

Matt Smith · Jean Emberlin

A 30-day-ahead forecast model for grass pollen in north London, United Kingdom

Received: 22 February 2005 / Revised: 4 October 2005 / Accepted: 6 October 2005
© ISB 2005

Abstract A 30-day-ahead forecast method has been developed for grass pollen in north London. The total period of the grass pollen season is covered by eight multiple regression models, each covering a 10-day period running consecutively from 21 May to 8 August. This means that three models were used for each 30-day forecast. The forecast models were produced using grass pollen and environmental data from 1961 to 1999 and tested on data from 2000 and 2002. Model accuracy was judged in two ways: the number of times the forecast model was able to successfully predict the severity (relative to the 1961–1999 dataset as a whole) of grass pollen counts in each of the eight forecast periods on a scale of 1 to 4; the number of times the forecast model was able to predict whether grass pollen counts were higher or lower than the mean. The models achieved 62.5% accuracy in both assessment years when predicting the relative severity of grass pollen counts on a scale of 1 to 4, which equates to six of the eight 10-day periods being forecast correctly. The models attained 87.5% and 100% accuracy in 2000 and 2002, respectively, when predicting whether grass pollen counts would be higher or lower than the mean. Attempting to predict pollen counts during distinct 10-day periods throughout the grass pollen season is a novel approach. The models also employed original methodology in the use of winter averages of the North Atlantic Oscillation to forecast 10-day means of allergenic pollen counts.

Keywords Aerobiology · Grass pollen counts · London · Forecast models · North Atlantic Oscillation

Introduction

Aerobiologists have often produced forecast models for allergenic pollen. The output of such models has either been supplied to allergy sufferers directly so that they can plan their medication and activities in advance, or to

medical professionals who plan treatment and schedule clinical trials. Forecasts are also of value to those who produce and stock health care products (Emberlin et al. 1999).

Short-term forecasts have been used to predict pollen concentrations for the next day or next few days, and concentrate mainly on those factors that affect the release and dispersal of pollen from the plant. Long-term forecasts, on the other hand, are used to calculate the principal characteristics of the main pollen season, such as the start date or severity and therefore focus on those factors that affect pollen production (Larsson 1993). The forecast model presented in this paper is novel because it incorporates factors that affect both pollen production and pollen release to predict grass pollen counts during distinct 10-day periods 1 month ahead throughout the grass pollen season.

The overall prevalence of hay fever in Europe is approximately 15–20% (Aas et al. 1997; Huynen et al. 2003). Grass pollen is considered to be one of the most important aeroallergens (Sanchez Mesa et al. 2003) affecting up to 90% of hay fever sufferers (Emberlin 1997). The highest prevalence occurs in late adolescence/early adulthood, with between 8% and 35% of young adults in EEC countries having IgE serum antibodies to grass pollen (Burr 1999; D'Amato 2000). An accurate month-ahead forecast would be of notable economic and social benefit as well as being a considerable asset to health care professionals.

Methods

Pollen monitoring sites

Daily average grass pollen counts were obtained from three pollen-monitoring sites in north London that have operated consecutively from 1961 to 2002. These sites were: St Mary's Hospital, Paddington (1961–1989); the University of North London, Holloway Road (1990–1994); and the Environmental and Public Protection offices, Islington (1995–2002). The pollen traps were positioned at a height (above ground level) of 29 m, 18 m and 15 m, respectively.

M. Smith · J. Emberlin (✉)
National Pollen and Aerobiology Research Unit,
University of Worcester, UK
e-mail: j.emberlin@worc.ac.uk
Tel.: +44-1905-855255

The furthest distance between sites was approximately 8 km. The sites are situated about 3.3 km to the north of the River Thames (Fig. 1).

The topography of the region is dominated by a geological structure known as the London Basin. The Marlborough and Berkshire Downs (approximately 115 km and 93 km to the east, respectively), the Chiltern Hills (approximately 47 km northeast) and the North Downs (approximately 30 km south) are part of an extensive chalk outcrop that forms the rim of the basin. The land slopes downward from the rim of the basin to the River Thames, which runs through its heart. Altitudes range from sea level to 245 m a.s.l. on the North Downs. Nearer to the pollen monitoring sites, this gradual descent toward the river is interrupted by a number of gently rising hills, such as Hampstead Heath (110 m a.s.l.) approximately 4.5 km to the northwest, and Primrose Hill (60 m a.s.l.) located about half-way between Paddington and Islington (all distances calculated from Environmental and Public Protection offices, Islington). This area of north London is mainly urban/suburban, with a mix of residential, commercial, retail and institution uses. There is a small amount of semi-natural vegetation, including grassland and secondary deciduous woodland (EA 1999).

18°C, respectively, and mean annual rainfall is about 584 mm (1971–2000 average) (METO 2004).

London has a complex urban climate that has an effect on pollen dispersion, transport and deposition. The presence of high buildings and intricate surfaces within cities increase turbulence, thereby causing pollen counts to differ considerably over short distances both vertically and horizontally (Emberlin and Norris-Hill 1991; Goudie 1994; Jones 1995; Emberlin 2000). In addition, cities create a microclimatic change in temperature known as an urban heat island. These thermal increases can influence plant phenology by causing plants to flower earlier in urban areas (Jones 1995; Goudie 1996). Furthermore, urban heat islands promote turbulence and the upward movement of air. Urban heat islands can also produce a local circulation created by the difference in temperature between cities and the surrounding areas, which is a generating mechanism similar to that of the sea-breeze circulation seen in coastal areas during the warmest months. The re-circulation of airborne pollen is made possible by means of this heat island breeze system, which can bring a regional component (pollen from 2 km to 200 km) to the pollen spectrum of urban areas (Gassman et al. 2002; Ohashi and Kida 2002).

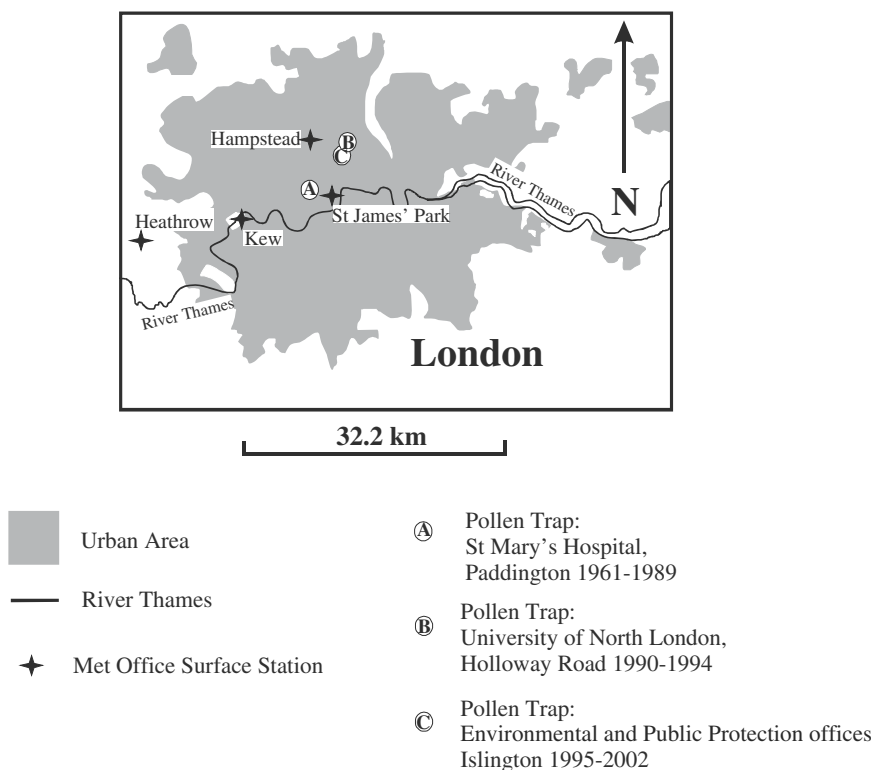
Climate

The British Isles has a temperate maritime climate that is influenced throughout the year by areas of low pressure moving in from the Atlantic. In London, mean temperatures for January and July are approximately 5.5°C and

Grass pollen data

The methodology used for collecting the grass pollen data followed the standard method of the United Kingdom National Pollen Monitoring Network described in the British Aerobiology Federation (BAF) guide for trapping

Fig. 1 Pollen monitoring sites and Met Office Surface Stations used



and counting airborne pollen and spores (BAF 1995). The three grass pollen datasets used in this investigation were spliced together to make a single dataset running from 1961 to 2002. The years 1991 and 2001 were omitted from the study because they contained insufficient data for analysis due to trap failure. The results of a Kruskal-Wallis non-parametric test showed that there was no significant difference in the amount of grass pollen recorded annually at the three sites and the datasets could therefore be treated as one population. Normality tests were performed on the grass pollen data using the Explore function in SPSS. Datasets that were not normally distributed were square rooted (normalised) (Jones 1995; Pallant 2001).

Grass pollen season statistics

Emberlin et al. (1993) showed that there was a marked decrease in the severity of the grass pollen seasons in north London from the early 1960s to the early 1970s, with a continued decline through the mid-1980s when the trend levelled off. This decline was explained by a number of factors, including the general reduction in grassland area, the preference of silage over hay production and the agricultural improvement of grasslands that resulted in a reduction in species diversity and an increase in the proportion of so-called preferred species in the sward, especially *Lolium perenne* (Fuller 1987; Emberlin et al. 1993; Hopkins and Davies 1994). Moreover, pollen concentrations typically decrease with increased height above ground level (Hart et al. 1994). The pollen trap in north London has been relocated three times over the last 40 years and has been lowered each time it was moved, which suggests that the trend toward less severe grass pollen seasons may be even more pronounced.

There has been a tendency toward earlier and more severe grass pollen seasons in north London in recent years. However, it is not known whether this is a short-term increase or a long-term trend. The forecast models were produced using grass pollen data from 1961 to 1999 and tested on data from 2000 and 2002. The amount of grass pollen recorded in 2000 was lower than the 1961–1999 mean, whereas 2002 was an above average year.

Forecasting methods

The 30-day-ahead forecast for grass pollen in north London employed eight 10-day models that ran consecutively from 21 May to 8 August. These dates allow for changes in plant phenology (early or later flowering times) that have been related to global warming (Menzel and Fabian 1999; Ahas et al. 2002; Frenguelli 2002; Van Vliet et al. 2002; Parmesan and Yohe 2003; Root et al. 2003). The use of 10-day periods or decades-of-days is a standard method employed in meteorological and aerobiological work (Jones 1995; Spieksma et al. 1995; Recio et al. 1998; Spieksma and Nikkels 1998; Emberlin et al. 1999; Adams-Groom et al. 2002).

The method of forecasting used throughout this investigation was standard multiple regression analysis, which is

an empirical technique often used in aerobiological studies. Dependent variables (normalised 10-day mean grass pollen counts) were regressed against a number of independent variables made up of environmental data (meteorological and grass pollen data). Each of the (eight) multiple regression models included environmental data from the months preceding the forecast period so that factors affecting pollen production were integrated into the model. Each model also included meteorological data taken from a corresponding period to the dependent variable, in order to incorporate factors that influence the release and dispersal of grass pollen from the plant. The strict adherence to this rule often complicated the forecasting process, especially when no significant relationships could be found. When this occurred, less significant variables were entered into the analysis.

The problem of presenting the data in such a way that it was straightforward to understand was considered. It was decided to construct forecasts defined by date (for example, 10–19 June) rather than a period defined numerically (such as days 161–170 from 1 January), because it was thought that this would be easier for people to relate to. However, it was not possible to use this method when dealing with meteorological data taken from the spring (1 January to the end of April) because of the inclusion of leap years. Spring data were therefore defined numerically, as days 1–120 from 1 January.

Every 30-day-ahead forecast contained three 10-day models. For example, a forecast for 20 June to 19 July would include separate regression models for: 20–29 June, 30 June–9 July and 10–19 July. Environmental data from the 20 days immediately preceding each model was excluded from the forecast. This meant that just one regression model was needed for each 10-day period, thereby simplifying the modelling process. Unfortunately, this also meant that data that might have increased model accuracy was left out of the forecast.

It should, however, be noted that this does not include meteorological data taken from the same period as the dependent variable, which were used to represent factors affecting the release and dispersal of pollen from the plant. When the 30-day-ahead forecast is produced in real time, these variables will be in the form of predicted values taken from weather forecasts.

The multiple regression models were constructed using grass pollen and environmental data from 1961 to 1999. Predicted values were tested against 10-day mean normalised grass pollen counts from the 2000 and 2002 grass

Table 1 Categories used to test the long-range forecast model for grass pollen in north London

Pollen data category: 25th and 75th percentiles and mean calculated using grass pollen data from 1961 to 1999	Relative severity	Numerical score
≤25th percentile	Very low	1
>25th percentile and <mean	Low	2
>Mean and <75th percentile	High	3
≥75th percentile	Very high	4

Table 2 Significant correlations between 10-day mean grass pollen counts and North Atlantic Oscillation (NAO) winter averages

10-day mean grass pollen count	NAO (winter averages)	Correlation coefficient
21–30 May	February–March average	0.368*
20–29 June	December–January average	−0.395*

*Correlation significant at the 0.05 level (2-tailed)

pollen seasons. Pollen data from 1961 to 1999 were categorised (Table 1) and given scores ranging from 1 to 4 in order to determine the severity (relative to the 1961–1999 dataset as a whole) of grass pollen counts in each of the eight forecast periods. Model accuracy was judged in two ways: the number of times the forecast model was able to successfully predict actual grass pollen counts on a scale of 1 to 4 (in relation to the 1961–1999 dataset); and the number of times the forecast model was able to predict whether grass pollen counts were higher or lower than the (1961–1999) mean.

Independent variables entered into the analysis

The Met Office, via the British Atmospheric Data Centre (BADC), supplied the meteorological data used to construct the forecast models. Meteorological data were taken from a number of different sites, situated in and around the city, in order to take into account spatial and temporal differences in the production, release and dispersal of grass pollen (Fig. 1). The variables considered were: maximum and minimum daily temperatures recorded at Heathrow and St. James' Park; and daily rainfall recorded at Hampstead, Heathrow, and Kew Royal Botanical Gardens.

Because of the problems associated with multicollinearity, it was decided not to enter more than one temperature or rainfall variable from the same 10-day period into the regression models simultaneously. Maximum and minimum daily temperatures are highly correlated and therefore unsuitable to be entered into regression analysis together.

The independent variables representing factors that affect pollen production were 10-day mean maximum and minimum temperatures and rainfall during spring (from 1 January to 21 May). Also entered into this analysis, and related to

Table 3 Significant positive correlations between 10-day mean grass pollen counts and 10-day mean meteorological data

10-day mean grass pollen count	10-day mean meteorological data	Correlation coefficient
21–30 May	St James' Park minimum temperature days 31–40 from 1 Jan	0.349*
	Heathrow minimum temperature days 51–60 from 1 Jan	0.377*
	St James' Park maximum temperature days 51–60 from 1 Jan	0.473**
	St James' Park minimum temperature days 51–60 from 1 Jan	0.537**
	St James' Park maximum temperature days 71–80 from 1 Jan	0.371*
	Hampstead rain days 101–110 from 1 Jan	0.307*
31 May–9 June	St James' Park minimum temperature days 111–120 from 1 Jan	0.349*
	Heathrow minimum temperature days 41–50 from 1 Jan	0.390*
	St James' Park minimum temperature days 41–50 from 1 Jan	0.332*
	Heathrow maximum temperature days 51–60 from 1 Jan	0.467**
	Heathrow minimum temperature days 51–60 from 1 Jan	0.512**
	St James' Park maximum temperature days 51–60 from 1 Jan	0.474**
10–19 June	St James' Park minimum temperature days 51–60 from 1 Jan	0.472**
	None	
	None	
	None	
	None	
	None	
20–29 June	None	
	None	
	None	
30 June–9 July	None	
	None	
	None	
10–19 July	Heathrow rain days 41–50 from 1 Jan	0.388*
	Kew rain days 41–50 from 1 Jan	0.354*
	Hampstead rain days 41–50 from 1 Jan	0.368*
20–29 July	St James' Park minimum temperature days 41–50 from 1 Jan	0.345*
	Hampstead rain days 1–10 from 1 Jan	0.358*
	Heathrow maximum temperature days 11–20 from 1 Jan	0.408*
30 July–8 August	Heathrow maximum temperature days 41–50 from 1 Jan	0.391*
	St James' Park minimum temperature days 41–50 from 1 Jan	0.354*
	Heathrow maximum temperature days 71–80 from 1 Jan	0.365*
	St James' Park minimum temperature days 71–80 from 1 Jan	0.347*
	Heathrow minimum temperature days 111–120 from 1 Jan	0.397*
	St James' Park minimum temperature days 111–120 from 1 Jan	0.457*

*Correlation significant at the 0.05 level (2-tailed)

**Correlation significant at the 0.01 level (2-tailed)

spring temperatures and rainfall, were winter averages of the North Atlantic Oscillation (NAO). The NAO is a mode of interannual variability in atmospheric circulation that is associated with changes in the surface westerlies across the North Atlantic and into Europe (Hurrell 1995). There is a well-documented relationship between the NAO and terrestrial ecosystems (Ottersen et al. 2001; D'Odorico et al. 2002; Stenseth et al. 2002; Walther et al. 2002) and D'Odorico et al. (2002) suggested using phases of the NAO to predict the onset of pollination. Data for the NAO were obtained from the NAO index, calculated from Gibraltar and southwest Iceland (Reykjavik), which is maintained by P. D. Jones at the University of East Anglia (D'Odorico et al. 2002; Osborn and Climatic Research Unit 2002).

Meteorological data taken from a period corresponding to that of the dependent variable were entered into the models in order to represent pollen release; these were 10-day means of maximum and minimum daily temperatures and daily rainfall from 21 May to 8 August. The influence

of grass pollen counts at the beginning of the season on grass pollen availability later in the year was explored by including past pollen counts (cumulative grass pollen counts from 21 May and 10-day means of daily grass pollen counts from 21 May to 9 July) in the analysis.

Results

Correlation analysis

The following environmental variables were entered into correlation analysis with 10-day mean normalised grass pollen counts from 21 May to 8 August (Tables 2, 3, 4, 5 and 6):

- NAO (winter averages)
- 10-day means of meteorological data (maximum and minimum daily temperatures and daily rainfall) from day 1 to day 120 from 1 January

Table 4 Significant negative correlations between 10-day mean grass pollen counts and 10-day mean meteorological data

10-day mean grass pollen count	10-day mean meteorological data	Correlation coefficient
21–30 May	Hampstead rain days 81–90 from 1 Jan	–0.362*
31 May–9 June	Heathrow rain days 81–90 from 1 Jan	–0.494**
	Kew rain days 81–90 from 1 Jan	–0.485**
	Hampstead rain days 81–90 from 1 Jan	–0.530**
	None	
10–19 June		
20–29 June	Heathrow maximum temperature days 61–70 from 1 Jan	–0.368*
	Heathrow minimum temperature days 61–70 from 1 Jan	–0.500**
	St James' Park maximum temperature days 61–70 from 1 Jan	–0.355*
	St James' Park minimum temperature days 61–70 from 1 Jan	–0.441**
	Hampstead rain days 61–70 from 1 Jan	–0.344*
	Heathrow maximum temperature days 81–90 from 1 Jan	–0.366*
	Heathrow minimum temperature days 81–90 from 1 Jan	–0.365*
	St James' Park maximum temperature days 81–90 from 1 Jan	–0.324*
	St James' Park minimum temperature days 81–90 from 1 Jan	–0.418**
	Heathrow maximum temperature days 91–100 from 1 Jan	–0.388*
	St James' Park maximum temperature days 91–100 from 1 Jan	–0.439**
	Heathrow minimum temperature days 1–10 from 1 Jan	–0.330*
	Heathrow minimum temperature days 71–80 from 1 Jan	–0.467**
	St James' Park minimum temperature days 71–80 from 1 Jan	–0.385*
	St James' Park minimum temperature days 81–90 from 1 Jan	–0.375*
	Heathrow minimum temperature days 101–110 from 1 Jan	–0.389*
	St James' Park minimum temperature days 101–110 from 1 Jan	–0.365*
	Heathrow minimum temperature days 21–30 May	–0.329*
10–19 July	Heathrow maximum temperature days 10–19 June	–0.370*
	Heathrow minimum temperature days 10–19 June	–0.413**
	St James' Park minimum temperature 10–19 June	–0.464**
20–29 July	Heathrow maximum temperature days 1–10 May	–0.377*
	Heathrow maximum temperature days 20–29 June	–0.355*
	Heathrow minimum temperature days 20–29 June	–0.543**
	St James' Park maximum temperature 20–29 June	–0.412*
	St James' Park minimum temperature 20–29 June	–0.592**
30 July–8 August	Heathrow rain days 51–60 from 1 Jan	–0.357*

*Correlation significant at the 0.05 level (2-tailed)

**Correlation significant at the 0.01 level (2-tailed)

Table 5 Significant correlations between 10-day mean grass pollen count and earlier grass pollen counts in the same season

10-day mean grass pollen count	Previous grass pollen counts	Correlation coefficient
30 June–9 July	Cumulative grass pollen from 21 May to 30 May	0.420*
9–18 August	10-day mean grass pollen count from 21–30 May	0.481**

*Correlation significant at the 0.05 level (2-tailed)

**Correlation significant at the 0.01 level (2-tailed)

- 10-day means of meteorological data (maximum and minimum daily temperatures and daily rainfall) from 1 May to 8 August
- 10-day means of daily grass pollen counts, from 21 May to 9 July
- Cumulative grass pollen counts from 21 May

A total of 191 variables were entered into the analysis. The results presented in Tables 2, 3, 4, 5 and 6 are therefore limited to those correlations that were significant ($P < 0.05$).

Regression analysis

The results of correlation analysis were used to help select variables for entry into the regression models (Tables 7, 8). It should be noted, however, that the likelihood is that a number of significant relationships were spurious, and came about by chance. For that reason, informed discretion was applied in the selection of variables for entry into regression analysis, and factors that are known to influence pollen production, release and dispersal were considered.

Testing the 30-day-ahead forecast model for north London

The construction of the long-range forecast model involved a certain amount of trial and error before the right combinations of variables were selected for the final model. It was frequently found that the most significant variables were not necessarily the best at forecasting grass pollen counts. As a result, it was often necessary to enter less significant variables into regression analysis instead.

Table 6 Significant correlations between 10-day mean grass pollen counts and corresponding meteorological data

Period	10-day mean meteorological data	Correlation coefficient
21–30 May	Heathrow maximum temperature 21–30 May	0.522**
	St James' Park maximum temperature 21–30 May	0.502**
	St James' Park minimum temperature 21–30 May	0.435*
31 May–9 June	Heathrow maximum temperature 31 May–9 June	0.518**
	Heathrow minimum temperature 31 May–9 June	0.573**
	St James' Park maximum temperature 31 May–9 June	0.577**
	St James' Park minimum temperature 31 May–9 June	0.488**
	Heathrow maximum temperature 10–19 June	0.485**
10–19 June	Heathrow minimum temperature 10–19 June	0.365**
	St James' Park maximum temperature 10–19 June	0.456**
	St James' Park minimum temperature 10–19 June	0.415**
	Heathrow rain 20–29 June	–0.456**
20–29 June	Kew rain 20–29 June	–0.388*
	Hampstead rain 20–29 June	–0.379*
	Heathrow maximum temperature 30 June–9 July	0.394*
30 June–9 July	Heathrow rain 30 June–9 July	–0.429*
	St James' Park maximum temperature 30 June–9 July	0.446**
	Hampstead rain 30 June–9 July	–0.448*
	Heathrow maximum temperature 10–19 July	0.343*
10–19 July	Heathrow rain 10–19 July	–0.445**
	St James' Park maximum temperature 10–19 July	0.389*
	Kew rain 10–19 July	–0.448**
	Hampstead rain 10–19 July	–0.438*
20–29 July	None	
30 July–8 August	Heathrow maximum temperature 30 July–8 August	0.432*
	Heathrow minimum temperature 30 July–8 August	0.533**
	St James' Park maximum temperature 30 July–8 August	0.464**
	St James' Park minimum temperature 30 July–8 August	0.520**

*Correlation significant at the 0.05 level (2-tailed)

**Correlation significant at the 0.01 level (2-tailed)

Table 7 Results of multiple regression analysis between 10-day mean grass pollen counts and various environmental variables

Regression model	Prediction period	Regression equation	Adjusted R^2
1	21–30 May	$(0.162 \times \text{HEATMAXT } 21\text{--}30 \text{ May}^*) + (0.236 \times \text{SJPMINT } 51\text{--}60^*) - 2.498$	0.375
2	31 May–9 June	$(0.736 \times \text{HEATMINT } 31 \text{ May--}9 \text{ June}^*) + (0.279 \times \text{HEATMINT } 51\text{--}60^*) - (0.868 \times \text{HAMPRAIN } 81\text{--}90^*) - 3.306$	0.629
3	10–19 June	$(0.666 \times \text{HEATMAXT } 10\text{--}19 \text{ June}^*) - (0.582 \times \text{HAMPRAIN } 61\text{--}70) - 4.816$	0.271
4	20–29 June	$8.770 - (0.422 \times \text{HAMPRAIN } 20\text{--}29 \text{ June}^*) - (0.525 \times \text{NAO FM}^*)$	0.229
5	30 June–9 July	$(0.445 \times \text{SJPMAXT } 30 \text{ June--}9 \text{ July}^*) + (0.02347 \times \text{CGPC } 21\text{--}30 \text{ May}^*) - 3.980$	0.325
6	10–19 July	$5.666 - (0.805 \times \text{KEWRain } 10\text{--}19 \text{ July}^*) - (0.450 \times \text{NAODJF}^*) + (0.571 \times \text{HAMPRAIN } 41\text{--}50^*)$	0.333
7	20–29 July	$7.634 + (0.111 \times \text{KEWRain } 20\text{--}29 \text{ July}) - (0.314 \times \text{HEATMAXT } 1\text{--}10 \text{ May}^*) + (0.303 \times \text{SJPMINT } 41\text{--}50^*)$	0.289
8	30 July–8 August	$(0.122 \times \text{HEATMINT } 30 \text{ July--}8 \text{ August}^*) + (0.262 \times \text{GRASSPOLL } 21\text{--}30 \text{ May}^*) + (0.07002 \times \text{HEATMAXT } 11\text{--}20^*) - 1.723$	0.493

*Variable significant at the 95% level when included in the regression model

The 30-day-ahead forecast models were tested by comparing predicted values against 10-day mean normalised grass pollen counts from the 2000 and 2002 grass pollen seasons (Figs. 2, 3). The forecast models successfully predicted the relative severity (relative to the 1961–1999 dataset) of grass pollen counts on a scale of 1 to 4 (Table 1) in six of the eight 10-day periods in both assessment years (62.5% accuracy). When assessed on their ability to predict whether grass pollen counts were higher or lower than the 1961–1999 mean, the models attained 87.5% and 100% accuracy in 2000 and 2002, respectively.

In order to take into account variations in plant phenology, the models included days not usually considered to be in the peak of the grass pollen season (in late May and early August). Therefore, forecast models for the peripheries of the grass pollen season proved rather difficult to

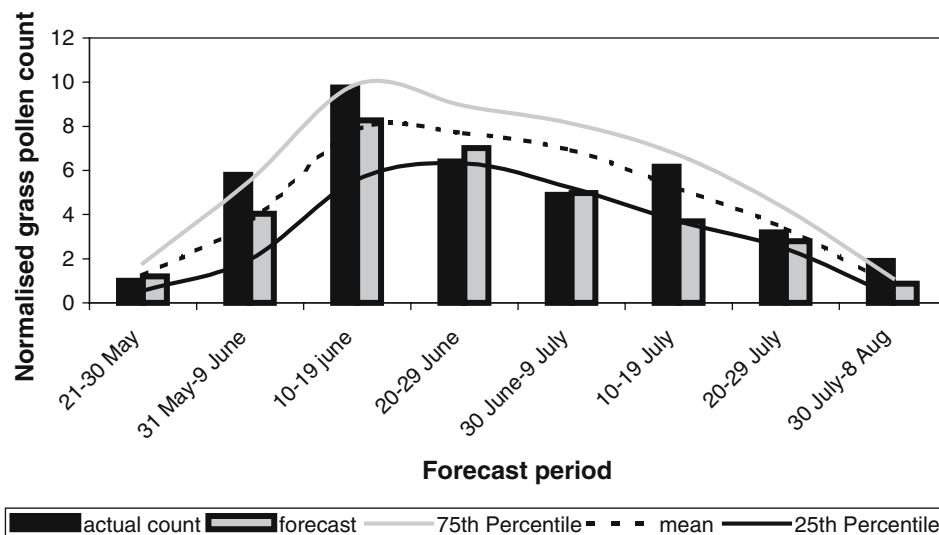
produce. For example, the forecast model for the period 21–30 May successfully predicted the relative severity of the grass pollen count (on a scale of 1 to 4) in 2000, but only managed to determine whether pollen counts were higher or lower than the mean in 2002. Conversely, the model for the period 30 July–8 August predicted whether counts would be higher or lower than average in 2000, and correctly forecast the relative severity of the grass pollen count in 2002.

Temporal variations in the peak of the grass pollen season also affected the performance of the models. The peak of the grass pollen usually occurs, based on the 1961–1999 mean, around the 10–19 June (Figs. 2, 3). However, the peak occurred later than usual in 2002, which caused the forecast model for the 20–29 June to under-predict the amount of pollen recorded.

Table 8 Independent variables entered into the eight regression models

Regression model	Abbreviation	Variables
1	HEATMAXT 21–30 May SJPMINT 51–60	Heathrow maximum temperature 21–30 May St James' Park minimum temperature days 51–60 from 1 Jan
2	HEATMINT 31 May–9 June HEATMINT 51–60 HAMPRAIN 81–90	Heathrow minimum temperature 31 May–9 June Heathrow minimum temperature days 51–60 from 1 Jan Hampstead rainfall days 81–90 from 1 Jan
3	HEATMAXT 10–19 June HAMPRAIN 61–70	Heathrow maximum temperature 10–19 June Hampstead rain days 61–70 from 1 Jan
4	HAMPRAIN 20–29 June NAO FM	Hampstead rain 20–29 June NAO February–March average
5	SJPMAXT 30 June–9 July CGPC 21–30 May	St James' Park maximum temperature 30 June–9 July Cumulative grass pollen count 21–30 May
6	KEWRain 10–19 July NAODJF HAMPRAIN 41–50	Kew rainfall 10–19 July NAO December–January–February average Hampstead rain days 41–50 from 1 Jan
7	KEWRain 20–29 July HEATMAXT 1–10 May SJPMINT 41–50	Kew Botanical Gardens rainfall 20–29 July Heathrow maximum temperature 1–10 May St. James' Park minimum temperature days 41–50 from 1 Jan
8	HEATMINT 30 July–8 August GRASSPOLL 21–30 May HEATMAXT 11–20	Heathrow minimum temperature 30 July–8 August 10-day mean daily grass pollen count 21–30 May Heathrow maximum temperature days 11–20 from 1 Jan

Fig. 2 Actual and predicted grass pollen counts in north London for 2000. Eight forecast periods from 21 May to 8 August. 25th percentile, mean and 75th percentile of 1961–1999 data set plotted for comparison



The forecast for the period 10–19 July was the most difficult to produce. This may also be related to the timing of peak pollen concentrations. Norris-Hill (1995) noted that, in north London, there was a second major peak in grass pollen counts around 10 July, which was probably caused by the simultaneous flowering of several common grass species. This second peak does not always occur. Consequently, the forecast model for this period was only able to predict whether counts would be higher or lower than the mean in one assessment year (2002).

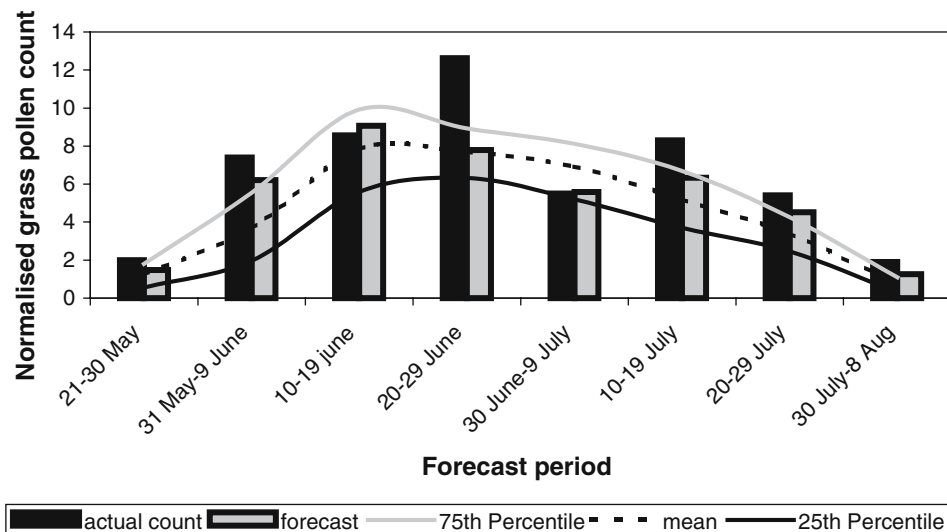
Discussion

Correlation analysis is an extremely useful tool but the forecaster must use an understanding of pollen production, release and dispersal to decide which variables to enter into the forecast model. For instance, each of the (eight) regression models included independent variables from the same period as the dependent variable in order to introduce pollen release and dispersal into the forecast. There were no

notable correlations between the dependent variable and corresponding meteorological data for the period 20–29 July (Table 6). The variable Kew Botanical Gardens rainfall from 20–29th July was chosen for the model because it produced the most satisfactory forecast even though correlation analysis showed only a poor relationship between it and the dependent variable ($R=0.077$). It should, however, be noted that the variable was not selected simply by chance; it is known that rainfall has a significant negative effect on the release and dispersal of grass pollen from the plant (Norris-Hill and Emberlin 1991; Hart et al. 1994; Emberlin 1997).

An unusual feature of this study is the use of a variety of different independent variables in an attempt to forecast 10-day mean grass pollen counts. Grass pollen counts at the start of the season had a significant effect on the quantity of grass pollen released later in the year. The amount of grass pollen recorded during the period 21–30 May played a part in the regression models covering the periods 30 June–9 July and 30 July–8 August. The positive relationship between grass pollen counts from 21 to 30 May and grass

Fig. 3 Actual and predicted grass pollen counts in north London for 2002. Eight forecast periods from 21 May to 8 August. 25th percentile, mean and 75th percentile of 1961–1999 data set plotted for comparison



pollen counts at the beginning of July and August (Table 5) means that the forecaster is given an early indication of the magnitude of grass pollen counts to expect later in the season. Furthermore, winter averages of the NAO were used to forecast 10-day mean grass pollen counts for the periods 20–29th June and 10–19th July. These results emphasise the importance of considering large-scale patterns of climate variability when forecasting grass pollen counts.

Conclusion

The 30-day-ahead forecast model for north London performed well, particularly when predicting whether the grass pollen count would be higher or lower than the 1961–1999 mean (attaining 87.5% and 100% accuracy in 2000 and 2002, respectively). It is important to remember that this level of accuracy was achieved at a site for which it is inherently difficult to produce forecast models because of the complexities associated with the urban climate.

Attempting to predict pollen counts during distinct 10-day periods throughout the main flowering period of grasses is a novel approach. However, probably the most original method employed here was the use of winter averages of the NAO. The use of winter averages of the NAO to model pollen concentrations is innovative and marks a significant advance in pollen forecasting. The relationship between grass pollen counts and the NAO has been examined (D'Odorico et al. 2002) but there is room for further research in this area.

The ability to predict allergenic pollen counts for a 30-day period will be of assistance to the medical profession, including allergists planning treatment and physicians scheduling clinical trials. Furthermore, because of the just-in-time method of stock control often used in industry, such information will also be useful for pharmaceutical companies and the health care industry who market and stock hay fever treatments (Emberlin et al. 1999).

Acknowledgements This work was partially supported by Andrew Kress, Surveillance Data Incorporated, Plymouth Meeting, PA, USA. The authors wish to thank the British Atmospheric Data Centre (BADC) for providing access to the Met Office Land Surface Observation Stations Data. We also thank the Met Office who supplied the initial data via the BADC. The authors are also grateful to Mike Savage, St. Mary's Hospital Paddington, and the Environmental and Public Protection offices, Islington for use of the pollen data.

References

- Aas K, Bachert C, Bergmann K, Bergmann R, Bonini S, Bousquet J, de Weck A, Farkas I, Hejdenberg K (1997) European white paper: allergic diseases as a health problem. UCB, Brussels, Belgium
- Adams-Groom B, Emberlin J, Corden JM, Millington W, Mullins J (2002) Predicting the start of the Betula pollen season at London, Derby and Cardiff, United Kingdom, using a multiple regression model, based on data from 1987 to 1997. *Aerobiologia* 18:117–123
- Ahas R, Aasa A, Menzel A, Fedotova VG, Scheifinger H (2002) Changes in European spring phenology. *Int J Climatol* 22:1727–1738
- BAF (1995) Airborne pollens and spores: a guide to trapping and counting. The British Aerobiology Federation ISBN 0-9525617-0-0
- Burr ML (1999) Grass pollen: trends and predictions. *Clin Exp Allergy* 29:735–738
- D'Amato G (2000) Urban air pollution and plant derived respiratory allergy. *Clin Exp Allergy* 30:628–636
- D'Odorico P, Yoo J, Jaeger S (2002) Changing seasons: an effect of the North Atlantic Oscillation? *J Climate* 15:435–445
- EA (1999) Thames Tideway (Teddington to Tower Bridge) Local Environment Agency Plan: action Plan
- Emberlin J (1997) Grass, tree and weed pollens. In: Kay AB (ed) *Allergy and allergic diseases*, vol 2. Blackwell, Oxford
- Emberlin J (2000) *Aerobiology*. In: Busse WW, Holgate ST (eds) *Asthma and rhinitis*, vol 2. Blackwell, Oxford
- Emberlin J, Norris-Hill J (1991) Spatial variation of pollen deposition in North London. *Grana* 30:190–195
- Emberlin J, Savage M, Jones S (1993) Annual variations in grass pollen seasons in London 1961–1990: trends and forecast models. *Clin Exp Allergy* 23:911–918
- Emberlin J, Mullins J, Cordon J, Jones S, Millington W, Brooke 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
- Frenguelli G (2002) Interactions between climatic changes and allergenic plants. *Monaldi Arch Chest Dis* 57:141–143
- Fuller RM (1987) The changing extent and conservation interest of lowland grasslands in England and Wales: A review of grassland surveys 1930–84. *Biol Conserv* 40:281–300
- Gassman MI, Perez CF, Gardiol JM (2002) Sea-land breeze in a coastal city and its effect on pollen transport. *Int J Biometeorol* 46:118–125
- Goudie A (1994) *The human impact on the natural environment*. Blackwell, Oxford
- Goudie A (1996) *The Nature of the Environment*, Blackwell
- Hart ML, Wentworth JE, Bailey JP (1994) The effects of trap height and weather variables on recorded pollen concentration at Leicester. *Grana* 33:100–103
- Hopkins A, Davies RR (1994) Changing grassland utilization in the United Kingdom and its implications for pollen production and hay fever. *Grana* 33:71–75
- Hurrell JW (1995) Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science* 269:676–679
- Huynen M, Menne B, Behrendt H, Bertollini R, Bonini S, Brandao R, Brown-Fährlander C, Clot B, D'Ambrosio, De Nuntiis P, Ebi KL, Emberlin J, Orbanne EE, Galan C, Jäger S, Kovats S, Mandrioli P, Martens P, Menzel A, Nyenzi B, Rantio Lehtimäki A, Ring J, Rybníček O, Traidl-Hoffmann, Van Vliet A, Voigt T, Weiland S, Wickman M (2003) Phenology and human health: allergic disorders. Rome, Italy, Report of a WHO meeting
- Jones S (1995) Allergenic pollen concentrations in the United Kingdom. PhD Thesis, University of North London
- Larsson KA (1993) Prediction of the pollen season with a cumulated activity method. *Grana* 32:111–114
- Menzel A, Fabian P (1999) Growing season extended in Europe. *Nature* 397:659
- METO (2004) 1971–2000 Averages. Met Office <http://www.meto.gov.uk/climate/uk/averages/index.html>
- Norris-Hill J (1995) The modelling of daily Poaceae pollen concentrations. *Grana* 34:182–188
- Norris-Hill J, Emberlin J (1991) Diurnal variation of pollen concentration in the air of north-central London. *Grana* 30:229–234
- Ohashi Y, Kida H (2002) Local circulations in the vicinity of both coastal and inland urban areas: a numerical study with a meso-scale atmospheric model. *J Appl Meteor* 41:30–45
- Osborn T, Climatic Research Unit (2002) North Atlantic Oscillation. http://www.cru.uea.ac.uk/~timo/projpages/nao_update.htm
- Ottersen G, Planque B, Belgrano A, Post E, Reid PC, Stenseth NC (2001) Ecological effects of the North Atlantic Oscillation. *Oecologia* 128:1–14
- Pallant J (2001) SPSS survival manual, Open University Press

- Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421:37–42
- Recio M, Cabezudo B, Trigo M, Toro FJ (1998) Pollen calendar of Malaga (southern Spain), 1991–1995. *Aerobiologia* 14:101–107
- Root TL, Price JT, Hall KR, Schneider SH, Rosenweig C, Pounds JA (2003) Fingerprints of global warming on wild animals and plants. *Nature* 421:57–60
- Sanchez Mesa JA, Smith M, Emberlin J, Allitt U, Caulton E (2003) Characteristics of grass pollen seasons in areas of southern Spain and the United Kingdom. *Aerobiologia* 19:243–250
- Spieksma FT, Nikkels AH (1998) Airborne grass pollen in Leiden, the Netherlands: annual variations and trends in quantities and season starts over 26 years. *Aerobiologia* 14:347–358
- Spieksma FTM, Emberlin JC, Hjelmroos M, Jäger S, Leuschner RM (1995) Atmospheric birch (*Betula*) pollen in Europe: trends and fluctuations in annual quantities and the starting dates of the seasons. *Grana* 34:51–57
- Stenseth NC, Mysterud A, Ottersen G, Hurrell JW, Chan K-S, Lima M (2002) Ecological effects of climate fluctuations. *Science* 297:1292–1296
- Van Vliet AJH, Overeem A, De Groot R, Jacobs AFG, Spieksma FTM (2002) The influence of temperature and climate change on the timing of pollen release in The Netherlands. *Int J Climatol* 22:1757–1767
- Walther G-R, Post E, Convey P, Menzel A, Parmesan C, Beebee TJC, Fromentin J-M, Hoegh-Guldberg O, Bairlein F (2002) Ecological responses to recent climate change. *Nature* 416:389–395