# Statistical Analysis on Daily Variations of Birch Pollen Amount with Climatic Variables in Sapporo

Masaru Inatsu[1], Satoshi Kobayashi[2], Shinji Takeuchi[2], and Anna Ohmori[3]

[1]*Faculty of Science*, *Hokkaido University*, *Sapporo*, *Japan*
[2]*Hokkaido Institute of Public Health*, *Sapporo*, *Japan*
[3]*School of Science*, *Hokkaido University*, *Sapporo*, *Japan*

## Abstract

Birch pollen mainly causes springtime allergy-related diseases, birch pollinoses, widely known in high-latitude countries. By utilizing the observation in Sapporo from 2001 to 2011, we found that the daily pollen amount almost follows the log-normal distribution with its characteristic time-scale of several days. The pollen amount itself was therefore taken as a major predictor for its day-to-day variations. Another predictor was chosen from climatic variables that were possibly related to the pollen amount such as temperature, rainfall, sunshine duration, wind, relative humidity, rainfall, and daily temperature difference to explain daily variations of the pollen amount. A resulting statistical equation with two independent predictors of lagged pollen amount and diurnal temperature range based on the multiple regression analysis provided a reasonable hindcast prediction with the correlation coeffcient with observation being 0.80. Moreover, the equation was better fitted to the observations in abundant years than in poor-yield years.

(**Citation**: Inatsu, M., S. Kobayashi, S. Takeuchi, and A. Ohmori, 2014: Statistical analysis on daily variations of birch pollen amount with climatic variables in Sapporo. *SOLA*, **10**, 172−175, doi:10.2151/sola.2014-036.)

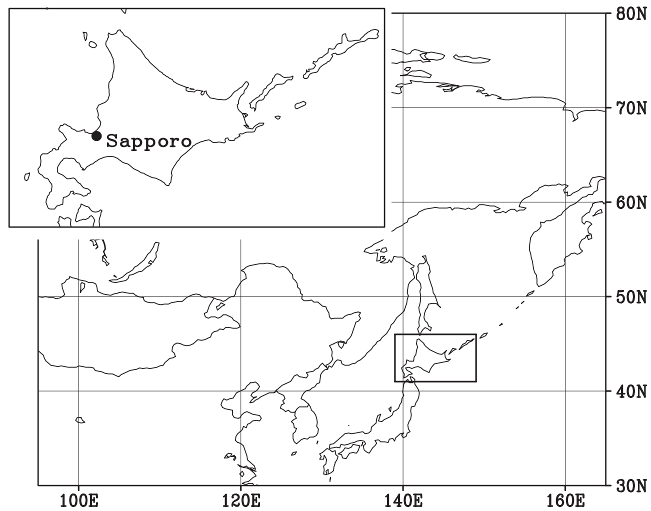


Fig. 1. Location of the pollen site in the map with a panel in the upper left zooming in Hokkaido with a mark at Sapporo, denoted by the rectangle box.

## 1. Introduction

Pollen of birch, the genus *Betula*, is a main cause of springtime allergy-related diseases such as rhinitis, itchy eyes, and sneezing in Hokkaido (Fig. 1), the northern island of Japan (Gotoda et al. 2001). Approximately 20% of population in Hokkaido were reported to be a patient of allergy caused by non-cedar pollens (Baba and Nakae 2008) and the patients have been recently increasing (Gotoda et al. 2001). More than 40% of birch pollinosis patients coincidently suffer from oral allergy syndrome to *Rosaceae* fruits (Yamamoto et al. 2004). However, because no medical treatment has been established for a complete recovery of pollinosis, an effective prescription for the patients is to avoid pollen exposure during the pollen season in order to alleviate their symptom. Hence if the daily prediction of the amount of airborne pollen were realized, it would be a key piece of information for the patients.

Airborne pollen survey has been operationally conducted in major cities of Hokkaido since the end of 1990s. Kobayashi et al. (2014) analyzed the data and revealed that the amount of the annual birch pollen varied with a significant biennial rhythm and gradually increased. Yasaka et al. (2009) moreover paid attention to a high correlation between the male catkin number in a year and the pollen amount in the following year (Ranta et al. 2008), and developed a practical prediction method for annual-sum pollen amount. This method is actually helpful for the prescription of pollen prophylactics before the pollen season. On the other hand, the amount of pollen also has significant day-to-day variations, which are much influenced by some meteorological conditions. Accumulated temperature corresponds to the flower and anther maturation (Larsson 1993) and high temperature triggers the pollen release to the air (Clot 2001; Ribeiro 2003). Daily temperature fluctuation and sunshine hours also have a positive effect on the pollen (Fehér and Járai-Komlódi 1998). Wind vector is related to the long-distance transport of wind-pollinated pollen with the air mass, which possibly increases the airborne pollen concentration at sites far from the origin (e.g., Porsbjerg et al. 2003; Ranta et al. 2006). Intermittent precipitation washes pollutants from the air; if airborne pollen is taken into rain or snow droplets, it is likely forced to quickly move down to the ground with them (cf., Sofiev et al. 2006). Ambient humidity may also suppress the flowering.

The daily pollen forecasts have been conventionally based on *in situ* aerobiological monitoring and phenological observations. This persistent effort is quite reasonable not only for real-time monitoring but for short-range forecasts of the pollen amount, because the pollen would stay in the air for approximately one day if only the dry deposition process worked (Sofiev et al. 2006). Even so, only by the monitoring without meteorological information, we can neither forecast the episodic increase in the pollen amount by high temperatures on the site nor forecast the sudden decrease by heavy rainfall. Moreover, monitoring the pollen amount at a single site would be insufficient to predict a long distance transport of pollen from the source plants. Recently, a numerical model of birch pollen emission and dispersion in the atmosphere has been developed, especially focusing on North Europe. Birch forest areas and flowering timing are both important to estimate pollen emission flux into the atmosphere and the latter is reasonably estimated by the temperature sum model (Linkosalo et al. 2010). Many processes such as wind advection, turbulent mixing, gravitational settling, and scavenging with precipitation have to be implemented into the model. The European group successfully reproduced the observed distribution patterns and absolute levels of concentration (Sofiev et al. 2013; Siljamo et al. 2013). However, because model and observation uncertainties prevent

Corresponding author: Masaru Inatsu, Rigaku 8th bldg., Hokkaido University, N10W8, Sapporo 060-0810 Japan. E-mail: inaz@mail.sci.hokudai.ac.jp. 

us from stabilizing the threshold-based statistics, a large-volume averaging is necessary to compare model results with observation. This averaging leaves out consideration of particular episodes and individual time series.

The purpose of this paper is to develop the statistical relation on daily variations of the birch pollen amount in Sapporo, the capital of Hokkaido. Although the long-term pollen observation has been conducted, a numerical modelling of the pollen amount has not been attempted over Hokkaido yet. As has been reviewed above, it is still difficult to predict the daily time-series of pollen at a specific site only by the numerical modelling. Therefore, as the first stage of daily pollen amount analysis in Sapporo, we will simply find a statistical relation based on the long-term pollen observation and meteorological information. Multiple regression analysis may systematically construct a predictor-predictand relation if one selects relevant predictors. Real-time pollen amount is *a priori* the most important predictor because of its residual timescale, and climatic variables such as temperature, wind vector, and precipitation are possible meteorological predictors that may *more or less* influence the atmospheric dispersion and pollen emission. However, since the onset date cannot be predicted without the temperature sum model, we limit our purpose to finding a statistical relation on the daily variations during the pollen season with the onset date given.

## 2. Airborne pollen and climatic data

We used the data of airborne pollen amount between 2001 and 2011 at Hokkaido Institute of Public Health in Sapporo, located at 43°05′N, 141°20′E (Fig. 1). A Durham-type sampler was installed at the roof floor of a three-story building at 16 m aloft from the ground surface. Airborne pollen surveys were conducted every day between April and June, and the pollen was collected on applied petrolatum glass slides with the sampler. The glass slides were stained with gentian violet, and the amount of pollen counted under a microscope in a 3.24-$cm^2$ square area. The pollen amount at a date is here defined as the pollen amount collecting from 9 a.m. of the date to 9 a.m. of the following date. The onset date is defined as the first observed date in a case that more than 1 grains/$cm^2$ has been observed for two consecutive days. The retreat date is defined as the final observed date in a case that no pollen has been counted for three consecutive days after the date. There are three major birch species in Hokkaido of *Betula platyphylla*, *B.ermanii*, and *B.maximowicziana*, all of which pollen shape is mostly comparable. This paper hence treated all *Betula* species as a single category. The pollen amount on the glass slides at the site is assumed to be representative of the airborne pollen concentration in Sapporo.

Climatic variables we used are surface air temperature at 1.5 m above the ground, surface wind vector at 60 m above the ground, solar radiation, relative humidity, and precipitation in Sapporo Meteorological Observatory (43°04′N, 141°20′E), approximately 2.5 km south to the pollen site. This is a site of the Japanese nationwide meteorological observation network, called AMeDAS, operated by the Japan Meteorological Agency (JMA). Because both the meteorological and pollen sites are located in the Sapporo urban area with a level land, it is reasonable to regard the meteorological data as those at the pollen site. All the climatic variables are possible predictors in finding an optimal statistical relation by the multiple regression analysis.

## 3. Results

The annual total amount of birch pollen is displayed by a bar graph in Fig. 2a. The average amount in the analysis period is 1,137 grains/$cm^2$. In 2008 total pollen amount was largest in the period we analyzed and was approximately 2.4 times as large as the average. By contrast, total pollen amount in 2007 and 2009 were much smaller than the other years. As has already reported by Kobayashi et al. (2014), annual pollen amount has a biennual

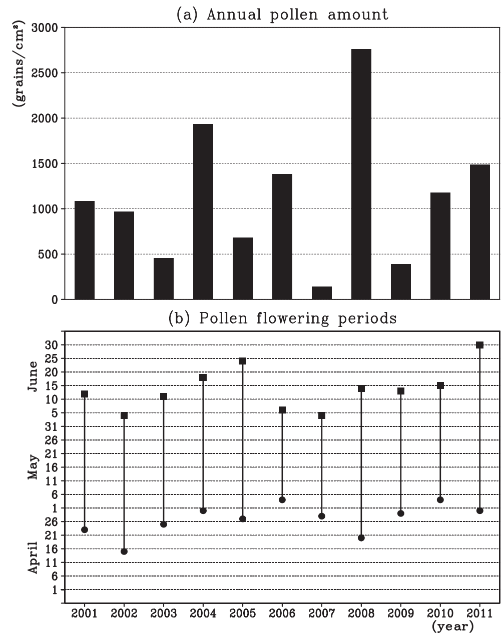


Fig. 2. (a) Annual-sum birch pollen amount (grains/$cm^2$) and (b) duration of birch pollen season from onset to retreat dates, in Sapporo from 2001 to 2011. The onset and retreat dates are marked by closed circle and square, respectively.

rhythm and the abundant amount was recently observed in even-number years. The pollen season from onset to retreat dates were regularly from the end of April to middle of June (Fig. 2b). The season irregularly started early in 2002 and irregularly ended late in 2005 and 2011. The season was quite short in 2006, even though the total pollen amount was larger than the average.

In order to avoid non-Gaussianity in the daily pollen amount distribution, we performed the goodness-of-fit test for normality by skewness and kurtosis[1]. Consequently, the distribution of logarithm of pollen amount was much closer to the Gaussian than the distribution of the pollen amount itself. Hence we replaced zero-pollen count with 0.1 and then took the common logarithm of the daily pollen amount [hereafter log($p$)] before the multiple regression analysis.

Developing a statistical relation on daily variations of pollen amount by multiple regression analysis, the auto-correlation function gave us a clue to look for an important predictor. The auto-correlation with a lag of 1 day of log($p$) attained 0.76 (Fig. 3). This indicates that log($p$) based on *in situ* observation can be directly used for its daily forecast. The simple linear regression of log($p$) with the lag onto log($p$) without lags brought one of statistical relations as

$$\log p(t) = 0.76 \log p(t-1) + 0.17, \tag{1}$$

where $p(t)$ denotes the pollen amount at the day $t$. This relation actually gave a reasonable forecast for a couple of days, and it is expected that this can be slightly modified by an addition of daily climatic variables as the predictors that are possibly related to the pollen amount. Potential climatic variables are mean temperature $w_1$, maximum temperature $w_2$, mininum temperature $w_3$, rainfall $w_4$, solar radiation $w_5$, relative humidity $w_6$, or/and mean wind speed $w_7$. The daily temperature difference $w_8(t) = w_1(t) - w_1(t-1)$ and the diurnal temperature range in a day $w_9 = w_2 - w_3$ are also thought as possible predictors. A more reasonable statistical equa-

[1] According to the formula given by Metz (1991), the effective degree of freedom (DOF) is a function of decorrelation time. Based on the auto-correlation function of pollen amount logarithm (Fig. 3) with their formula, the effective DOF for variance is 90.6 for its logarithm and is much less than the total observed days. Since there are no simple formula for the effective DOF for skewness or kurtosis, we substituted the effective DOF for vaiance to that for skewness and kurtosis.

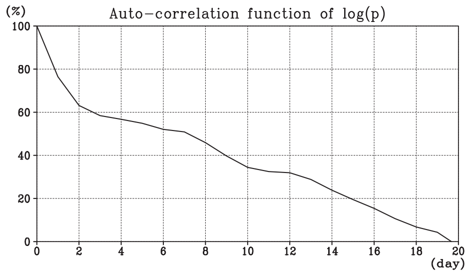


Fig. 3. Auto-correlation function (%) of the common logarithm of daily birch pollen amount.

tion than Eq. (1) may be developed with predictors selected from above as

$$\log p(t) = \alpha \log p(t-1) + \Sigma_{\text{selected } k}\, \beta_k w_k(t) + \varepsilon, \quad (2)$$

where $\alpha$ and $\beta_k$ are partial regression coefficients and $\varepsilon$ is residual. Climatic variables at the forecast day $t$ are based on observed data in this paper, but they would be practically given from a short-term weather forecast.

Table 1 shows a Pearson's correlation coefficient matrix among predictand and possible predictors in the multiple regression analysis. Since mean temperature $w_1$, maximum temperature $w_2$, mean wind speed $w_7$ were uncorrelated with the pollen amount, these were first rejected as a predictor. Moreover, the relative humidity $w_6$ was correlated with the log($p$) at the previous day. This means that the relative humidity cannot be used as the second predictor together with the log($p$) at the previous day. For a similar reason, we rejected the daily minimum temperature as well. The remaining possible predictors $w_4$, $w_5$, $w_8$, and $w_9$ were mutually correlated, and then we selected one of them in order to improve the relation of Eq. (1). Table 2 shows the result of one-day hindcast prediction by the multiple regression relations with two predictors of the lagged pollen amount and one of remaining climatic variables above. All the bivariate regressions reasonably improved the statistical relation compared with the simple regression. The case for the use of diurnal temperature range as the second predictor provided the best prediction among them. Consequently we decided to fix the multiple regression relation as

$$\log p(t) = 0.785 \log p(t-1) + 0.094\, w_9(t) + 0.127. \quad (3)$$

It is remarked that the trivariate regression of

$$\log p(t) = 0.753 \log p(t-1) - 0.041\, w_3(t) + 0.097\, w_9(t) + 0.541, \quad (4)$$

Table 1. Pearson's correlation coefficient matrix (%) among the predictand and possible predictors in the multiple regression analysis. The predictand log $p(t)$ is denoted by $p_1$ and the most important predictor log $p(t-1)$ is denoted by $p_0$. Other possible predictors are mean temperature $w_1$, maximum temperature $w_2$, mininum temperature $w_3$, rainfall $w_4$, solar radiation $w_5$, relative humidity $w_6$, mean wind speed $w_7$, daily temperature difference $w_8$, and the diurnal temperature range in a day $w_9$. The shading denotes the statistical significance at 5% level with the effective degree of freedom being 90.6.

| $p_0$ | $w_1$ | $w_2$ | $w_3$ | $w_4$ | $w_5$ | $w_6$ | $w_7$ | $w_8$ | $w_9$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| 76.5 | −11.6 | 3.7 | −30.8 | −22.3 | 23.0 | −40.8 | 5.0 | 32.7 | 20.8 | $p_0$ |
| | −11.4 | −3.3 | −21.3 | −9.2 | 1.5 | −29.2 | 13.2 | 14.7 | −3.6 | $p_1$ |
| | | 92.5 | 85.0 | −17.5 | 24.7 | −15.0 | −21.0 | 43.0 | 33.4 | $w_1$ |
| | | | 63.9 | −29.5 | 45.9 | −32.9 | −28.7 | 71.8 | 44.6 | $w_2$ |
| | | | | 2.3 | −8.5 | 23.2 | −8.5 | −7.6 | 3.0 | $w_3$ |
| | | | | | −50.0 | 39.4 | 17.4 | −40.3 | −31.5 | $w_4$ |
| | | | | | | −55.2 | −8.6 | 67.2 | 44.0 | $w_5$ |
| | | | | | | | 1.2 | −63.6 | −40.5 | $w_6$ |
| | | | | | | | | −29.6 | −16.8 | $w_7$ |
| | | | | | | | | | 55.5 | $w_8$ |

Table 2. Multiple regression relations with two predictors of the lagged pollen amount and a single climatic variable. Partial regression coefficients $\alpha$ and $\beta_k$ and the residual $\varepsilon$ are given in Eq. (2). The correlation coefficient (%) between observation and one-day hindcast prediction by the relation are shown in the rightmost column.

| $k$ (climatic variable) | $\alpha$ | $\beta_k$ | $\varepsilon$ | Correlation with observation |
|---|---|---|---|---|
| 4 (rainfall) | 0.762 | −0.033 | 0.207 | 78.0 |
| 5 (solar radiation) | 0.773 | 0.024 | −0.321 | 79.5 |
| 8 (daily temperature difference) | 0.744 | 0.052 | −0.313 | 79.6 |
| 9 (diurnal temperature range) | 0.785 | 0.094 | 0.127 | 80.0 |

actually gave a bit better statistical relation, but we did not use this because the minimum temperature $w_3$ is not entirely independent of the lagged pollen amount log[$p(t-1)$] with their correlation being near the statistically significance level (Table 1).

One-day hindcast prediction for the pollen amount was then performed based on the multiple regression relation of Eq. (3) with the observed log($p$) and the observed temperature at the prediction day. Figure 4 shows the Pearson's correlation coefficient between observation and hindcast prediction for each year. It was higher than 0.6 throughout the period, with surpassing the statistical significant level. The best prediction was realized for year 2006 with the correlation coefficient being 87.1%, while the prediction was worst for year 2009 with the correlation coefficient being 61.0%. The interannual variation of pollen amount predictability seemed biennial and was closely related to the total pollen amount (Fig. 2). The daily pollen amount was more predictable in abundant years of 2004, 2006, 2008, 2010, and 2011, while it was less predictable in poor-yield years of 2003, 2007 and 2009. Hence, combined with a prediction for annual-sum pollen amount established by Yasaka et al. (2009), the statistical method that we proposed in this paper could be utilized in a practical application to the short-term pollen prediction during the flowering season.

Finally, in order to exclude the effect of predicting year from the multiple regression relation, we performed a one-day hindcast prediction for year 2011 by using a statistical relation based on 10-year pollen data from 2001 to 2010. The statistical relation excluding 2011 was however almost same as Eq. (3). The time-series of sequential one-day prediction was then quite similar to that of observation (Fig. 5); their correlation coefficient is 84.2% (cf. Fig. 4). Extending this statistical prediction for more days, the correlation was monotonically decreasing, but it was still near 0.75 for three-day prediction in this case. It is remarked that multiple-day prediction would not change if we used a single regression relation by Eq. (1).

## 4. Concluding remarks

The statistical analysis on daily variations of the birch pollen amount has been developed. By utilizing the airborne pollen survey in Sapporo from 2001 to 2011, we found that the lagged pollen amount was chosen as the most important predictor. Among

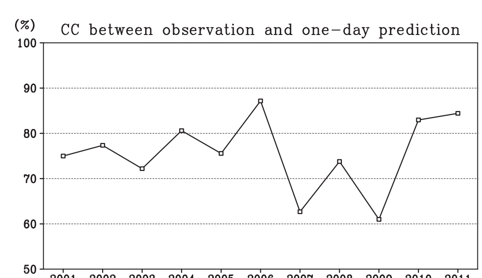


Fig. 4. Correlation coefficient between observation and one-day hindcast prediction by Eq. (3) for each year.

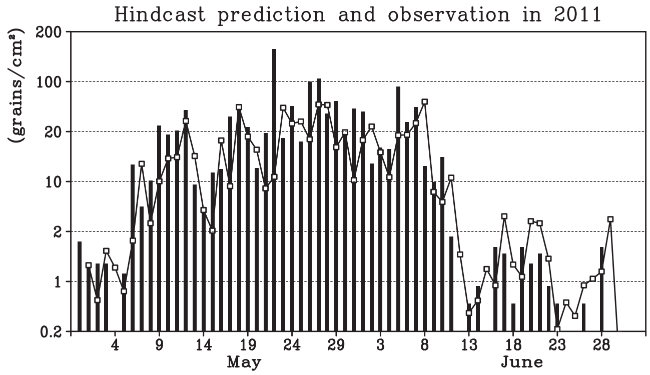


Fig. 5. Time-series of the pollen amount (grains/cm$^2$) from 30 April to 30 June in 2011 in the bar graph. One-day hindcast of the pollen amount in 2011, shown in the line with square marks, by the multiple regression relation based on the pollen data from 2001 to 2010.

several climatic variables such as daily, minimum and maximum temperature, sunshine duration, wind speed, rainfall, and relative humidity, we selected the diurnal temperature range as the second predictor. A simple equation with these two predictors based on the multiple regression analysis provided a reasonable statistics on daily pollen variations. The statistical equation proposed in this paper was better fitted to the observations in abundant years than in poor-yield years. As stated in Introduction, the annual-sum pollen amount can be predicted based on male catkin numbers observed in the previous year. Therefore, in years where the total pollen amount would be abundant by this prediction method, the statistical equation gave daily variations of the pollen amount; in years where the total pollen amount would be poor, the statistical equation was not necessary because the daily variations were likely to be smaller than in normal years.

Practically a qualitative warning rather than numerical information may be helpful for patients who suffer from the pollen allergy. This could be realized if one reasonably transformed the numerical information to a qualitative warning message. Now we attempted to do it in a one-day hindcast prediction of pollen amount by the criterion of 10 grains/cm$^2$ on the glass slide. As a result there were 159 hit days, 61 missing days, 33 false-alarmed days. Since the skill score was then 63%, this simple practical warning system may moderately work. The score of course depends on the criterion and it decreases if the criterion level increases. However, this warning method can be posed as a useful idea in the short-term pollen prediction in a practical sense.

The statistical relation established here based on the pollen data in Sapporo is not always applicable to the other region. For example, the wind speed is not a suitable predictor in Sapporo, but it becomes a more important one in the area where most of the pollen grains is advected from the upstream [See Kawashima and Takahashi (1995) for airborne cedar pollen analysis]. As have reviewed in Introduction, moreover, another way to predict the pollen amount is to run an atmospheric transport model. In contrast with this statistical manner, the atmospheric transport model can directly predict the pollen amount with a physical principal that can be universally applied to any area. Even with a particular set of parameters as accurate as possible, however, a model cannot always realize a spatio-temporally accurate prediction of the pollen amount. We could then extend our research toward more accurate daily prediction of pollen amount by combining the statistical relation with an atmospheric transport model.

## Acknowledgments

The first author is partly supported by Research Program on Climate Change Adaptation funded by the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

## References

Baba, K., and K. Nakae, 2008: National epidemiological survey of nasal allergy 2008 (compared with 1998) in otolaryngologists and their family members. *Progress in Medicine*, **28**, 2001−2012 (in Japanese).

Clot, B., 2001: Airborne birch pollen in Neuchâtel (Switzerland): Onset, peak and daily patterns. *Aerobiologia*, **17**, 25−29.

Fehér, Z., and M. Járai-Komlódi, 1998: A new weather factor predicting airborne pollen concentration: Péczely's macrosynoptic weather types. *Aerobiologia*, **14**, 171−177.

Gotoda, H., S. Maguchi, H. Kawahara, Y. Terayama, and S. Fukuda, 2001: Spring time pollinosis and oral allergy syndrome in Sapporo. *Auris Nasus Larynx*, **28**, 49−52.

Kawashima, S., and Y. Takahashi, 1995: Modelling and simulation of mesoscale dispersion processes for airborne cedar pollen. *Grana*, **34**, 142−150.

Kobayashi, S., S. Takeuchi, and M. Yasaka, 2013: Trends in annual counts of Betula pollen from six cities in Hokkaido. *Japanese J. Palynology*, **59**, 59−67.

Larsson, K. A., 1993: Prediction of the pollen season with a cumulated activity method. *Grana*, **32**, 111−114.

Linkosalo, T., H. Ranta, A. Oksanen, P. Siljamo, A. Luomajoki, J. Kukkonen, and M. Sofiev, 2010: A double-threshold temperature sum model for predicting the flowering duration and relative intensity of Betula pendula and B. pubescens. *Agr. Forest Meteor.*, **150**, 6−11.

Metz, W., 1991: Optimal relationship of large-scale flow patterns and the barotropic feedback due to high-frequency eddies. *J. Atmos. Sci.*, **48**, 1141−1159.

Porsbjerg, C., A. Rasmussen, and V. Backer, 2003: Airborne pollen in Nuuk, Greenland, and the importance of meteorological parameters. *Aerobiologia*, **19**, 29−37.

Ranta, H., E. Kubin, P. Siljamo, M. Sofiev, T. Linkosalo, and A. Oksanen, 2006: Long distance pollen transport cause problems for determining the timing of birch pollen season in Fennoscandia by using phenological observations. *Grana*, **45**, 297−304.

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

Ribeiro, H., M. Cunha, and I. Abreu, 2003: Airborne pollen concentration in the region of Braga, Portugal, and its relationship with meteorological parameters. *Aerobiologia*, **19**, 21−27.

Siljamo, P., and co-authors, 2013: A numerical model of birch pollen emission and dispersion in the atmosphere: Model evaluation and sensitivity analysis. *Int. J. Biometeor.*, **57**, 125−136.

Sofiev, M., P. Siljamo, H. Ranta, and A. Rantio-Lehtimäki, 2006: Towards numerical forecasting of long-range air transport of birch pollen: Theoretical considerations and a feasibility study. *Int. J. Biometeor.*, **50**, 392−402.

Sofiev, M., P. Siljamo, H. Ranta, T. Linkosalo, S. Jaeger, A. Rasmussen, A. Rantio-Lehtimaki, E. Severova, and J. Kukkonen, 2013: A numerical model of birch pollen emission and dispersion in the atmosphere. Description of the emission module. *Int. J. Biometeor.*, **57**, 45−58.

Yamamoto, T., K. Asakura, H. Shirasaki, T. Himi, H. Ogasawara, S. Narita, and A. Kataura, 2004: Oral allergy syndrome among patients with birch pollinosis in Sapporo. *Japanese J. Allergology*, **53**, 435−442 (in Japanese).

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

WHO, 2003: Phenology and human health: allergic disorders. Copenhagen. WHO Regional Office for Europe 55 pp.

*Manuscript received 31 July 2014, accepted 22 September 2014*
*SOLA*: *https://www.jstage.jst.go.jp/browse/sola/*