Predicting the severity of the grass pollen season and the effect of climate change in Northwest Europe

Alexander Kurganskiy1,‡, Simon Creer2, Natasha de Vere3,4, Gareth W. Griffith4, Nicholas J. Osborne5,6, Benedict W. Wheeler5, Rachel N. McInnes7, Yolanda Clelow7, Adam Barber7, Georgina L. Brennan2,8, Helen M. Hanlon7, Matthew Hegarty4, Caitlin Potter4, Francis Rowney5,9, Beverley Adams-Groom1, Geoff M. Petch1, Catherine H. Pashley10, Jack Satchwell7, Letty A. de Weger11, Karen Rasmussen12, Gilles Oliver13, Charlotte Sindt13, Nicolas Bruffaerts14, The PollerGEN Consortium‡, Carsten A. Skjøth17

Allergic rhinitis is an inflammation in the nose caused by overreaction of the immune system to allergens in the air. Managing allergic rhinitis symptoms is challenging and requires timely intervention. The following are major questions often posed by those with allergic rhinitis: How should I prepare for the forthcoming season? How will the season’s severity develop over the years? No country yet provides clear guidance addressing these questions. We propose two previously unexplored approaches for forecasting the severity of the grass pollen season on the basis of statistical and mechanistic models. The results suggest annual severity is largely governed by preseasonal meteorological conditions. The mechanistic model suggests climate change will increase the season severity by up to 60%, in line with experimental chamber studies. These models can be used as forecasting tools for advising individuals with hay fever and health care professionals how to prepare for the grass pollen season.

INTRODUCTION
Pollen allergy, primarily causing allergic rhinitis (AR; also known as hay fever), affects up to 40% of the European population (1). AR is an inflammation in the nose caused by overreaction of the immune system to allergens in the air. According to the European Academy of Allergy and Clinical Immunology: “Allergic rhinitis represents a global health care problem affecting 10 to 20% of the total population, making AR the most prevalent chronic non-communicable disease”. Grass (Poaceae) pollen is particularly important as the prevalence of sensitization is greater than other pollen types in most (European) countries (2). Pollen allergy can have substantial negative impacts on quality of life, for example, affecting sleep and performance at work or school (3). Only a small fraction of patients with symptoms have been diagnosed by a specialist (3), although efficient treatment of patients with hay fever during the pollen season has been shown to reduce symptoms and improve quality of life (4). AR is also often comorbid with asthma (5–7), and hay fever symptoms, asthma exacerbations, and related hospital admissions increase with the severity of the pollen season (8). The pollen season severity is defined as the total amount of pollen per season in this study. A general increase in severity of the pollen season has also been observed to increase the number of sensitized patients (9). The severity of the pollen season is therefore important for patients with hay fever. Major questions posed by those with hay fever often include the following: How bad is this year’s season going to be? How should I prepare for the forthcoming season? How will the season’s severity (the amount of pollen) develop over the years? However, reliable methods for answering these questions have not previously been developed. A telephone-based screening of 7004 patients with self-reported allergic disease from 10 European countries revealed that one-third of patients were not satisfied with their treatment, and two-thirds experienced restrictions in daily activities (3). Forecasts of the pollen season severity are consequently important for patients and the health sector on seasonal and decadal time scales, respectively. In addition, longer-term forecasting is useful for health system planning, for example, in preparing for climate change impacts. Climate change can lead to higher pollen concentrations and longer pollen seasons, resulting in a doubling of the population sensitization to pollen (9).

The seasonal pollen integral (SPIn), the sum of pollen concentrations obtained at a site during the grass pollen season, is the main parameter used to describe the severity of the pollen season (10). The SPIn can be used to quantify pollen exposure, related health outcomes, and associated costs among the population (11), and comprises two main components: annual pollen production (APP) and atmospheric transport (AT). Studies suggest that AT contributes only 10 to 20% of the SPIn for grasses (12). The remaining 80 to 90% is therefore dominated by variations in APP, which is a parameter that can potentially be forecast for both the upcoming season and for future decades, e.g., in relation to climate change.
Here, we propose two novel mathematical approaches for studying and forecasting the severity of the grass pollen season using statistical and mechanistic models. Approach 1 is a statistical approach, which is site based and designed for seasonal forecasting. The statistical approach is based on building a regression model using the SPIₙ, air temperature, and precipitation observation data. Approach 2 is a mechanistic approach, which is designed for long-term assessments, such as the impact of climate change or mitigation scenarios. The concept of the mechanistic approach relies on describing the interannual variation of SPIₙ through variation in grass growth, measured primarily via net primary production (NPP), the net production of organic carbon by plants in an ecosystem (13). The two novel approaches will have a practical usage in daily forecasting routines. We use the approaches in the study to test the following scientific hypotheses:

1) The severity of the grass pollen season is a regional-scale (i.e., 10 to 1000 km) phenomenon.
2) The severity of the grass pollen season can be forecast using preseasonal meteorological conditions.
3) Long-term changes in the severity of the grass pollen season can be simulated by using land surface models, e.g., due to increased CO₂ under future climate change scenarios.

Addressing these hypotheses, we provide novel methods for predicting the severity of the grass pollen season and reveal the current limitations of these methods both in relation to the upcoming season and in relation to scenarios by using the Northwest European region (Fig. 1 and Table 1) as a study area.

**RESULTS**

Modeling grass pollen interannual variation: A statistical model

We built a statistical regression model to simulate and predict the SPIₙ and pollen exposure at 28 pollen-monitoring stations located in Northwest Europe (Fig. 1 and Table 1; Materials and Methods). The correlation coefficients between the individual sites (Fig. 1) in relation to seasonal severity (SPIₙ) are typically below 0.6 to 0.7 (Fig. 2A). Only 65 of 291 correlation coefficients are statistically significant (P < 0.05), and the linear regression line shows a decreasing correlation with distance between stations (Fig. 2B). The result of the χ² = 3.865 at P = 0.01 is not statistically significant (P = 0.049304). The null hypothesis, that there is a connection between the sites, is therefore rejected. Each site should therefore be considered individually when building a statistical regression model.

The start of the season varies by approximately a month, from late April at the French sites La Roche-sur-Yon and La Rochelle (latitude 46.2°N) in the south, to the end of May or early June at the UK sites East Riding, Belfast, and Invergowrie (latitude 56.5°N) in the north (Fig. 1 and fig. S1). Thus, we can consider that the preseasonal period (March to April) chosen in the study is applicable for the entire region, and it can be used for building a statistical (regression) model.

The four different statistical regression models yielded coefficients of determination varying from 0.63 (model 1, Fig. 3A) to 0.75 (model 4, Fig. 3D). The number of statistically significant stations increases when preseasonal meteorological conditions...
Table 1. Selected pollen/meteorological observation sites and temporal data coverage used for the statistical and mechanistic models. The first 28 stations are used in both statistical and mechanistic models, whereas the last 6 stations are used in mechanistic only. NL, The Netherlands; BE, Belgium; DK, Denmark; FR, France; UK, United Kingdom.

<table>
<thead>
<tr>
<th>Pollen station, country</th>
<th>Station code</th>
<th>Station code</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Nearest meteorological station</th>
<th>Years, statistical model</th>
<th>Years, mechanistic model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ipswich, UK</td>
<td>GBIPSW</td>
<td>52.06°N</td>
<td>1.20°E</td>
<td></td>
<td>Wattisham</td>
<td>2011–2018</td>
<td>2011–2016</td>
</tr>
<tr>
<td>East Riding, UK</td>
<td>GBEROY</td>
<td>53.84°N</td>
<td>0.43°W</td>
<td></td>
<td>Bridlington MRSC</td>
<td>2011–2018</td>
<td>2011–2016</td>
</tr>
<tr>
<td>Derby, UK</td>
<td>GBDERB</td>
<td>52.92°N</td>
<td>1.50°W</td>
<td></td>
<td>Church Lawford</td>
<td>1996–2005</td>
<td>1996–2005</td>
</tr>
<tr>
<td>Poitiers, FR</td>
<td>FRPOIT</td>
<td>46.58°N</td>
<td>0.34°E</td>
<td></td>
<td>Biard</td>
<td>2010–2018</td>
<td>2010–2016</td>
</tr>
<tr>
<td>La Rochelle, FR</td>
<td>FRROCH</td>
<td>46.17°N</td>
<td>1.15°W</td>
<td></td>
<td>Ile de Re</td>
<td>2010–2018</td>
<td>2010–2016</td>
</tr>
<tr>
<td>Kings College, UK</td>
<td>GBLON2</td>
<td>51.51°N</td>
<td>0.12°E</td>
<td></td>
<td>–</td>
<td>2012–2016</td>
<td>–</td>
</tr>
<tr>
<td>Eskdalemuir, UK</td>
<td>GBESKD</td>
<td>55.31°N</td>
<td>3.21°W</td>
<td></td>
<td>–</td>
<td>2011–2016</td>
<td>–</td>
</tr>
<tr>
<td>Bath, UK</td>
<td>GBBATH</td>
<td>51.38°N</td>
<td>2.36°W</td>
<td></td>
<td>–</td>
<td>2011–2016</td>
<td>–</td>
</tr>
<tr>
<td>Exeter Uni, UK</td>
<td>GBEXEU</td>
<td>50.74°N</td>
<td>3.53°W</td>
<td></td>
<td>–</td>
<td>2014–2016</td>
<td>–</td>
</tr>
</tbody>
</table>
are included, from 4 (only SPIn, table S1) to 14 (SPIn + preseas-
onal air temperature, table S2), to 12 (SPIn + preseasional precip-
itation, table S3) and 20 (SPIn + preseasional precipitation and air temperature, table S4). This shows that including both
preseasional meteorological conditions and the measured SPIn provides the most robust way of forecasting the severity of the forthcoming grass pollen season.

The modeled SPIn values interpolated to full geographic coverage of the region for 2014 (Fig. 4A and fig. S2) show that forecast variations in seasonal severity mainly varied within ±20% of the mean modeled SPIn. The observation-based map (Fig. 4B) shows a similar picture with the majority of variation being within ±20%, but parts of central United Kingdom and Denmark showed areas above 1.2 (i.e., SPIn interannual variation is higher than 20% of the mean value) and most of the Netherlands below 0.8 (i.e., SPIn interannual variation is lower than 20% of the mean value). This suggests that extreme variations are more difficult to capture. It was also supported by the results from the cross correlation procedure using 24 pollen stations: the cross correlation yielded $R^2$ values of 0.02 and 0.05 and root mean square error values of 0.28 and 0.12 for the modeled and observed data, respectively. This sensitivity study demonstrates the high dependency of all data points in the mapping procedure; hence the local signal of the SPIn, and further supports the rejection of the null hypothesis. This also suggests that there is no connection (i.e., no spatial correlation for SPIn) between the stations within the entire region, and each site should be considered individually when building a statistical (regression) model.

The regression model was tested to identify the number of years providing the best correlation between the modeled and observed SPIn data. The tests were performed using the model with the highest coefficient of determination (model 4; fig. S3 and table S4), taking into account all correlation coefficients (fig. S3A) and only those that were significant ($P < 0.05$; fig. S3B). The results indicate that the average number of years used to obtain the maximum correlation values is equal to 8 (SD = 1).

Modeling grass pollen interannual variation: A mechanistic model

We studied the interannual variations in grass SPIn through the interannual variations in NPP in grasslands with C3 grasses at 34 pollen monitoring stations located in Northwest Europe (Fig. 1 and Table 1). The Joint UK Land Environmental Simulator (JULES) model (14) was used to simulate NPP for grasses over 407 pollen seasons at the selected stations for the years 1996 to 2016. Comparison of NPP and SPIn interannual variations for all sites (Fig. 5) demonstrated positive and significant associations between NPP and SPIn ($R = 0.2$, $P = 1.55 \times 10^{-5}$). The correlation values (table S5) were positive and significant at six UK stations, with values varying from 0.5 at Worcester to 0.94 at Ipswich. Moreover, the SPIn varied within a factor of 2, whereas NPP interannual variation was lower and within 50% at the individual sites and when considering all sites together (Fig. 5). These data suggest that small variations in NPP can cause large variations in SPIn.

We also performed an additional set of sensitivity model runs to study the influence of doubled atmospheric CO$_2$ concentration on NPP values (fig. S4 and data file S1) at each station and for the same years. Doubling the initial value of global averaged CO$_2$ atmospheric concentrations (i.e., from $5.241 \times 10^{-4}$ mmr to $10.482 \times 10^{-4}$ mmr) in the JULES parameters led to an increase in the NPP values of up to 60% (fig. S4 and data file S1). However, it did not affect the interannual variations of NPP and, hence, the correlation values between the simulated NPP and SPIn since the doubled value was used for all years in the model.
DISCUSSION

Our study sheds new light on the current feasibility and limitations of predicting the grass SPIn and thereby grass pollen exposure, using statistical and mechanistic models in the Northwest European region. We do this by testing three scientific hypotheses.

Hypothesis 1 tests whether grass SPIn is a regional-scale (i.e., 10 to 1000 km) phenomenon, and the analyses indicate that this hypothesis should be rejected. Analyzing the SPIn time series did not reveal any strong and statistically significant connection between the stations in the region. Similar results have also been obtained...
studying SPIn at three UK stations: London, Derby, and Cardiff (15). The differences in the interannual SPIn variation were explained by the geographical variation of local meteorological conditions and changes in grassland cover in the areas where the stations were located. The SPIn has been considered as a regional-scale phenomenon for birches (*Betula* spp.) in the boreal region of Northern and Northeastern Europe (16), where a single regression model has been applied to describe interannual variation of the birch SPIn. In the context of birch, the single regression approach is feasible because the annual pollen productivity in birch within the boreal region is synchronized over large areas (17). However, the statistical approach showed poor performance outside the region (e.g., in Brussels) due to differences in climate, i.e., humid continental, boreal versus maritime (18). Our studies also show that the
We found that preseasonal air temperature and precipitation are the parameters driving the grass SPIn in the statistical model. However, the North Atlantic Oscillation (NAO) index has been shown to be an important parameter influencing the grass SPIn in recent studies, for example, (20). Therefore, including the NAO index in the regression models could potentially be used to predict the severity of the grass pollen season. The NAO is, however, a single index describing overall synoptical-scale weather conditions within the study region and is related to large-scale phenomena such as the strength of the jet stream that, in turn, also affects local meteorology such as rain, temperature, and wind. The initial hypothesis that grass pollen concentration was a large-scale phenomenon was rejected. This suggests that models with strong predictive power should not rely on large-scale phenomena like the NAO, which is supported by previous findings (20), revealing limited predictive power for regression models using this approach. The approach presented here offers spatial and temporal advances in the context of forecasting. First, it allows for the usage of local environmental data instead of a single index. Second, it may be applied before the pollen season commences and can therefore be a tool in a forecasting environment.

Using different years for different stations could be construed to be a limitation of the study. However, the statistical approach applied in the study was developed for modeling and potentially forecasting of SPIn for the next year, consistent with recent statistical model development for birch (16). In contrast, the pollen data are not used for estimating the impact of climate change as this question is addressed using the JULES model, a tool commonly applied in the context of climate change scenarios (21).

We extended the statistical model results from point-based locations to full geographic coverage by interpolating them on a map using the year 2014 as an example. Maps illustrating the severity of the grass pollen season are an end-user product, which will help predict grass pollen exposure and provide advice to people with grass pollen allergies to help prevent/minimize the allergy symptoms and, consequently, reduce hospital admissions and health-related costs. The results (Fig. 4) demonstrate that the mapping approach is sensitive for all of the data points for the year of 2014. This suggests again that the grass SPIn is not a regional-scale phenomenon in the Northwest European region. Introducing more observation points located in the region could improve the agreement between the model- and observation-based maps. It also suggests that an observational-based approach predicting the severity of the grass pollen season nationwide would require a high density of observation sites. For instance, one should take with caution the area with the big data gap between Denmark and the rest of the observation sites (Figs. 1 and 4) since the values are obtained by interpolation using stations located far away from each other. In contrast to the regional-scale synchronization of pollen productivity seen in birches (17), grass pollen productivity is a local-scale phenomenon affected by the impact of local environmental variables (e.g., air temperature, precipitation, and land cover).

We identified that the number of years generating the best correlation using the statistical model was between 7 and 9, dependent on site. The latter suggests that time series of varying length may be applied, but the use of longer time series should generally be avoided. Similar results were obtained for the birch SPIn using a regression model for Northern Europe (17). It was shown that the model could explain up to 92% of the birch SPIn interannual

---

**Fig. 5.** Scatter plot for comparison of the SPIn and NPP interannual variations at the selected pollen sites. NPP variations are calculated using sums of daily NPP from 1 March until the grass pollen season start. The season start is calculated as the day when the accumulated sum of daily grass pollen concentrations reached 2.5% of the annual pollen sum. The results are significant with $P = 2.38 \times 10^{-5}$. The $R$ value is based on the calculations of the Spearman correlation coefficient between the time series of NPP (x axis) and SPIn (y axis) interannual variation.

---


**SCIENCE ADVANCES | RESEARCH ARTICLE**

regional-scale synchronization found for birches is not present for grasses (19). We therefore accept the alternative hypothesis: There is no connection between SPIn for grass pollen sites beyond 20 km, suggesting that the season severity is a local-scale phenomenon for grass pollen. New methods for describing this are therefore needed.

Hypothesis 2 tests whether the severity of the grass pollen season can be forecast using preseasonal meteorological conditions. The analyses reveal that the hypothesis should be accepted. Including preseasonal meteorological conditions in the statistical model provided the best model performance with the highest statistical significance throughout the stations. The statistical regression model based on the SPIn, preseasonal temperature, and precipitation data (model 4, table S4) provided the highest number of stations (71% of sites) with positive and significant correlations between the modeled and observed SPIn time series in comparison with models 1 to 3. However, it should be noted that for some sites (e.g., Worcester, UK), it is enough to use either preseasonal temperature (model 2) or precipitation alone (model 3) to ensure positive and significant correlation between modeled and observed SPIn. The previous study showed high accuracy of the regression model ($R^2 > 0.9$) at three UK sites (15). However, in the latter case, the approach was based on building a model with individual equation and regression coefficients using temperature and precipitation data. The approach presented here is more unified: We use one equation with individual regression coefficients at each site. To our knowledge, this is the first ever attempt at applying this approach to multiple stations located in different parts of the Northwest European region.
enhanced CO₂ concentrations and the future dynamics of pollen production (22). An increase in timothy grass (Phleum pratense) growth and pollen production at rural and urban sites (23). Doubling CO₂ concentration in the model showed a large increase in NPP, which enhanced the SPIn and pollen exposure. A similar increase in NPP is supported by a study showing that ragweed pollen production increased by up to 55% associated with high CO₂ levels (24). An increase in timothy grass (Phleum pratense) pollen production (about 50% per flower) caused by enhanced atmospheric CO₂ levels has also been found in an experimental chamber study (25). However, the responses of grasses to elevated CO₂ concentrations are species specific and vary heterogeneously according to other environmental variables, such as soil nutrients (e.g., nitrogen) (25–27). Overall, there are complex relationships between grass species, ecosystems, and environmental variables (28), with large variations in both growth and flowering of different species (29) in response to CO₂ and nitrogen. Grass pollen productivity is therefore likely to vary according to species and soil nutrient conditions. Current vegetation models cannot replicate chamber and field studies, for grasses in general. However, our study has shown the robustness of vegetation models to simulate grass pollen production by applying the mechanistic approach. To our knowledge, it is the first ever mechanistic modeling attempt made for grasses using multiple stations located in different parts of the Northwest European region. Nevertheless, further work is needed to cover the most important grass species and their response to the combined effects of climate change and nutrient availability in order to capture the large variations between species.

The results presented here show the applicability of the mechanistic approach at some stations, but not across broader geographically and environmentally heterogeneous sites. Improvements are, therefore, needed to increase the utility of the mechanistic approach for the entire Northwest European region. As an extension of the current approach, local variations of meteorological and environmental variables (e.g., CO₂) should be taken into account for both NPP and the SPIn. Moreover, footprint modeling (30) can be performed to estimate local AT and the distribution of local grass pollen sources for more detailed investigation of spatial attributes of the SPIn variations. The mechanistic model thus has the potential practical application of being used for the estimation of local variation in grass pollen productivity, hence pollen exposure, throughout the Northwest European region.

This study has been carried out using data for grass pollen identified at the family level, i.e., including all grass species. Different grass species flower at different times and cause a range of allergic reactions (19). Currently, there is no dataset for grass pollen concentrations at the species level covering a long-term period (N = 20 to 30 years). The models presented here will be applicable for predicting the SPIn at species level, when the observational data with a sufficient number of years are available in the future. However, performing the season severity forecast at the family level is useful since the main allergen groups (groups 1 and 13) are present in most allergenic grasses and up to 90% of grass-sensitized people react to them (31).

The pollen measurement may have large uncertainty on daily observed concentrations (32), but the large number of daily observations can reduce the uncertainty of the SPIn to less than 10% for a whole season of observations, where the exact value depends on the number of daily observations and counting method (33). Each instrument is affected by a systematic instrumental bias between 5 and 72% (34), potentially increasing the observed uncertainty of the measured SPIn. Nevertheless, the relative annual variation in SPIn will not be affected by this systematic bias. Both the statistical and mechanistic models explain the relative variation between years, and their performance will therefore only be affected by the random uncertainty of the annual SPIn, which is assessed to be less than 10%.

We found that the grass SPIn varies from year to year and from station to station by a factor of 2 to 4. The SPIn has been used in atmospheric models to take into account the seasonal magnitude of pollen emission and its spatial and interannual variation for different pollen types (35, 36). Therefore, our findings have the potential to be used in atmospheric dispersion models for Northwest Europe or larger regions throughout the world where sufficient robust pollen data are available. Our findings will improve forecasting of the grass SPIn and grass pollen exposure, with concomitant socioeconomic benefits to global society and health care systems.

### MATERIALS AND METHODS

**Pollen and meteorological data**

Thirty-four pollen observation stations located in Northwest Europe were selected for the study (Fig. 1). The region covers stations located in the United Kingdom, Denmark, The Netherlands, Belgium, and Northern France. The period of the study corresponds to the grass pollen seasons 1996–2018 mostly covering May to September. The observational data contain daily mean grass pollen concentrations obtained using 7-day volumetric samplers of the Hirst type (37) and were analyzed according to standard methods in aerobiology (10). The UK grass pollen data were quality controlled by excluding the seasons with substantial numbers of gaps (i.e., >7 days) in the observed time series within the main pollen season (38). The main grass pollen season was identified using the dates when accumulated grass pollen concentrations reached 2.5 and 97.5% of the annual pollen integral, thus delimiting the start and the end of the season. Several approaches such as the accumulated catch of 50, 75 pollen grains as well as a fraction of the catch are possible definitions of the start of the pollen season (39). The percentage approach chosen
here has the advantage, in contrast to the accumulated catch, that the known bias associated with each observing instrument (34) will, by using the fractional approach, not affect the calculated relative variation in SPIn, thereby keeping uncertainty low in the regression modeling. Extraction of SPIn and the dates of the grass pollen season start/end were performed using a predefined extraction routine on the European Aeroallergen Network webpage (https://ean.polleninfo.eu/Ean/) for the rest stations and suitable pollen seasons (i.e., no gaps >7 days).

Meteorological data were obtained from the Global Surface Summary of the Day (GSOD) dataset provided by National Oceanic and Atmospheric Administration. The dataset has global geographic coverage and provides meteorological data worldwide. The selected meteorological stations met two main requirements: (i) They provided meteorological data corresponding to preseason conditions and (ii) were located close to the selected pollen sites. Thus, precipitation and the maximum daily air temperature data were extracted and matched with the pollen stations for the selected years.

**SPIn sensitivity within the region: Connection between the stations**

A correlation matrix was calculated to investigate whether a connection existed between interannual variations of the SPIn at the stations within the region studied. Since the stations covered different numbers of years (Table 1), the matrix includes only sites with eight or more overlapping years (38). The correlation coefficients were also analyzed depending on the distance between the stations (Fig. 2). The correlation coefficients were separated into two groups defined by the distance between stations and two classifications used in mesoscale meteorology and air pollution studies: meso-beta (20 to 200 km) and meso-alpha (200 to 2000 km) defined by (40). The number of significant \((P < 0.05)\) and nonsignificant coefficients was calculated within each group. A \(\chi^2\) test was performed using the calculated numbers to determine the degree of significance (difference of significance) between each scale.

**Modeling grass pollen interannual variation: A regression model**

The statistical approach used in this study was devoted to building a geostatistical regression model that goes beyond traditional approaches. The grass pollen SPIn, preseason air temperature, and precipitation observations were used as input data for the 28 pollen observation sites covering Northwest Europe, totaling 386 pollen seasons (Table 1). The selected pollen sites contained a minimum of 8 years of pollen data according to the requirements of the aerobiological studies (38). The regression model followed the approach proposed by (41) and later applied by (42) for grasses in Cordoba, Spain. The model used here to simulate SPIn was

\[
SPIn(t) = \exp \left( \ln(SPIn(t-1) - rm + \exp(a*\ln(SPIn(t-1)) + b + c*\ln(T_{max}) + c*\ln(N_{prec})) \right)
\]

where SPIn(t) and SPIn(t – 1) are the SPIn referring to the current \((t)\) and previous \((t – 1)\) pollen seasons; \(rm\) is a constant representing the maximum productive rate observed; \(a, b, c,\) and \(d\) are regression coefficients; \(T_{max}\) is maximum daily air temperature averaged for the period 1 March to 30 April for each considered year; and \(N_{prec}\) is the number of precipitation days during the same period as \(T_{max}\).

The number of precipitation days \((N_{prec})\) was selected instead of the amount of precipitation (in millimeter scale) since days with rain events were considered a better predictor for water availability used for grass growth and pollen production. There are some days with high precipitation volumes (in millimeter scale), but because of surface water runoff in flood events, the water available to plants is the same as that in much lower days, biasing the result.

While the equation is general, both constants and input data are specific for each site. Four versions of the model (Eq. 1) were applied in this study: (i) model 1, a model including SPIn only; (ii) model 2, a model taking into account SPIn and \(T_{max}\); (iii) model 3, a model with SPIn and \(N_{prec}\); and (iv) model 4, a model including SPIn, \(T_{max}\), and \(N_{prec}\) (as shown above). The models were built using the nonlinear least square function in R software version 3.6 (https://r-project.org). The choice of the period for \(T_{max}\) and \(N_{prec}\) applied for each region is justified by analyzing the starting dates of grass pollen season (Fig. 1). The modeled SPIn values were compared with observations by means of statistical (correlation) analysis for each individual station as well as globally, i.e., including all stations. The global scatter plots with corresponding \(R^2\) values are shown in Results, whereas the statistical summaries with modeled and observed SPIn time series are available in the Supplementary Materials. The optimal number of years for building the regression model was identified using the stations with 20 years of pollen data for the period 1999–2018: Worcester (UK), Leicester (UK), Leiden (The Netherlands), Brussels (Belgium), De Haan (Belgium), Copenhagen (Denmark), and Viborg (Denmark). Model 4 was used to determine the optimal number of years. The model was run for various numbers of years starting with the eight most recent years then increasing the number by 1 until it reached 20 years for each of the seven selected pollen stations. Thus, the number of years providing the maximum correlation between the modeled and observed SPIn was identified for individual stations and averaged over the stations with corresponding SD (fig. S3).

The modeled and observed SPIn interannual variation was extended from the point-based locations to full geographic coverage within the study region using interpolation with the inverse distance method. The sensitivity to individual points was assessed using cross correlations, using similar approaches that have been implemented in related pollen studies on ragweed [see, e.g., (43) and references therein] developed by using standard recommendations by the U.S. Environmental Protection Agency.

**Modeling grass pollen interannual variation: A mechanistic model**

The mechanistic approach used in this study describes the interannual variation of SPIn through variation in grass growth, measured primarily via NPP. NPP is defined as the net production of organic carbon by plants in an ecosystem (13). The main assumption used in the approach suggests that SPIn varies as APP from year to year. APP interannual variation, in turn, can be described through variation of NPP. This can be written as

\[
APP(x,y) = veg(x,y) * NPP_{var}(x,y) * N_{total}
\]

where APP is the annual pollen production at the given location with geographical coordinates \(x\) and \(y\) representing longitude and latitude, \(veg\) is the fraction of the grid cell covered by grass vegetation with coordinates \(x\) and \(y\), \(NPP_{var}\) is the interannual variation...
of NPP for the given year, and \( N_{\text{total}} \) is the total amount of grass pollen emitted per season represented by a constant value equal to \( 10^{12} \) pollen/m\(^2\)year.

The JULES model has been used to simulate NPP. JULES is a process-based model simulating land-atmosphere interactions represented by fluxes of carbon, water, momentum, and energy between the surface and the atmosphere (14). The model is globally applicable, and it can be used to perform simulations for different plant functional types: broadleaf trees, needle-leaf trees, C3 and C4 grasses, shrubs, and crops (44). JULES (v5.1) has been configured to represent different plant functional types: broadleaf trees, needle-leaf trees, C3 and C4 grasses, shrubs, and crops (44). JULES (v5.1) has been configured to

represented by fluxes of carbon, water, momentum, and energy be-

Inconsistent accounting of variations in grass pollen seasons in Northern and North-Eastern Europe.

**Statistical analysis**

The statistical analysis was performed by calculating either the Spearman or Pearson correlation coefficients (\( R \)) after first testing the hypothesis that the annual SPIn could be non-normally distributed using the Shapiro-Wilk test (with the level of significance \( \alpha = 0.01 \)) using the “R Stats package” in R software. The correlation coefficients (\( R \)) and coefficients of determination – \( R^2 \) (the statistical model) were calculated as described in the corresponding subsections of Material and Methods. The observed SPIn time series covering all available years and stations was also analyzed by calculating the mean, median, SD values, and the Shapiro-Wilk normality test. The results are shown in table S6. \( P \) and \( N \) values are reported in the main text, figures, and tables in the Supplementary Materials where applicable.

**REFERENCES AND NOTES**


et al., 7, 28. 30. 32. (FACE) at two levels of M. Implications of of C. area within H. 1602 (2017). R. J. C. L. reproduction of ozone levels. M. 152 B. K. of and of J. var. variations M. V. R. B. J. in (2017). M. V. R. A. generalized phenological model. Allergy Clin. Immunol. 29, 29–33 (2017). M. (2017). E. Smyshlyaev, E. M. V. R. A. Interaction of PLOS ONE 9, 114 E. Bazzaz, Interaction of A. K. and C. A. S. designed the study. A. K. installed the models, performed the model developments for the NWE region, the model runs, and analysis of the results. A.K. and C.A.S. prepared initial paper draft. A.B., B.A.-G., G.M.P., C.H.P., J.S., J.A.d.W., K.R., G.O., C.S., N.B., J.A., J.Ba., J.Be., C.A.F., R.N.M., S.P., K.S., J.T., J.Z.-C. contributed to the writing of the paper. Acknowledgments: We thank the National Oceanic and Atmospheric Administration for supplying global summary of the day meteorological data. T. Richard Matthews (CEH) and A. Martinez-de-la Torre (CEH) are acknowledged for fruitful discussions of the JULES modeling issues. We are thankful to A.-P. Holm (Asthma Allergy Denmark) for extracting and providing the grass SPIN data for two Danish sites (Copenhagen and Viborg) involved in the study. We are also very grateful to R. Jones at the University of Exeter for coordinating and supporting the Exeter PollerGen project. We thank P. Harvey at the Met Office, Eskdalemuir Geomagnetic Observatory, UK, for collecting the pollen data. We acknowledge all pollen counters that over the years collected data at the involved monitoring sites. Funding: This work was supported by the Natural Environment Research Council (grant numbers NE/N002375/1, NE/N002431/1, NE/N002105/1, and NE/N001710/1). C.H.P. is supported by the NIHR Leicester Biomedical Research Centre and the Midlands Asthma and Allergy Research Association (MAARA). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. The James Hutton Institute receives financial support from the Scottish Government Rural and Environmental Science and Analytical Services (RESAS) division. Author contributions: A.K. and C.A.S. designed the study. A.K. installed the models, performed the model developments for the NWE region, the model runs, and analysis of the results. A.K. and C.A.S. prepared initial paper draft. A.B., B.A.-G., G.M.P., C.H.P., J.S., J.A.d.W., K.R., G.O., C.S., N.B., J.A., J.Ba., J.Be., C.A.F., R.N.M., S.P., K.S., J.T., J.Z.-C. collected samples, counted pollen, and provided grass pollen observation data for their respective sites. All authors contributed to the writing of the paper. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The GSDO meteorological dataset used in this study is freely available at https://catalog.data.gov/dataset/global-surface-summary-of-the-day-gsd. Additional data are available from the authors and/or data owners upon request. The WFDEI meteorological dataset is available at ftp://rfdatalab/forcedata/ftp.iiasa.ac.at [see details in (43)]. Submitted 12 September 2020 Accepted 5 February 2021 Published 26 March 2021 10.1126/sciadv.abd7658 Citation: A. Kurganskiy, S. Creer, N. de Vere, G. W. Griffith, N. J. Osborne, B. W. Wheeler, R. N. McNines, Y. Clewlow, A. Barber, G. L. Brennan, H. M. Hanlon, M. Hegarty, C. Potter, F. Rowney, B. Adams-Groom, G. M. Petch, C. H. Pashley, J. Satchwell, L. A. de Weer, K. Rasmussen, G. Oliver, C. Sindt, N. Bruffaerts, The PollerGEN Consortium, C. A. Skjeth, Predicting the severity of the grass pollen season and the effect of climate change in Northwest Europe. Sci. Adv. 7, eabd7658 (2021).
Predicting the severity of the grass pollen season and the effect of climate change in Northwest Europe


Sci Adv 7 (13), eabd7658.
DOI: 10.1126/sciadv.abd7658