

A biology-driven receptor model for daily pollen allergy risk in Korea based on Weibull probability density function

Kyu Rang Kim¹  · Mijin Kim¹ · Ho-Seong Choe¹ · Mae Ja Han¹ · Hye-Rim Lee¹ · Jae-Won Oh² · Baek-Jo Kim¹

Received: 13 October 2015 / Revised: 24 June 2016 / Accepted: 27 June 2016
© ISB 2016

Abstract Pollen is an important cause of respiratory allergic reactions. As individual sanitation has improved, allergy risk has increased, and this trend is expected to continue due to climate change. Atmospheric pollen concentration is highly influenced by weather conditions. Regression analysis and modeling of the relationships between airborne pollen concentrations and weather conditions were performed to analyze and forecast pollen conditions. Traditionally, daily pollen concentration has been estimated using regression models that describe the relationships between observed pollen concentrations and weather conditions. These models were able to forecast daily concentrations at the sites of observation, but lacked broader spatial applicability beyond those sites. To overcome this limitation, an integrated modeling scheme was developed that is designed to represent the underlying processes of pollen production and distribution. A maximum potential for airborne pollen is first determined using the Weibull probability density function. Then, daily pollen concentration is estimated using multiple regression models. Daily risk grade levels are determined based on the risk criteria used in Korea. The mean percentages of agreement between the observed and estimated levels were 81.4–88.2 % and 92.5–98.5 % for oak and Japanese hop pollens, respectively. The new models estimated daily pollen risk more accurately than the original statistical models because of the newly integrated biological response

curves. Although they overestimated seasonal mean concentration, they did not simulate all of the peak concentrations. This issue would be resolved by adding more variables that affect the prevalence and internal maturity of pollens.

Keywords Pollen · Allergy · Concentration · Risk warning · Weibull PDF

Introduction

Reactions to pollen such as allergic rhinitis and allergic conjunctivitis are disorders of a hypersensitive immune system. These reactions are caused by specific airborne pollen species. Medical treatments, such as taking antihistamines before severe symptoms appear, can preemptively reduce the risk of allergic symptoms. Economic losses because of allergy-related reduced work efficiency have been estimated to be as much as \$200 million USD per year in Korea (Kim et al. 2009). It has been reported that approximately 10–25 % of Koreans are affected by airborne pollen (Bousquet et al. 2008; Oh et al. 2006). The lifetime prevalence of asthma was reported to be 5.3 % in 2000, which was a significant increase from 1995 when the reported prevalence was 2.7 % (Hong et al. 2004). Allergic rhinitis among the young children is increasing even more rapidly; its prevalence in elementary school students increased from 16 % in 1995 to 20 % in 2000 and 28 % in 2006 (Hong et al. 2008; Jee et al. 2009). In 2009, the prevalence of allergic rhinitis in preschool children was 40.7 % (Kim et al. 2013), and the sensitization rate of pollen allergen in students with allergic rhinitis was 41.3 % (Hwang et al. 2013). During the last 30 years, skin reactivity to tree pollens of respiratory allergic patients was significantly increased to 36.4 % in the 2010s from 8.8 % in the 1980s (Park et al. 2014).

✉ Kyu Rang Kim
krk9@kma.go.kr

¹ National Institute of Meteorological Sciences, Seogwipo-si, Republic of Korea

² College of Medicine, Hanyang University, Seoul, Republic of Korea

The potential risk of pollen allergies, therefore, is expected to increase in the future because of this increase in young patients (Oh et al. 2012) in addition to the prolonging and intensification of the pollen season associated with global warming (D'Amato et al. 2015). In Japan, more than 25 % of the population suffers from allergies to Japanese cedar pollen (Okamoto et al. 2009). Daily forecasts for major allergenic pollens are provided in Japan as well as in many developed countries in Europe and North America.

Since 1997, daily concentrations of pollen in Korea have been recorded by the Korea Meteorological Administration (KMA) in collaboration with the Korean Academy of Pediatric Allergy and Respiratory Disease. Pollen calendars for each species and location were developed based on 10-day average pollen concentration (Park et al. 2008; Oh et al. 2012). The KMA observes daily concentrations of pollen in 12 cities in Korea (Fig. 1).

Weather conditions play a crucial role in the spread of pollen by influencing processes such as flowering, release, transportation, and deposition of pollen; therefore, weather conditions are key variables in the prediction of pollen allergy risks (Lee et al. 2012; Kim et al. 2014). There are two categories of models for forecasting pollen based on meteorological conditions: one category estimates the timing of pollen season, and the other estimates daily concentrations of pollen. Models in the former category estimate either the date of the first pollen appearance (Driessen et al. 1989), or the dates of the start, peak, and end of pollen seasons (Larsson 1993). For predictors, they analyze air temperature data as well as data for previously reported pollen concentrations. Jato et al. (2006) reviewed various criteria used to constrain the dates of pollen seasons. Definitions of pollen season include, but are not limited to, when the sum of daily mean pollen concentrations reaches 1, 2.5, or 5 % of the yearly total pollen, or when the daily pollen count exceeds 3, 20, or 30 grains m^{-3} . Jato et al. concluded that the criteria selected should meet the purpose of the survey.

Many studies have attempted to estimate these forecasts quantitatively. For example, seasonal and annual total pollen counts have been estimated based on cumulative air temperature and rainfall (Emberlin et al. 1999) and on monthly total rainfall and mean humidity (Subiza et al. 1992), respectively. Recent studies have estimated daily pollen concentrations using a number of methods, e.g., cypress tree pollen concentration predicted based on daily maximum temperature (Sabariego et al. 2012); birch and grass pollen concentration predicted based on daily air temperature, solar radiation, and wind speed (Khwarahm et al. 2014); and grass pollen concentration based on weather conditions and previous pollen observations (De Weger et al. 2014). Models that estimate daily pollen concentrations are typically developed through multiple regression analysis that incorporates stepwise methods for the selection of independent variables (Toro et al. 1998).

Pollen season can be divided into two sub-seasons: the pre- and post-peak periods. These two periods either were analyzed separately (Khwarahm et al. 2014) or only the pre-peak season was targeted for model development (Galan et al. 1995; Toro et al. 1998) because this period is considered to be more important. Values representing meteorological conditions can be used to forecast pollen concentration and to issue allergy warnings.

There are two modeling approaches for forecasting the daily variations of pollen concentrations: source-oriented methods and receptor-oriented methods. Receptor-oriented pollen forecasting is the focus of this study. Because of the complex nature of the distribution of vegetation and the physiological characteristics of the allergenic plants, and receptor-oriented methods are more relevant for forecasting daily observed pollen concentrations at street level. This approach estimates concentration in a systematic way based on observed and forecasted variables without calculating source and dispersion conditions (Norris-Hill 1995). In contrast, the source-oriented approach is based on atmospheric transport and dispersion modeling and associated emission modeling.

Fig. 1 Pollen observational network in Korea, which comprises twelve Burkard 7-day recording volumetric spore traps (Map: Google Earth 2016)

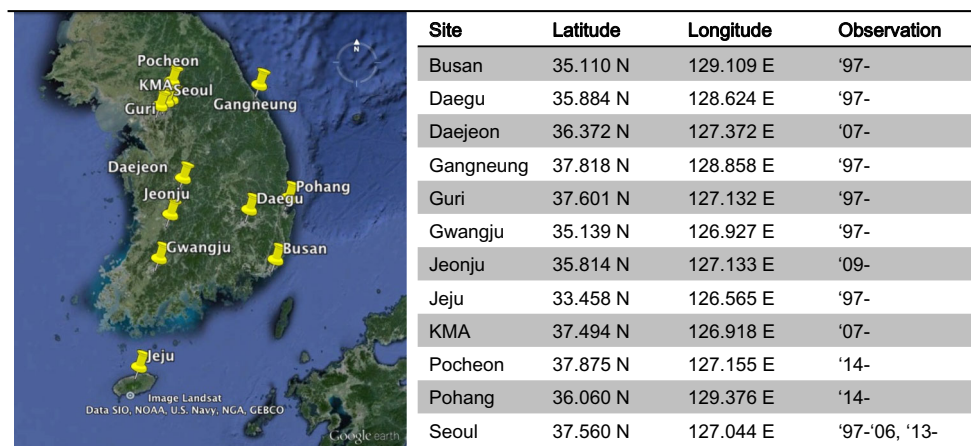


Table 1 Variables utilized in the development of the model

Variable name	Unit	Description
Ta	°C	Daily mean air temperature
Tmax	°C	Daily maximum air temperature
Tmin	°C	Daily minimum air temperature
DTR	°C	Daily temperature range (Tmax–Tmin)
DD	°C	Degree-days (accumulated temperature above 0 °C from 1 January)
RH	%	Daily mean relative humidity
WS	m s ⁻¹	Daily mean wind speed
Sh	hours	7-day total sunshine hours
P	mm	Daily total precipitation
Ph	hours	Daily total hours of measurable precipitation (hourly $P \geq 0.1$ mm)

Furthermore, receptor-oriented models can be applied in more complex source-oriented models through improving the modeling of emissions.

Until 2013, two-day forecasts of daily pollen warnings were provided for seven cities in Korea: Busan, Daegu, Daejeon, Gangneung, Gwangju, Jeju, and Seoul. The forecasting models were developed from the multiple regression analysis based on daily observed pollen concentrations and meteorological variables. The models were developed separately for different periods and pollen species; tree pollen models were developed for April and May, while weed pollen models were developed for September and October. Tree pollens were categorized as pine or other species (Kim et al. 2012). These models were able to predict daily pollen risk levels. However, the parameters of the regression models do not necessarily explain the response of the biological processes of pollen production. In addition, there are spatio-temporal limitations to the application of these models: such a model developed for a specific location and month cannot be generalized to the other places or times.

For this study, new models were developed in two parts. The first part is a set of probability density functions (PDFs) that can be used to estimate the optimal meteorological

Table 2 Re-fitted original regression models using the observation from 1997 to 2009 for the daily pollen concentrations of tree and weed species

Site	Species	Month	Daily pollen concentration model (1997–2009)	R ²
Busan	Trees	Apr.	$-0.51 + 0.002*DD + 0.019*P - 0.027*Ph$	0.116
	Trees	May	$2.64 - 0.001*DD - 0.011*Sh$	0.057
	Weeds	Sep.	$1.70 - 0.001*DD + 0.026*Tmax + 0.011*Sh$	0.086
	Weeds	Oct.	$-0.080 + 0.017*Tmin + 0.003*Sh$	0.019
Daegu	Trees	Apr.	$0.036 + 0.002*DD - 0.14*WS$	0.117
	Trees	May	$2.86 - 0.035*Ta - 0.001*DD - 0.007*RH - 0.16*WS$	0.080
	Weeds	Sep.	$3.50 - 0.001*DD - 0.012*RH + 0.007*Sh$	0.065
	Weeds	Oct.	$-0.11 + 0.051*Ta - 0.006*RH$	0.103
Daejeon	Trees	Apr.	$-0.23 + 0.13*Tmin + 0.43*WS - 0.06*Ph$	0.337
	Trees	May	$3.73 - 0.003*DD + 0.040*Tmin - 0.024*Ph$	0.574
	Weeds	Sep.	No variable met the 0.15 significance level for entry into the model	–
	Weeds	Oct.	$-1.30 + 0.044*DTR + 0.021*RH + 0.21*WS - 0.058*P - 0.012*Sh$	0.278
Gang-neung	Trees	Apr.	$-0.15 + 0.002*DD - 0.11*WS + 0.006*Sh$	0.152
	Trees	May	$0.59 + 0.006*Sh$	0.010
	Weeds	Sep.	$0.71 + 0.005*Sh$	0.009
	Weeds	Oct.	$2.85 - 0.001*DD - 0.012*Tmax$	0.035
Gwangju	Trees	Apr.	$0.41 + 0.001*DD - 0.009*RH + 0.20*WS - 0.025*Ph$	0.099
	Trees	May	$3.36 - 0.25*Ta - 0.001*DD + 0.11*Tmax + 0.10*Tmin - 0.018*RH + 0.030*Ph - 0.004*Sh$	0.181
	Weeds	Sep.	$0.085 + 0.033*Tmax$	0.017
	Weeds	Oct.	$-0.012 + 0.024*Ta$	0.020
Jeju	Trees	Apr.	No variable met the 0.15 significance level for entry into the model	–
	Trees	May	$2.14 - 0.001*DD - 0.029*Tmin$	0.118
	Weeds	Sep.	$4.02 - 0.001*DD$	0.032
	Weeds	Oct.	$3.18 - 0.001*DD - 0.005*RH - 0.003*Sh$	0.059
Seoul	Trees	Apr.	$0.069 + 0.003*DD - 0.011*RH + 0.033*Ph$	0.191
	Trees	May	$1.43 - 0.001*DD - 0.009*P$	0.032
	Weeds	Sep.	$-0.35 + 0.071*Tmin + 0.11*WS - 0.013*Ph$	0.095
	Weeds	Oct.	$2.23 - 0.001*DD + 0.018*Tmax + 0.14*WS - 0.017*Ph - 0.011*Sh$	0.117

Table 3 Criteria for risk grade levels of tree and weed pollen concentrations (grains m⁻³) in Korea

Risk grade	Trees	Weeds
Mild	0–49	0–49
Moderate	50–99	50–99
Severe	100–199	100–299
Extreme	≥200	≥300

conditions for pollen. The second part is a multiple regression model for forecasting daily concentrations of pollen based on the daily probabilities that were estimated in the first part. Thus, the optimal conditions for pollen concentration were modeled based on biologically meaningful functions and were integrated as regression models for daily pollen concentration.

Materials and methods

Daily pollen concentrations reported by the pollen observational network in Korea

Airborne pollen was collected using 7-day recording volumetric spore traps (Burkard Manufacturing Co. Ltd., Hertfordshire, UK), which were installed at the 12 sites in Korea (Fig. 1). The sampling drums from each site were collected weekly at a pre-determined time. The samples were stained with Calberla's fuchsin dye by combining 10 ml of glycerin, 20 ml of 95 % ethanol, 30 ml of distilled water, and 0.2 ml of basic fuchsin with each sample. Pollen grains were then identified and counted under a light microscope (×400). These counts were converted to yield daily concentration values (grains m⁻³). Allergenic pollen species identified in Korea include trees such as oak, pine, birch, alder, elm, plane tree, maple, and Japanese cedar, and weeds such as ragweed, Japanese hop, sagebrush, and plantain, as well as grass species. For the development of the statistical models, daily records from 1997 to 2009 from the seven sites (Busan, Daegu, Daejeon, Gangneung, Gwangju, Jeju, and Seoul) were considered. Data from 2010 to 2012 were selected for the

evaluation. The missing period of the Seoul data was filled with the KMA data from 2007 to 2012.

Weather data

The KMA weather stations near each pollen sampling site were selected to provide data to represent the daily weather conditions at each site. The weather variables analyzed were daily mean, maximum, and minimum air temperatures, daily temperature range, degree-days (accumulated temperature above 0 °C from the first day of the year), daily mean relative humidity, daily mean wind speed, 7-day total sunshine hours, daily total rainfall, and rainfall hours (Table 1).

Original pollen forecasting models for the operational service in the KMA

The original pollen forecasting system of the KMA became operational in 2008. It was developed based on multiple regression models, which were separately developed for the seven sites (Busan, Daegu, Daejeon, Gangneung, Gwangju, Jeju, and Seoul) and 4 months (April, May, September, and October) using observational data from 2001 to 2006 (Kim et al. 2012). These models used daily weather conditions as independent variables and estimated daily concentrations of tree pollens in April and May and of weed pollens in September and October. Several meteorological variables, listed in the weather data section, were considered for the development of these models. Regression models were developed using the stepwise and GLM procedures of SAS software. All of these regression models used the identity link function, which means that the residuals were assumed to be normally distributed. Models were developed separately for each site and month. As indicated in Table 2, for comparison with the new models proposed below, the original models' equations were re-fitted using the same methods, but with data from 1997 to 2009. The parameters determined with the models do not necessarily represent collective responses to weather conditions. Species of the allergenic trees and weeds

Table 4 Percentages of agreement for the daily pollen risk grades using the re-fitted original regression models with data collected by the KMA from 2010 to 2012

Species	Month	Site/accuracy (unit %)							
		Busan	Daegu	Daejeon	Gangneung	Gwangju	Jeju	Seoul	No. of days
Trees	Apr.	48.9	81.1	44.4	78.9	39.8	–	72.2	90
Trees	May	58.1	69.9	48.4	55.9	68.8	61.3	55.9	93
Weeds	Sep.	63.3	53.3	–	77.5	48.3	71.1	52.2	90
Weeds	Oct.	88.2	100.0	97.7	100.0	100.0	86.0	93.6	93
Mean		64.6	76.1	63.5	78.1	64.2	72.8	68.5	91.5

were not differentiated. All tree species except for pine and all weeds species were grouped together to forecast overall tree pollen and weed pollen allergy risks. Pine pollen was

excluded because it is known to be less allergenic than other tree pollens, although it is abundantly distributed on the Korean Peninsula.

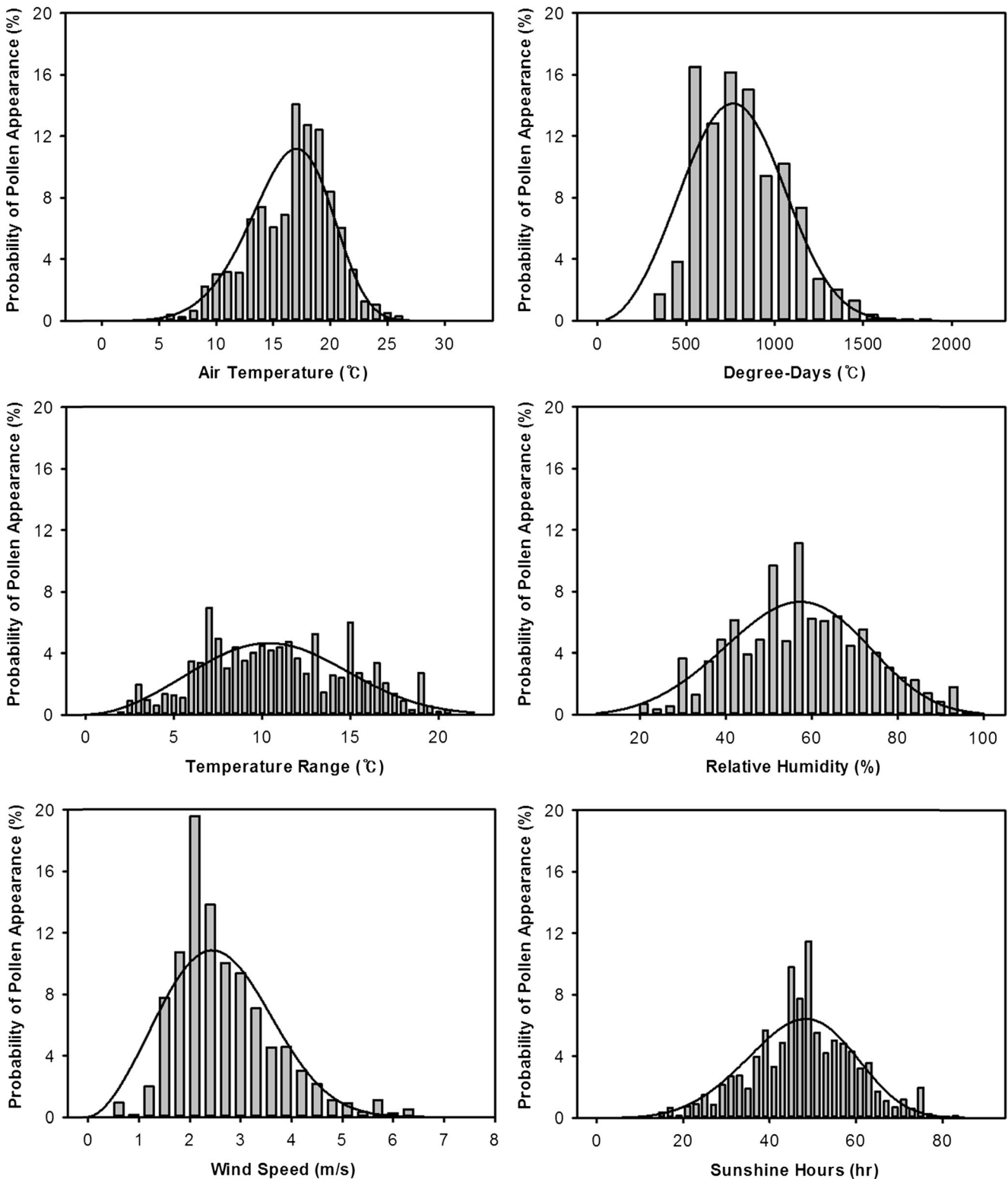


Fig. 2 Frequency distributions of observed oak pollen for all sites combined over 1997–2009 for air temperature, degree-days, daily temperature range, relative humidity, wind speed, and sunshine hours. Curves are the probability density functions for fitted Weibull distribution

The accuracy of the re-fitted original models was determined based on the percentage (0–100 %) of agreement of risk levels following the prediction of daily pollen concentrations during the period of 2010–2012. Risks were classified

into four levels on the basis of observed allergic symptoms after exposure to various levels of pollen (Table 3). The percentage of agreement ranged from 40 to 100 % accuracy depending on the site, plant species, and month; the mean

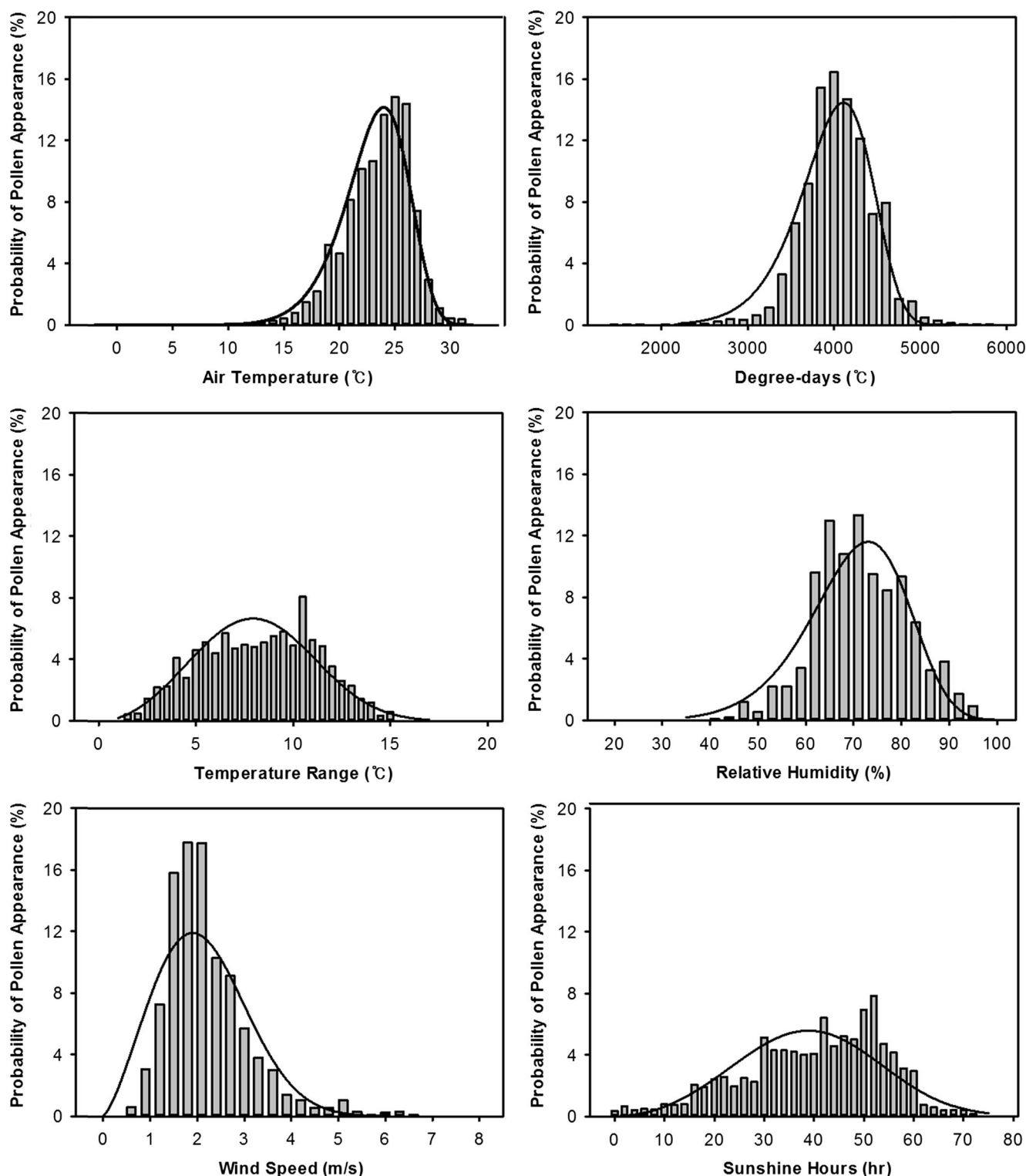


Fig. 3 Frequency distributions of observed Japanese hop pollen for all sites combined over 1997–2009 for air temperature, degree-days, daily temperature range, relative humidity, wind speed, and sunshine hours. Curves are the probability density functions for fitted Weibull distribution

Table 5 Parameter (σ and c) values determined with the Weibull probability density function (PDF) for oak and Japanese hop pollens based on daily weather elements from 1997 to 2006 for all sites

Meteorological element	Oak			Japanese hop		
	σ	c	p value*	σ	c	p value*
Ta	17.68	5.60	<0.01	24.25	9.99	<0.01
DD	864.24	3.12	<0.01	4145.87	10.65	<0.01
Tmax	23.70	5.50	<0.01	28.90	9.11	<0.01
Tmin	12.54	5.21	<0.01	20.64	9.50	<0.01
DTR	12.03	2.87	<0.01	9.06	3.01	<0.01
RH	61.74	3.97	<0.01	74.36	7.80	<0.01
WS	2.90	2.70	<0.01	2.39	2.37	<0.01
Sh	51.34	4.35	<0.01	43.74	3.23	<0.01
P	1.53	0.80	<0.01	2.29	0.64	<0.01
Ph	1.72	0.93	<0.01	1.96	0.91	<0.01

* $\Pr > \omega^2$ by the Cramer–von Mises test

agreement was 69.2 % (Table 4). Predictive accuracy was higher for weeds in October than for trees or weeds in September because site-based differences in weed species phenophases are reduced later in the season when it is beyond the most suitable period for pollen production.

A novel modeling scheme for unified pollen allergy risk

A novel approach for predicting daily pollen concentration

Each pollen count was observed under the microscope independently. Therefore, one count of the observed airborne pollen is equivalent to one sample of the daily weather conditions that have successfully produced and delivered a pollen grain from a plant to the pollen sampler and microscope. The production of pollen involves enzymatic chemical activities, with rates determined by substrate properties such as temperature and pH. The delivery of pollen depends on physical processes such as wind and rainfall. The distribution of the recorded weather conditions therefore represents the optimal overall conditions for pollen concentration in the air. Suitable weather conditions for pollen can therefore be analyzed based on their frequency distributions.

Table 6 Multiple regression models to predict daily pollen concentrations for oak and Japanese hop based on the Weibull probability (e.g., w_Ta for air temperature) for weather conditions of all sites from 1997 to 2006

	Model	R^2
Oak	$-14.97 + 196.12*w_Ta + 33,245*w_DD - 271.08*w_DTR + 1061.18*w_RH + 642.29*w_Sh$	0.035
J. hop	$-15.80 + 34,151*w_DD + 253.03*w_Tmin - 69.19*w_DTR + 187.94*w_RH + 476.50*w_Sh$	0.096

The distributions of oak and Japanese hop pollen for all sites combined from 1997 to 2009 show asymmetrical bell shapes with optimal conditions in the middle (Figs. 2 and 3). These distribution data were fitted to the Weibull probability density function (PDF, Eq. 1, curves in Figs. 2 and 3). The parameters of the function for each weather element and pollen species are listed in Table 5.

$$f(x) = \frac{c}{\sigma} \left(\frac{x}{\sigma}\right)^{c-1} e^{-\left(\frac{x}{\sigma}\right)^c} \text{ where } (x > 0, c > 0, \sigma > 0) \quad (\text{Eq.1})$$

The Weibull function has a scale parameter (σ) that represents relative optimal conditions, and a shape parameter (c) that represents the skewness. Fitted functions represent the probability of the pollen appearance under selected weather conditions. The maximum probability depends on how the conditions are clustered; if the favorable weather conditions were confined to a relatively small range, then the maximum probability values in the distribution would be relatively higher and vice versa (Figs. 2 and 3).

Table 7 Mean concentrations of observed oak and Japanese hop pollens and rainfall compensation factors calculated as the ratio between the amount of pollen accumulated during rainfall divided by the amount of pollen without rainfall

	Oak	Japanese hop
Rainfall ≥ 0.1	7.98	8.02
Rainfall < 0.1	25.77	21.45
Rainfall compensation factor	0.31	0.37

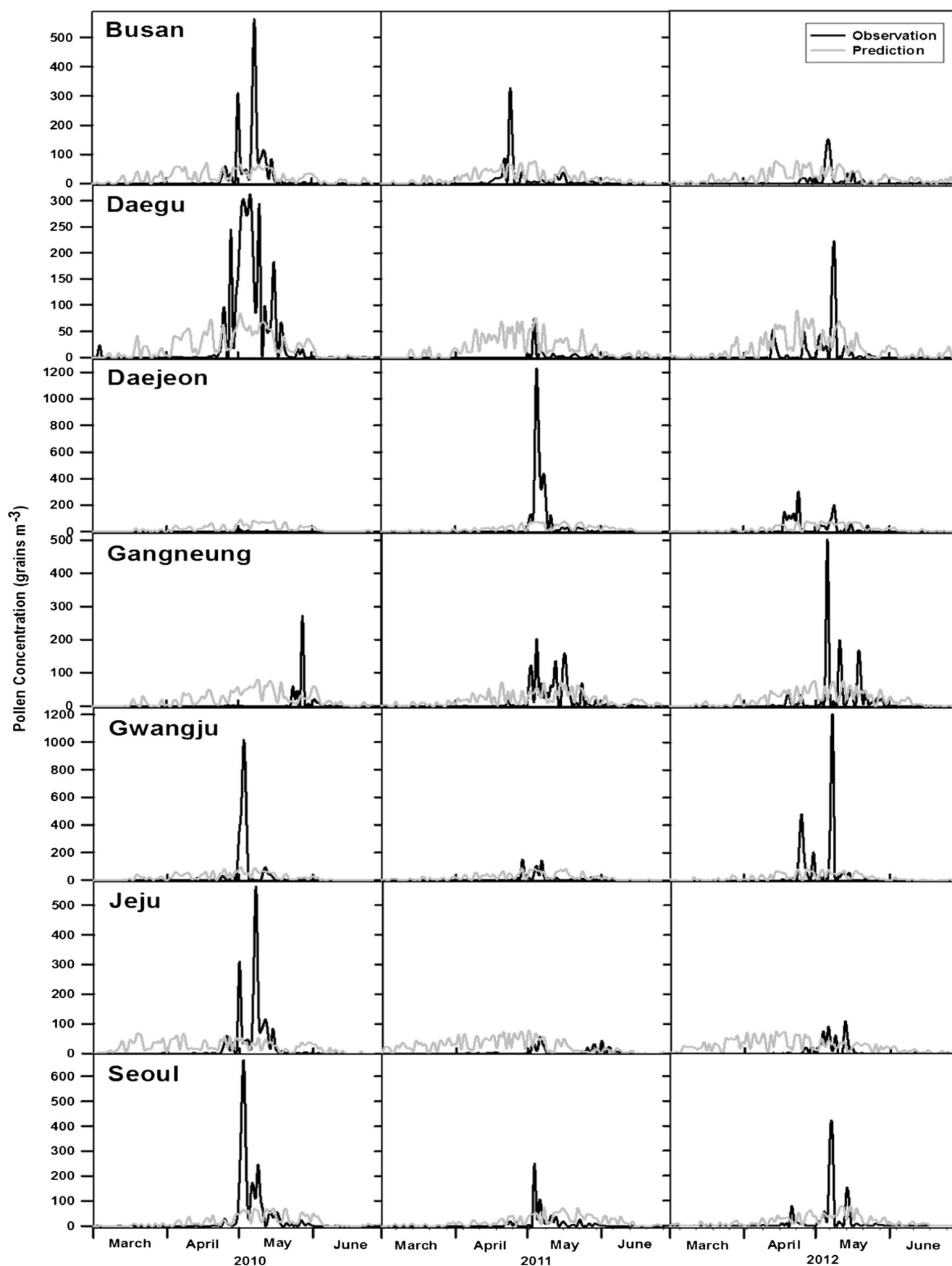


Fig. 4 Predicted and observed daily oak pollen concentrations during the evaluation period from 2010 to 2012 at the seven validation sites in Korea

Pollen concentration and risk models

Multiple regression analysis was performed to estimate daily pollen count using probabilities obtained from the Weibull PDF. Relevant points from the Weibull PDF curves calculated using daily weather values were utilized as the predictor variables instead of the original values. Daily observed pollen concentrations from 1997 to 2009 were used to identify suitable variables following the stepwise procedure of the SAS software (Table 6). Although the regression models did not include rainfall variables, the observed daily pollen concentrations were significantly lower during rainy days, which is a phenomenon known as the washout effect. Therefore, to compensate for this effect, the observed ratio of pollen concentration on rainy vs. non-rainy days was incorporated into the daily estimates of pollen concentration for rainy days (daily rainfall ≥ 0.1 mm). For daily forecasting, rainfall compensation factors were applied to daily estimates regardless of the rainfall amount. These factors were 0.31 and 0.37 for oak and Japanese hop pollens, respectively (Table 7).

Evaluation of the new models

The new models were evaluated using the daily observed data from 2010 to 2012 for the seven cities. The calculated daily risk levels from the observed pollen concentrations and estimated risks from the new models were compared. Monthly percentages of agreement were calculated for comparison with the original forecast models.

Another important characteristic of a pollen season for its allergenicity is its start date. To exclude the effects of long-range transport, the predicted and observed first dates with pollen concentrations greater than 5 % of yearly total pollen in grains m^{-3} were compared.

Results

Optimal condition models using the Weibull probability density function

Each Weibull probability density model was used to determine the distribution of optimal conditions for a pollen outbreak for the selected weather variable. The optimal ranges are indicated by the parameter σ of the Weibull PDF. For example, the σ for air temperature for oak pollen was 17.7, whereas that for Japanese hop was 24.3, and the distribution for Japanese hop was skewed to the right compared to that of oak (Table 5, Figs. 2 and 3). The fitted curves are reasonable matches for the histograms in Figs. 2 and 3. The Weibull parameters were significant at $p < 0.05$, except for the rain variables; because there were too many cases of zero

rainfall, the Weibull PDF was not able to match the observed patterns very well for rainfall.

Multiple regression models for daily pollen concentration

Daily pollen concentrations of oak and Japanese hop pollen at the seven sites from 2010 to 2012 were estimated using multiple regression models and then compared with the observed data, as shown in Figs. 4 and 5. The models fit the data for all sites. The observed concentrations of oak pollen were much higher than those of Japanese hop pollen. Overall, predictions also matched the observational data, and Japanese hop pollen concentration was fairly well-predicted across sites and years, but the oak pollen concentration predictions were overestimated early in the season and underestimated later.

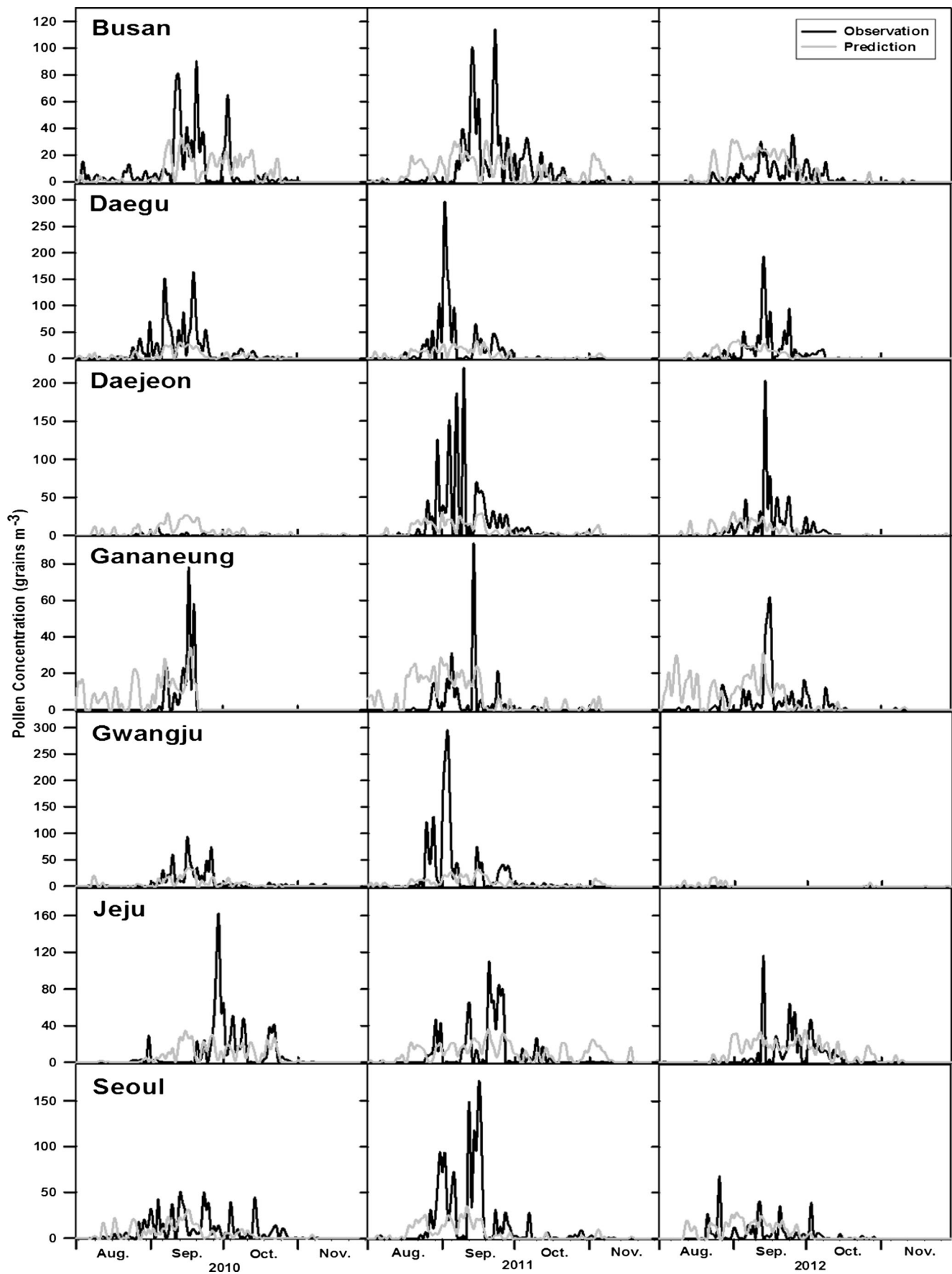
The comparison of the observed and predicted start dates with cumulative pollen concentrations greater than 5 % of the annual total pollen indicates that the new models were able to predict the beginning of pollen seasons in most cases (Table 8). However, the difference between the predicted and actual date varied across the sites and years. The predicted start date for oak pollen was estimated to be much earlier than was observed, whereas that for Japanese hop pollen was relatively accurate.

Pollen concentration and risk models

In Korea, pollen allergy risk levels have been divided into four grades based on the relationship between daily pollen concentration and the severity of patients' symptoms (Kim et al. 2009; Oh et al. 2009; Han et al. 2012; Kim et al. 2014). These grades are used as standard warning levels in daily pollen allergy forecasting by the KMA (Kim et al. 2012). It is reasonable to compare the observed and predicted pollen risk grades for the assessment of model performance. The risk grade criteria (Table 3) were applied to the estimated daily pollen concentration values to determine daily pollen risk grades.

The differences between predicted and observed risk grades are shown in Table 9. The models successfully predicted the risk grades for 85 to 95 % of the evaluation period in 2010–2012. They underestimated for 3 to 4 % of the days and overestimated for 2 to 11 % of the days during the evaluation period.

The monthly percentages of agreement for each site and month are shown in Table 10; these values are used to assess model accuracy. The lowest predictive accuracy for oak pollen was found for either in April or May depending on the site, whereas Japanese hop pollen concentration was least accurately predicted in September. The minimum accuracy, 47.3 %, was observed for oak pollen at Daejeon in May. The mean accuracy was in the range of 81 to 98 % by site and 65 to 100 % by month.



◀ **Fig. 5** Predicted and observed daily Japanese hop pollen concentrations during the evaluation period from 2010 to 2012 at the seven validation sites in Korea

Discussion and conclusion

The observed pollen concentrations show nonlinear relationships with meteorological variables. Although, as reviewed by Motulsky and Ransnas (1987), nonlinear regression is regarded as superior to linear regression of transformed data, it is more challenging to fit the nonlinear regression curves with the data and to interpret the resulting parameters. The development of the new models involved data-driven rescaling of predictor variables in a regression; the frequency distributions of pollens based on meteorological parameters were the independent variables, which showed unique distribution patterns with optimal ranges for each weather condition analyzed. The Weibull PDF was used as a transformation function to account for the distribution patterns associated with each meteorological variable; then, multiple regression was performed with the transformed meteorological variables to estimate daily pollen concentrations. Nonlinear transformation of the predictor variables can enhance the predictive ability of linear regression models. For example, a nonlinear transformation was applied for forecasting the daily peak electrical load based on weather conditions (Haida and Muto 1994). The weather–load relationship from the previous year was transformed, translated, and then fitted using the latest daily load and weather data. Although this method yielded better performance than simple linear regression, the study year’s recorded data and the previous year’s weather–load relationship data were needed to develop operational forecasting models. This approach is regarded as impractical for pollen prediction because the critical information of the timing of the pollen’s first appearance of the year cannot be predicted in this way. In

contrast, the novel models developed for this study by design incorporate information from many years of observation, and therefore, yearly updates are not necessary for their application in operational pollen forecasting.

The new models were able to estimate daily risk levels for oak and Japanese hop pollens based on Weibull PDFs that represent the favorability of the weather conditions for pollen concentration. The Weibull PDF was chosen to mimic the effects of biological processes on pollen production as they vary with weather conditions. The influence of biological processes on the production and emission of allergenic pollen is substantial. Although a comprehensive understanding of these processes has not been achieved, data from previous years and for different pollen species have been used in some forecasting models (e.g., Recio et al. 2012; De Weger et al., 2014). These models have had limited applicability, however, because they require either spatial distribution data for the previous years’ observations or real-time observational data, which can be time consuming and laborious to collect.

The entire data set for the seven sites included in the national pollen observational network over the model development period (1997–2009) was used with the Weibull PDF. The multiple regression models, which predict daily pollen concentrations using the Weibull probability densities as independent variables, produced wide variation in the differences between the predicted and observed concentrations by year and by site. In fact, pollen concentrations obtained both from the models and from direct observations varied substantially from site to site and year to year. Therefore, the developed models are forced to predict the mean response of the plants to weather conditions so that they are free from overfitting.

In addition, the patterns in the differences between predicted and observed daily pollen concentrations of oak and Japanese hop (Figs. 4 and 5) are nearly indifferent among the sites and years. This finding suggests that the models can

Table 8 Observed and predicted starts of the pollen seasons for oak and for Japanese hop at the seven validation sites in Korea from 2010 to 2012 based on a start date defined as when yearly cumulative pollen concentration rises to greater than 5 % of the yearly total

Species	Year	Date (month/date)	Busan	Daegu	Daejeon	Gangneung	Gwangju	Jeju	Seoul
Oak	2010	Obs	4/26	4/26	5/1	5/24	4/29	4/26	5/1
		Pred	3/20	3/23	4/2	4/4	3/22	3/14	4/6
	2011	Obs	4/19	5/4	5/3	5/2	4/29	5/3	4/26
		Pred	3/25	4/1	4/4	3/30	3/22	3/10	4/1
	2012	Obs	4/26	4/13	4/18	4/25	4/24	4/27	4/21
		Pred	3/22	3/28	4/4	3/31	3/27	3/12	3/28
J. hop	2010	Obs	8/13	8/28	8/27	9/7	9/5	9/20	8/27
		Pred	8/28	8/18	8/25	8/4	8/19	8/21	8/24
	2011	Obs	9/9	8/25	8/27	8/28	8/25	8/29	8/29
		Pred	8/22	8/18	8/23	8/19	8/23	8/18	8/23
	2012	Obs	8/30	8/31	9/1	8/27	8/11	9/12	8/21
		Pred	8/23	8/21	8/22	8/8	8/11	8/20	8/22

Table 9 Percentage differences between observed risk grades and risk grades predicted with the new models at the seven validation sites in Korea from 2010 to 2012

Difference (Predicted/observed)	Oak	Japanese hop
-3	0.66	0.00
-2	1.80	1.11
-1	1.95	2.05
0	85.01	94.72
1	10.58	2.13

be applied to new sites in Korea without significant bias. Systematic assessment of the spatial applicability of these models, such as through leave-one-out cross-validation, may be important for validation of the models, but this assessment was not explicitly carried out for this study because of the indifferent distribution patterns among the sites and years, as well as the diminished possibility of overfitting.

The reason why there is overestimation early in seasons and underestimation later in seasons is that the Weibull PDF estimates the favorability of each weather element for pollen production and dispersal; the actual pollen released by an individual plant would be suppressed before the completion of flower maturation, then increased with full maturation, and then finally diminished after the peak release. The model's favorability early in the season is therefore overestimated because of lower favorability for pollen production at that stage in the plants' growth. The incorporation of such "negative" effects on pollen production may improve the process of determining daily pollen favorability. Negative impacts may include detrimental factors over previous seasons such as drought in the late fall or low temperatures in the early spring. However, this seasonal variability may require separate modeling with longer and/or denser observational data.

For oak pollen, there may be additional factors, such as differences in species composition and distribution: typical

oak forests include several species that differ in their flowering periods and pollen maturation, but the pollens of which cannot be differentiated under a microscope. Jato et al. (2006) concluded that the number of species included within the same pollen type significantly affected the duration of the pollen season. They also noted that the distribution of oak trees across different altitudes, which is common in Korea, would lead to longer pollen seasons. Data from a broad range of oak species and vegetation altitudes are utilized in the models development so that prediction at an individual site is less dependent upon specific local conditions.

Spatial differences, which are currently not considered in the models, may play an important role in the models' remaining imprecision. Expanding the capability of the current statistical modeling scheme to examine spatial characteristics would require substantial effort; for example, wind speed and direction as well as detailed spatial distribution of vegetation should be considered at each grid point, among other parameters. These spatial considerations are also potential explanations for underestimates of pollen concentration.

The predicted start dates for oak pollen were well before the observed start date, whereas that of Japanese hop pollen was reasonably predicted. The early predicted start date for the oak pollen season can be explained by its broader distribution pattern with lower peak concentrations in the developed model. Early prediction is regarded as preferable to a missed start date for the purpose of allergy preparedness. The deviation in start date prediction was affected by many sources of uncertainty, and it has also been observed in other specialized models (Andersen 1991).

Models that depend solely on weather conditions, such as that of Norris-Hill (1995), are more practical for operational forecasting because they do not require real-time pollen observation. Norris-Hill (1995) subtracted the influences of daily variation associated with weather conditions to estimate daily pollen concentration from a potential pollen concentration determined solely based on the calendar date. The composition

Table 10 Percentage agreement between the daily predicted pollen risks from the new models and the observed pollen risks from 2010 to 2012 at the seven validation sites

Species	Month	Site/accuracy (unit %)							
		Busan	Daegu	Daejeon	Gangneung	Gwangju	Jeju	Seoul	All
Oak	Mar.	100.0	100.0	100.0	100.0	100.0	92.5	100.0	98.9
Oak	Apr.	63.3	74.4	83.3	92.2	72.2	56.7	95.6	76.8
Oak	May	79.6	72.0	47.3	55.9	64.5	76.3	57.0	64.7
Oak	Jun.	100.0	100.0	100.0	98.9	100.0	100.0	100.0	99.8
Mean		85.7	86.6	82.7	86.8	84.2	81.4	88.2	85.1
J. hop	Aug.	98.9	96.8	98.9	100.0	95.6	98.9	96.8	98.0
J. hop	Sep.	74.4	74.4	84.4	93.8	80.0	73.3	86.7	81.0
J. hop	Oct.	98.9	100.0	100.0	100.0	100.0	97.9	100.0	99.5
J. hop	Nov.	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Mean		93.1	92.8	95.8	98.5	93.9	92.5	95.9	94.6

of the new forecast models in this study is largely comparable to his models, but the daily potential pollen concentration is instead determined dynamically based on biological responses of pollen production and transportation to weather conditions.

Using these novel models, operational pollen forecasting by the KMA has been simplified. This methodology has been used to provide information about the distribution of the pollen allergy risks in Korea through the KMA's Digital Weather Forecast since 2014. Because these models incorporate biological responses into pollen production forecasting, it is expected that the risk forecast will grade more reliably across various weather conditions and environments. This system also integrates all information of observed pollen concentrations from the seven cities across the country into one model. It now provides expanded pollen information to improve the quality of life of allergy patients. It is a significant improvement to a nationwide pollen forecasting compared to the previous system, which was only able to provide pollen forecasting for the seven cities included in the pollen monitoring network.

The two-part composition of the new models, i.e., the potential and actual pollen concentrations, has an advantage for application beyond daily operational forecasting: because the models were developed based on analysis of observed pollen and meteorological conditions in Korea without considering vegetational variation, the models can easily be applied with local climate change scenarios to estimate future vulnerability to pollen allergies. However, it must first be noted that the models are fitted based on the present climate; the statistical relationships for present-day conditions may not hold in the future.

The main advantage of the new models is their incorporation of biological processes such as flowering and pollen production. The best available probability functions, which have been developed and integrated in this study, may further be improved by incorporating limits for application. Such limits may include regional and cultivar differences in the areas of vegetation, flower production, and overwintering quality.

Acknowledgments This research was supported by the "Research and Development for KMA Weather, Climate, and Earth system Services: Biometeorology" of the National Institute of Meteorological Sciences (NIMS) of the Korea Meteorological Administration (KMA). The authors appreciate the thoughtful comments and suggestions of the two anonymous reviewers that improved the quality of the paper significantly.

References

- Andersen TB (1991) A model to predict the beginning of the pollen season. *Grana* 30:269–275
- Bousquet J, Khaltaev N, Cruz AA, et al. (2008) Allergic rhinitis and its impact on asthma (ARIA) 2008. *Allergy* 63:8–160
- D'Amato G, Holgate ST, Pawankar R, et al. (2015) Meteorological conditions, climate change, new emerging factors, and asthma and related allergic disorders. A statement of the World Allergy Organization. *World Allergy Organ J* 8:25
- De Weger LA, Beerthuizen T, Hiemstra PS, Sont JK (2014) Development and validation of a 5-day-ahead hay fever forecast for patients with grass-pollen-induced allergic rhinitis. *Int J Biometeorol* 58:1047–1055
- Driessen MNBM, Van Herpen RMA, Moelands RPM, Spijksma FTM (1989) Prediction of the start of the grass pollen season for the western part of the Netherlands. *Grana* 28:37–44
- Emberlin J, Mullins J, Corden J, Jones S, Millington W, Brook M, Savage M (1999) Regional variations in grass pollen seasons in the UK, long-term trends and forecast models. *Clin Exp Allergy* 29:347–356
- Galan C, Emberlin J, Dominguez E, Bryant RH, Villamandos F (1995) A comparative analysis of daily variations in the Gramineae pollen counts at Cordoba, Spain and London, UK. *Grana* 34:189–198
- Google Earth (2016) Republic of Korea 35°36'42.25"N, 127°29'47.81"E, elevation 393M. <<http://www.google.com/earth/>> [Viewed 26 April 2016]
- Haida T, Muto S (1994) Regression based peak load forecasting using a transformation technique. *IEEE T Power Syst* 9:1788–1794
- Han MJ, Han DH, JW O, Lee HB, Kim SW, Kook MH, Park KS, Kim BS, Kim KR, Choi YJ (2012) The relation between pollen concentration and symptom index of seasonal allergic rhinitis for risk grade for pollen allergy in Korea. *Jpn J Polynol* 58:76–77
- Hong SJ, Lee MS, Sohn MH, Shim JY, Han YS, Park KS, Ahn YM, Son BK, Lee HB (2004) Self-reported prevalence and risk factors of asthma among Korean adolescents: 5-year follow-up study, 1995–2000. *Clin Exp Allergy* 34:1556–1562
- Hong SJ, Ahn KM, Lee SY, Kim KE (2008) The prevalence of asthma and allergic diseases in Korean children. *Pediatr Allergy Respir Dis* 18:15–25 [in Korean with English abstract]
- Hwang SH, Jung SY, Lim DH, Son BK, Kim JH, Yang JM, IB O, Kim Y, Lee JH, Lee KH, Kim SY, Hong SC, Lee HS (2013) Epidemiology of allergic rhinitis in Korean children. *Allergy Asthma Respir Dis* 1: 321–332 [in Korean with English abstract]
- Jato V, Rodriguez-Rajo FJ, Alcazar P, De Nuntis P, Galan C, Mandrioli P (2006) May the definition of pollen season influence aerobiological results? *Aerobiologia* 22:13–25
- Jee HM, Kim KW, Kim CS, Sohn MH, Shin DC, Kim KE (2009) Prevalence of asthma, rhinitis and eczema in Korean children using the International Study of Asthma and Allergies in Childhood (ISAAC) questionnaires. *Pediatr Allergy Respir Dis* 19:165–172 [in Korean with English abstract]
- Khwahram N, Dash J, Atkinson PM, Newnham RM, Skjoth CA, Adams-Groom B, Caulton E, Head K (2014) Exploring the spatio-temporal relationship between two key aeroallergens and meteorological variables in the United Kingdom. *Int J Biometeorol* 58:529–545
- Kim HY, Kwon EB, Baek JH, Shin YH, Yum HY, Jee HM, Yoon JW, Han MY (2013) Prevalence and comorbidity of allergic diseases in preschool children. *Korean J Pediatr* 56:338–342
- Kim JH, Oh JW, Lee HB, Kim SW, Chung HL, Kook MH, Park KS, Kim BS, Kim JK, Lee DJ, Paik WK, Kim KR, Lee HL, Choi YJ, Yu SD, Kim JH, Cho YS (2014) Evaluation of the association of vegetation of allergenic plants and pollinosis with meteorological changes. *Allergy Asthma Respir Dis* 2:48–58 [in Korean with English abstract]
- Kim KR, Oh JW, Lee JW (2009) Characteristics of pollen concentration in Korea and survey of its health impact. Final report for the project of National Institute of Environmental Research [in Korean]
- Kim KR, Park KJ, Lee HR, Kim M, Choi YJ, Oh JW (2012) Development and evaluation of the forecast models for daily pollen allergy. *Korean J Agr and Forest Meteorol* 14:265–268 [in Korean with English abstract]
- Larsson KA (1993) Prediction of the pollen season with a cumulated activity method. *Grana* 32:111–114
- Lee HR, Kim KR, Choi YJ, Oh JW (2012) Meteorological impact on daily concentration of pollen in Korea. *Korean J Agr and Forest Meteorol* 14:99–107 (in Korean with English abstract)

- Motulsky HJ, Ransnas LA (1987) Fitting curves to data using nonlinear regression: a practical and nonmathematical review. *FASEB J* 1: 365–374
- Norris-Hill J (1995) The modeling of daily Poaceae pollen concentrations. *Grana* 34:182–188
- Oh JW, Kang IJ, Kim SW, Kook MH, Kim BS, Shin KS, Hahn YS, Lee HB, Shon MH, Cheong JT, Lee HR, Kim KE (2006) The correlation between increased sensitization rate to weeds in children and the annual increase in weed pollen in Korea. *Pediatr Allergy Respir Dis* 16:114–121 [in Korean with English abstract]
- Oh JW, Lee HB, Kang IJ, Kim SW, Park KS, Kook MH, Kim BS, Baek HS, Kim JH, Kim JK, Lee DJ, Kim KR, Choi YJ (2012) The revised edition of Korean calendar for allergenic pollen. *Allergy, Asthma Immunol Res* 4:5–11
- Oh YC, Kim HA, Cheong JT, Kang IJ, Kim SW, Kim BS, Kook MH, Lee HB, Oh JW (2009) Evaluation of the relationship between pollen count and the outbreak of allergic diseases. *Allergy Asthma Respir Dis* 19:354–364 [in Korean with English abstract]
- Okamoto Y, Horiguchi S, Yamamoto H, Yonekura S, Hanazawa T (2009) Present situation of cedar pollinosis in Japan and its immune responses. *Allergol Int* 58:155–162
- Park HJ, Lim HS, Park KH, Lee JH, Park JW, Hong CS (2014) Changes in allergen sensitization over the last 30 years in Korea respiratory allergic patients: a single-center. *Allergy, Asthma Immunol Res* 6: 434–443
- Park KJ, Kim HA, Kim KROJW, Lee SY, Choi YJ (2008) Characteristics of regional distribution of pollen concentration in Korean Peninsula. *Korean J Agr Forest Meteorol* 10:167–176 [in Korean with English abstract]
- Recio M, Trigo MM, Docampo S, Melgar M, Garcia-Sanchez J, Bootello L, Cabezudo B (2012) Analysis of the predicting variables for daily and weekly fluctuations of two airborne fungal spores: *Alternaria* and *Cladosporium*. *Int J Biometeorol* 56:983–991
- Sabariego S, Cuesta P, Fernandez-Gonzalez F, Perez-Badia R (2012) Models for forecasting airborne Cupressaceae pollen levels in Central Spain. *Int J Biometeorol* 56:253–258
- Subiza J, Masiello JM, Subiza JL, Jerez M, Hinojasa M, Subiza E (1992) Prediction of annual variations in atmospheric concentrations of grass pollen. A method based on meteorological factors and grain crop estimates *Clin Exp Allergy* 22:540–546
- Toro FJ, Recio M, Trigo MM, Cabezudo B (1998) Predictive models in aerobiology: data transformation. *Aerobiologia* 14:179–184