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

Masaru Inatsu¹, Satoshi Kobayashi², Shinji Takeuchi³, and Anna Ohmori¹

¹Faculty of Science, Hokkaido University, Sapporo, Japan
²Hokkaido Institute of Public Health, Sapporo, Japan
³School 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 coefficient with observation being 0.80. Moreover, the equation was better fitted to the observations in abundant years than in poor-year 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.)

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 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 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 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 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 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 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 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.
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 aolft 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-cm² 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/cm² has been observed for two consecutive days. The retreat date is defined as the final observed date 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/cm². 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 biennial 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. 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

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

(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 w₁, maximum temperature w₂, minimum temperature w₃, rainfall w₄, solar radiation w₅, relative humidity w₆, or/and mean wind speed w₇. The daily temperature difference w₈(t) = w₄(t) − w₄(t − 1) and the diurnal temperature range in a day w₉ = w₅ − w₆ are also thought as possible predictors. A more reasonable statistical equa-
Consequently we decided to fix the multiple regression relation as the second predictor provided the best prediction among them. The case for the use of diurnal temperature range as a predictor. The daily hindcast prediction by the multiple regression relations with two predictors of the lagged pollen amount and one of remaining climatic variables above as

$$\log p(t) = \alpha \log p(t-1) + \sum_{k=1}^{9} \beta_k w_k(t) + \varepsilon,$$

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 $w_3$, $w_4$, $w_5$, 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

$$\log p(t) = 0.785 \log p(t-1) + 0.094 w_3(t) + 0.127.$$  

It is remarked that the trivariate regression of

$$\log p(t) = 0.753 \log p(t-1) - 0.041 w_3(t) + 0.097 w_6(t) + 0.541,$$

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.

<table>
<thead>
<tr>
<th>$k$ (climatic variable)</th>
<th>$\alpha$</th>
<th>$\beta_k$</th>
<th>$\varepsilon$</th>
<th>Correlation with observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 (rainfall)</td>
<td>0.762</td>
<td>-0.033</td>
<td>0.207</td>
<td>78.0</td>
</tr>
<tr>
<td>5 (solar radiation)</td>
<td>0.773</td>
<td>0.024</td>
<td>-0.321</td>
<td>79.5</td>
</tr>
<tr>
<td>8 (daily temperature difference)</td>
<td>0.744</td>
<td>0.052</td>
<td>-0.313</td>
<td>79.6</td>
</tr>
<tr>
<td>9 (diurnal temperature range)</td>
<td>0.785</td>
<td>0.094</td>
<td>0.127</td>
<td>80.0</td>
</tr>
</tbody>
</table>

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 78.1%, 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-year 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
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 allergy. This could be realized if one reasonably transformed the numerical information to a qualitative warning message. Now

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.

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


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