woody plants in deciduous broad-leaved forests. 1. Tall trees. Canadian Journal of Botany 1983; 61: 2133-2139.

Kim J-H, Oh J-W, Lee H-B, Kim S-W, Kang I-J, Kook M-H, et al. Changes in Sensitization Rate to Weed Allergens in Children with Increased Weeds Pollen Counts in Seoul Metropolitan Area. J Korean Med Sci 2012; 27: 350-355.

Kishikawa R, Sahashi N, Saitoh A, Kotoh E, Shimoda T, Shoji S, et al. Japanese cedar airborne pollen monitoring by Durham’s and burkard samplers in Japan-estimation of the usefulness of Durham’s sampler on Japanese cedar pollinosis. Global Environ Res 2009; 13: 55-62.

Knapp EE, Goedde MA, Rice KJ. Pollen-limited reproduction in blue oak: implications for wind pollination in fragmented populations. Oecologia 2001; 128: 48-55.

Knight TM, Steets JA, Vamosi JC, Mazer SJ, Burd M, Campbell DR, et al. Pollen limitation of plant reproduction: pattern and process. Annu. Rev. Ecol. Evol. Syst. 2005; 36: 467-497.

Kobayashi S, Jin K, Ogawa H, Tsuzuki T. Airborne pollen survey in North Ward, Sapporo in 1997. Report of the Hokkaido Institute of Public Health 1998; 48: 45-48.

Koenig WD. Global patterns of environmental synchrony and the Moran effect. Ecography 2002; 25: 283-288.

Koenig WD, Ashley MV. Is pollen limited? The answer is blowin'in the wind. Trends in Ecology & Evolution 2003; 18: 157-159.

Koenig WD, Knops JMH. Scale of mast-seeding and tree-ring growth. Nature 1998; 396: 225.

Koenig WD, Knops JMH. Patterns of annual seed production by northern hemisphere trees: A global perspective. American Naturalist 2000; 155: 59-69.

Koenig WD, Knops JMH, Carmen WJ, Pearse IS. What drives masting? The phenological synchrony hypothesis. Ecology 2015; 96: 184-192.

Koike T. Leaf Structure and Photosynthetic Performance as Related to the Forest Succession of Deciduous Broad-Leaved Trees1. Plant Species Biology 1988; 3: 77-87.

Langrock R, King R, Matthiopoulos J, Thomas L, Fortin D, Morales JM. Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. Ecology 2012; 93: 2336-2342.

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

Lavee S. Biennial bearing in olive (Olea europaea). Annales, Series Historia Naturalis 2007.

Liao T, Wang S, Ai J, Gui K, Duan B, Zhao Q, et al. Heavy pollution episodes, transport pathways and potential sources of PM2.5 during the winter of 2013 in Chengdu (China). Science of The Total Environment 2017; 584-585: 1056-1065.

Lyles D, Rosenstock TS, Hastings A, Brown PH. The role of large environmental noise in masting: general model and example from pistachio trees. Journal of Theoretical Biology 2009; 259: 701-713.

Madeja J, Wypasek E, Plytycz B, Sarapata K, Harmata K. Quantification of airborne birch (Betula sp.) pollen grains and allergens in Krakow. Archivum immunologiae et therapiae experimentalis 2005; 53: 169-174.

Maekawa F. Origin and characteristics of Japan's flora. The flora and vegetation of Japan 1974.

Malkiewicz M, Drzeniecka-Osiadacz A, Krynicka J. The dynamics of the Corylus, Alnus, and Betula pollen seasons in the context of climate change (SW Poland). Science of the Total Environment 2016; 573: 740-750.

Masaka K, Maguchi S. Modelling the masting behaviour of Betula platyphylla var. japonica using the resource budget model. Annals of Botany 2001; 88: 1049-1055.

Matthews JD. The influence of weather on the frequency of beech mast years in England. Forestry: An International Journal of Forest Research 1955; 28: 107-116.

Maya-Manzano JM, Sadyś M, Tormo-Molina R, Fernández-Rodríguez S, Oteros J, Silva-Palacios I, et al. Relationships between airborne pollen grains, wind direction and land cover using GIS and circular statistics. Science of The Total Environment 2017; 584-585: 603-613.

Mesa JAS, Galan C, Hervas C. The use of discriminant analysis and neural networks to forecast the severity of the Poaceae pollen season in a region with a typical Mediterranean climate. International Journal of Biometeorology 2005; 49: 355-362.

Miao YC, Zheng YJ, Wang S, Liu SH. Recent advances in and prospects of the formation research of haze over Beijing-Tianjin-Hebei, China. Environ. Res Climatic 2015; 20: 356-368.

Miquel P. Les organismes vivants de l'atmosphère: Gauthier-Villars, 1883.

Moreira X, Abdala-Roberts L, Linhart YB, Mooney KA. Masting promotes individual-and population-level reproduction by increasing pollination efficiency. Ecology 2014; 95: 801-807.

Nakae K, Baba K. Update on epidemiology of pollinosis in Japan: changes over the last 10 years. Clinical & Experimental Allergy Reviews 2010; 10: 2-7.

Nilsson SG, Wastljung U. Seed predation and cross-pollination in mast-seeding beech (Fagus Sylvatica) patches. Ecology 1987; 68: 260-265.

Norton DA, Kelly D. Mast seeding over 33 years by Dacrydium cupressinum Lamb.(rimu)(Podocarpaceae) in New Zealand: the importance of economies of scale. Functional ecology 1988: 399-408.

Oei HD, Spieksma FTM, Bruynzeel PLB. Birch pollen asthma in the Netherlands. Allergy 1986; 41: 435-441.

Oke TR. Boundary layer climates: Routledge, 2002.

Okuda M. Epidemiology of Japanese cedar pollinosis throughout Japan. Annals of Allergy, Asthma & Immunology 2003; 91: 288–296.

Oteros J, Bartusel E, Alessandrini F, Nunez A, Moreno DA, Behrendt H, et al. Artemisia pollen is the main vector for airborne endotoxin. Journal of Allergy and Clinical Immunology 2019; 143: 369-377.e5.

Oteros J, Garcia-Mozo H, Hervas-Martinez C, Galan C. Year clustering analysis for modelling olive flowering phenology. International Journal of Biometeorology 2013; 57: 545-555.

Oteros J, Orlandi F, Garcia-Mozo H, Aguilera F, Ben Dhiab A, Bonofiglio T, et al. Better prediction of Mediterranean olive production using pollen-based models. Agronomy for Sustainable Development 2014; 34: 685-694.

Oteros J, Pusch G, Weichenmeier I, Heimann U, Moller R, Roseler S, et al. Automatic and Online Pollen Monitoring. International Archives of Allergy and Immunology 2015; 167: 158-166.

Palau JL, Pérez-Landa G, Diéguez JJ, Monter C, Millán MM. The importance of meteorological scales to forecast air pollution scenarios on coastal complex terrain. Atmospheric Chemistry and Physics 2005; 5: 2771-2785.

Patterson TA, Thomas L, Wilcox C, Ovaskainen O, Matthiopoulos J. State–space models of individual animal movement. Trends in ecology & evolution 2008; 23: 87-94.

Pawankar R, Holgate ST, Canonica GW, Lockey RF, Blaiss MS. WAO white book on allergy 2013 update. World Allergy Organization 2013.

Pearse IS, Koenig WD, Kelly D. Mechanisms of mast seeding: resources, weather, cues, and selection. New Phytologist 2016; 212: 546-562.

Phillips PCB. Bayesian model selection and prediction with empirical applications. Journal of Econometrics 1995; 69: 289-331.

Piovesan G, Adams JM. Masting behaviour in beech: linking reproduction and climatic variation. Canadian Journal of Botany 2001; 79: 1039-1047.

Puc M. Artificial neural network model of the relationship between Betula pollen and meteorological factors in Szczecin (Poland). International Journal of Biometeorology 2012; 56: 395-401.

Rabiner L, Juang B. An introduction to hidden Markov models. ieee assp magazine 1986; 3: 4-16.

Raftery AE, Madigan D, Hoeting JA. Bayesian model averaging for linear regression models.

Journal of the American Statistical Association 1997; 92: 179-191.

Ranta H, Hokkanen T, Linkosalo T, Laukkanen L, Bondestam K, Oksanen A. Male flowering of birch: Spatial synchronization, year-to-year variation and relation of catkin numbers and airborne pollen counts. Forest Ecology and Management 2008; 255: 643-650.

Ranta H, Oksanen A, Hokkanen T, Bondestam K, Heino S. Masting by Betula-species; applying the resource budget model to north European data sets. International Journal of Biometeorology 2005; 49: 146-151.

Rasmussen A. The effects of climate change on the birch pollen season in Denmark. Aerobiologia 2002; 18: 253-265.

Recio M, Picornell A, Trigo MM, Gharbi D, Garcia-Sanchez J, Cabezudo B. Intensity and temporality of airborne Quercus pollen in the southwest Mediterranean area: Correlation with meteorological and phenoclimatic variables, trends and possible adaptation to climate change. Agricultural and Forest Meteorology 2018; 250: 308-318.

Ritenberga O, Sofiev M, Siljamo P, Saarto A, Dahl A, Ekebom A, et al. A statistical model for predicting the inter-annual variability of birch pollen abundance in Northern and North-Eastern Europe. Science of the Total Environment 2018; 615: 228-239.

Rogers CA, Wayne PM, Macklin EA, Muilenberg ML, Wagner CJ, Epstein PR, et al. Interaction of the onset of spring and elevated atmospheric CO_2 on ragweed (Ambrosia artemisiifolia L.) pollen production. Environmental Health Perspectives 2006; 114: 865-869.

Ruiz-Valenzuela L, Aguilera F. Trends in airborne pollen and pollen-season-related features of anemophilous species in Jaen (south Spain): A 23-year perspective. Atmospheric Environment 2018; 180: 234-243.

Satake A, Iwasa YOH. Pollen coupling of forest trees: forming synchronized and periodic reproduction out of chaos. Journal of Theoretical Biology 2000; 203: 63-84.

Silvertown JW. The evolutionary ecology of mast seeding in trees. Biological Journal of the Linnean Society 1980; 14: 235-250.

Skjøth CA, Baker P, Sadyś M, Adams-Groom B. Pollen from alder (Alnus sp.), birch (Betula sp.) and oak (Quercus sp.) in the UK originate from small woodlands. Urban Climate 2015; 14: 414-428.

Skjøth CA, Sommer J, Stach A, Smith M, Brandt J. The long-range transport of birch (Betula) pollen from Poland and Germany causes significant pre-season concentrations in Denmark. Clinical & Experimental Allergy 2007; 37: 1204-1212.

Smagorinsky J. Global atmospheric modeling and the numerical simulation of climate. Weather and Climate modification 1974: 633-686.

Smith CC, Hamrick JL, Kramer CL. The advantage of mast years for wind pollination. The American Naturalist 1990; 136: 154-166.

Sofiev M, Siljamo P, Ranta H, Rantio-Lehtimäki A. Towards numerical forecasting of long-range air transport of birch pollen: theoretical considerations and a feasibility study. International journal of biometeorology 2006; 50: 392.

Spieksma FTM. Pollinosis in Europe: new observations and developments. Review of Palaeobotany and Palynology 1990; 64: 35-40.

Stach A, Emberlin J, Smith M, Adams-Groom B, Myszkowska D. 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 2008; 52: 311-321.

Stultz CM, White JV, Smith TF. Structural analysis based on state-space modeling. Protein Science 1993; 2: 305-314.

Subiza J, Masiello JM, Subiza JL, Jerez M, Hinojosa M, Subiza E. Prediction of annual variations in atmospheric concentrations of grass-pollen - a method based on meteorological factors and grain crop estimates. Clinical and Experimental Allergy 1992; 22: 540-546.

Sun W, Zhang H, Palazoglu A. Prediction of 8 h-average ozone concentration using a supervised hidden Markov model combined with generalized linear models. Atmospheric environment 2013a; 81: 199-208.

Sun W, Zhang H, Palazoglu A, Singh A, Zhang WD, Liu SW. Prediction of 24-hour-average PM2.5 concentrations using a hidden Markov model with different emission distributions in Northern California. Science of the Total Environment 2013b; 443: 93-103.

Sánchez-Mesa JA, Galán C, Martínez-Heras JA, Hervás-Martínez C. The use of a neural network to forecast daily grass pollen concentration in a Mediterranean region: the southern part of the Iberian Peninsula. Clinical & Experimental Allergy 2002; 32: 1606-1612.

Takeuchi S, Kobayashi S, Jin K, Ogawa H, Tsuzuki T. Airborne pollen survey in North Ward of Sapporo City in 1998. Report of the Hokkaido Institute of Public Heath 1999; 49: 148–151.

Takeuchi S, Kojima H, Kobayashi S. Comprehensive Survey of Airborne Pollen in Sapporo,

Japan in 2012. Rep.HokkaidoInst.Pub.Health 2013; 524: 1-56.
Thibaudon M, Šikoparija B, Oliver G, Smith M, Skjøth CA. Ragweed pollen source inventory for France–The second largest centre of Ambrosia in Europe. Atmospheric Environment 2014; 83: 62–71.
Thomas LC, Allen DE, Morkel-Kingsbury N. A hidden Markov chain model for the term structure of bond credit risk spreads. International Review of Financial Analysis 2002; 11: 311-329.
Thyer M, Kuczera G. A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach. Journal of Hydrology 2003a; 275: 12-26.
Thyer M, Kuczera G. A hidden Markov model for modelling long-term persistence in multi-site rainfall time series. 2. Real data analysis. Journal of Hydrology 2003b; 275: 27-48.
Triantafyllou AG, Kassomenos PA. Aspects of atmospheric flow and dispersion of air pollutants in a mountainous basin. Science of the total environment 2002; 297: 85-103.
Tseng Y-T, Kawashima S. Applying a pollen forecast algorithm to the Swiss Alps clarifies the influence of topography on spatial representativeness of airborne pollen data. Atmospheric Environment 2019; 212: 153-162.
Tseng Y-T, Kawashima S, Kobayashi S, Takeuchi S, Nakamura K. Algorithm for forecasting the total amount of airborne birch pollen from meteorological conditions of previous years. Agricultural and Forest Meteorology 2018; 249: 35-43.
Tseng Y-T, Kawashima S, Kobayashi S, Takeuchi S, Nakamura K. Forecasting the seasonal pollen index by using a hidden Markov model combining meteorological and biological factors. Science of The Total Environment 2019: 134246.
Tucker BC, Anand M. On the use of stationary versus hidden Markov models to detect simple versus complex ecological dynamics. Ecological Modelling 2005; 185: 177-193.
Vacchiano G, Hacket-Pain A, Turco M, Motta R, Maringer J, Conedera M, et al. Spatial patterns and broad-scale weather cues of beech mast seeding in Europe. New Phytologist 2017; 215: 595-608.
Van Hulst R. On the dynamics of vegetation: Markov chains as models of succession. Vegetatio 1979; 40: 3-14.
Van Wart J, Grassini P, Yang H, Claessens L, Jarvis A, Cassman KG. Creating long-term weather data from thin air for crop simulation modeling. Agricultural and Forest Meteorology 2015; 209: 49-58.

Visser I, Raijmakers MEJ, van der Maas HLJ. Hidden Markov Models for Individual Time Series. In: Valsiner J, Molenaar PCM, Lyra MCDP, Chaudhary N, editors. Dynamic Process Methodology in the Social and Developmental Sciences. Springer US, New York, NY, 2009, pp. 269-289.

Voukantsis D, Niska H, Karatzas K, Riga M, Damialis A, Vokou D. Forecasting daily pollen concentrations using data-driven modeling methods in Thessaloniki, Greece. Atmospheric Environment 2010; 44: 5101-5111.

Wagatsuma Y, Matsuyama R, Noto K, Ito K. Japanese white birch pollinosis in Sapporo, Japan. Japanese Journal of Allergology 1972; 21: 710-717.

Waggoner PE, Stephens GR. Transition probabilities for a forest. Nature 1970; 225: 1160-&.

Wayne P, Foster S, Connolly J, Bazzaz F, Epstein P. Production of allergenic pollen by ragweed (Ambrosia artemisiifolia L.) is increased in CO_2-enriched atmospheres. Annals of Allergy, Asthma & Immunology 2002; 88: 279-282.

White JV, Stultz CM, Smith TF. Protein classification by stochastic modeling and optimal filtering of amino-acid sequences. Mathematical Biosciences 1994; 119: 35-75.

WHO. Phenology and human health: Allergic disorders. Copenhagen: WHO Reg. Office Europe. 2003.

Woillez M, Fablet R, Ngo T-T, Lalire M, Lazure P, De Pontual H. A HMM-based model to geolocate pelagic fish from high-resolution individual temperature and depth histories: European sea bass as a case study. Ecological modelling 2016; 321: 10-22.

Yasaka M, Kobayashi S, Takeuchi S, Tokuda S, Takiya M, Ohno Y. Prediction of birch airborne pollen counts by examining male catkin numbers in Hokkaido, northern Japan. Aerobiologia 2009; 25: 111-117.

Yli-Panula E, Fekedulegn DB, Green BJ, Ranta H. Analysis of airborne betula pollen in Finland; a 31-year perspective. International journal of environmental research and public health 2009; 6: 1706-1723.

Yoon B-J. Hidden Markov models and their applications in biological sequence analysis. Current genomics 2009; 10: 402-415.

Zhang H, Zhang W, Palazoglu A, Sun W. Prediction of ozone levels using a Hidden Markov Model (HMM) with Gamma distribution. Atmospheric environment 2012; 62: 64-73.

Zhang Y, Bielory L, Mi Z, Cai T, Robock A, Georgopoulos P. Allergenic pollen season variations in the past two decades under changing climate in the United States. Global Change Biology 2015; 21: 1581-1589.

Ziello C, Sparks TH, Estrella N, Belmonte J, Bergmann KC, Bucher E, et al. Changes to airborne pollen counts across Europe. PloS one 2012; 7: e34076.

Ziska LH, Beggs PJ. Anthropogenic climate change and allergen exposure: The role of plant biology. Journal of Allergy and Clinical Immunology 2012; 129: 27-32.

Ziska LH, Caulfield FA. Rising CO_2 and pollen production of common ragweed (Ambrosia artemisiifolia L.), a known allergy-inducing species: implications for public health. Functional Plant Biology 2000; 27: 893-898.

Šaulienė I, Šukienė L, Daunys G, Valiulis G, Vaitkevičius L, Matavulj P, et al. Automatic pollen recognition with the Rapid-E particle counter: the first-level procedure, experience and next steps. Atmospheric Measurement Techniques 2019; 12: 3435-3452.