Representativeness of point-wise phenological Betula data collected in different parts of Europe

Pilvi Siljamo1, Mikhail Sofiev2, Hanna Ranta3, Tapio Linkosalo3, Eero Kubin4, Rein Ahas5, Eugene Genikhovich6, Katarzyna Jatczak7, Victoria Jato8, Jiří Nekovář9, Alexander Minin10, Elena Severova11 and Valentina Shalaboda12

ABSTRACT

Aim We examine issues of uncertainty regarding the spatial and temporal representativeness of phenological observations using a newly compiled Europe-wide data base of phenological observations for Betula species.

Location Europe.

Methods A new data base was compiled from national phenological observations covering 15 European countries, with the longest observational periods exceeding several decades for some sites. From this, the spatial and temporal representativeness of phenological observations were evaluated via statistical analysis.

Results The results showed that there was a significant and irreducible uncertainty related to the use of data of a single station, which varied from 3 to 8 days depending on the station location. In more continental and northern climatic zones the uncertainty was lower, probably due to faster spring-time weather developments. In mild climatic conditions, the uncertainty of dates of the phenological phases registered by a single station exceeded 1 week. The considerable number of data allowed us to preliminarily estimate the features of some stations, marking them as ‘late’, ‘early’, ‘representative’ or ‘random’, depending on the dates reported by these sites and the corresponding regional means.

Main conclusions The uncertainties discovered in single-site phenological observations are significant for virtually any potential application. Possible approaches for handling the uncertainty problem are station pre-averaging and spatial regularization of the data set, pre-selection (down-sampling) or changing the description of the phenomena from deterministic to probabilistic.

Keywords Betula, Europe, flowering, leaf unfolding, phenological observations, phenological stages, station representativeness, sub-grid variability.

INTRODUCTION

Phenological observations – the dates of leaf bud burst, leaf unfolding, start of flowering, etc. – are one of the most important (and sometimes the only) sources of information on the physiological condition of plants and their reactions to external forcing (Sparks & Carey, 1995; Sparks & Menzel, 2002; Menzel et al., 2006). Consequently, the number of studies in which these data are used for evaluation of the integrated characteristics of climate, and its changes during recent decades, is growing rapidly (Heikinheimo & Lappalainen, 1997; Parmesan & Yohe, 2003; Chen et al., 2005; Črepinšek et al., 2006). These data have also been used for the development, parametrization and evaluation of various (semi-)empirical models of phenological phases (e.g. Häkkinen et al., 1998; Linkosalo, 2000; Rötzer & Chmielewski, 2001; Schaber & Badeck, 2003). Many national and international phenological networks provide monitoring data for such investigations. An extensive list of these networks is presented at the website of the European Phenological Network (http://www.dow.wau.nl/msa/epn).
A significant limitation of phenological archives is that their features vary between different countries. As a result, many European regions are represented by just a few monitoring sites. This raises questions of the spatial representativeness of existing networks for the corresponding regions and, in particular, the representativeness of a single observational point for its neighbourhood. Such types of problems are well known in meteorology and air pollution, where the cost of measurements is often high and thus problems concerning the optimal number of stations and their optimal locations are important (e.g., Berg & Schaug, 1994). However, the phenomena related to meteorology and air quality have different features from those of phenological processes. First of all, they are much more dynamic and have much wider variability than the slowly progressing phenomenological developments. The spatial scales of meteorological changes are mainly affected by synoptic processes that are hundreds of kilometres in size. A similarity, however, between these two types of processes, is that the observations of a station are largely affected by its surroundings (WMO, 1996). Therefore, although the practices accepted in meteorology cannot be automatically applied to phenological networks, they can give important guidelines for the assessment of phenological observations.

There are several mechanisms limiting or promoting the synchronization of space-separated biological systems and, consequently, the representativeness of single-site observations. One of the most important mechanisms is large-scale forcing by meteorological and geophysical factors. From the theory of differential equations, it is known that the evolution of linear systems under external forcing will generally follow the evolution of the forcing, which works as a synchronizing agent for these systems. In population dynamics and ecology it is referred to as the Moran effect (after the pioneering study of Moran, 1953). Numerous studies (e.g., Blasius & Stone, 2000; Ripa, 2000; Sparks & Braslavská, 2001; Engen & Sæther, 2005) have shown that the Moran effect plays a crucial role in the spatial synchronization of biological systems. For the pollen and seed production of trees, Koenig & Knops (1998, 2000) demonstrated that synchronous seed reproduction over large areas in the Northern Hemisphere is caused by a common environmental fluctuation, such as rainfall and temperature.

However, the situation is more complicated and there are other processes in play. Thus, Satake & Iwasa (2000, 2002) indicated that in the case of pollen limitation, coupling of trees through pollen exchange may synchronize reproduction with a scale a few times larger than the range of direct pollen exchange, which is assumed to be limited in forest trees (Smouse & Sork, 2004). In the case of pollen coupling, the synchronization of flowering intensity could be connected to the resource allocation of trees (Satake & Iwasa, 2000, 2002). The effects of small-scale pollen-induced coupling in comparison with large-scale climate forcing were compared in the follow-up work by Satake (2004).

From an evolutionary point of view, effective pollination of anemophilous plants requires adaptations that cause scattered individuals to release pollen at the same time over large areas. For wind-pollinated Betulaceae trees, the benefits of releasing a great amount of pollen at the same time seem obvious: an exponential positive relationship exists between the amount of pollen produced, pollination efficiency and seed viability (Sarvas, 1952; Shibata et al., 1998).

One of the ultimate practical outcomes of coupling between biological systems is that observations made at any phenological station have a limited representativeness over the surrounding region. The term ‘representativeness’ reflects the uncertainty introduced by an extrapolation of data in time and/or space beyond the time period and area when/where they were obtained. Specific quantitative measures can vary depending on the application. For instance, representativeness can be quantified via the spatial correlation radius (as done in kriging analysis), via the pair-wise correlation coefficient and its dependence on the distance between the correlated points, via the standard deviation of a spatially averaged field, spatial structure functions, etc. More quantitative descriptions of representativeness can be found, for example, in the classic textbook of Yaglom (1987).

The purpose of the current paper is to provide the first quantitative assessment of spatial representativeness of phenological observations in Europe using the birch (Betula) taxon as an example. We also demonstrate how the assessments of representativeness can be used as quantitative indicators of variability of the phenological processes and their spatial and temporal scales.

The study was performed within the scope of the POLLEN project (http://pollen.fmi.fi), which has developed a grid-based numerical model combining meteorological, phenological and other types of information (Sofiev et al., 2006). We therefore focused our attention on a grid-based type analysis of the spatial station representativeness. In particular, we discuss how the local-scale or point-wise data of phenological station(s) correspond to grid-cell medians, and how the observations from single stations compare with each other if the sites are located in the same grid cell (assuming that some grid is imposed over the domain of interest).

**MATERIAL AND METHODS**

The collected phenological data base and its quality control

A main pre-requisite for the assessment of the spatial variability of plant processes over Europe is a comprehensive phenological data base. There are many national data bases in Europe, but they have not yet been combined. The ongoing COST (European Co-operation in the Field of Scientific and Technical Research) action 725 ‘Establishing a European Phenological Data Platform for Climatological Applications’ is collecting Europe-wide phenological data but has not yet produced the data base. The information for the current study was therefore collected on a country-by-country basis with subsequent conversion to a common format. The current status of the Betula data base is presented in Table 1 and the station locations are shown in Fig. 1(a).

The data base PhenoData contains observations of Betula from 15 countries from various sources. The taxonomy of Betula in Europe is disputed, but of the four species in *Flora Europaea* (Tutin et al., 1993) the observed trees are known or assumed to
Table 1 Current status and content of the phenological data base PhenoData.

<table>
<thead>
<tr>
<th>Country</th>
<th>No. of stations</th>
<th>Years</th>
<th>Species</th>
<th>Data provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belarus</td>
<td>5</td>
<td>1967–98/2002–05</td>
<td><em>Betula</em></td>
<td>University of Tartu, Institute of Geochemistry and Geophysics</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>206</td>
<td>1955–2004, a few older</td>
<td><em>Betula pendula</em></td>
<td>Czech Hydrometeorological Institute</td>
</tr>
<tr>
<td>Estonia</td>
<td>19</td>
<td>1947–2003</td>
<td><em>Betula</em></td>
<td>Estonian Meteorological and Hydrological Institute</td>
</tr>
<tr>
<td>Finland 1</td>
<td>39</td>
<td>1997–2005</td>
<td><em>Betula pendula,</em></td>
<td>Finnish Forest Research Institute</td>
</tr>
<tr>
<td>Finland 2</td>
<td>547</td>
<td>1973–2004, station specific</td>
<td><em>Betula</em></td>
<td>Finnish Society of Science and Letters</td>
</tr>
<tr>
<td>Germany</td>
<td>2119</td>
<td>1985–2004</td>
<td><em>Betula pendula</em></td>
<td>German Weather Service DWD</td>
</tr>
<tr>
<td>Latvia</td>
<td>2</td>
<td>1958–93</td>
<td><em>Betula</em></td>
<td>University of Tartu</td>
</tr>
<tr>
<td>Lithuania</td>
<td>3</td>
<td>1962–86</td>
<td><em>Betula</em></td>
<td>University of Tartu</td>
</tr>
<tr>
<td>Norway</td>
<td>1</td>
<td>1927–2004, holes</td>
<td><em>Betula pubescens</em></td>
<td>Institute of Meteorology and Water Management</td>
</tr>
<tr>
<td>Poland</td>
<td>20</td>
<td>1980–92, 2005</td>
<td><em>Betula pendula</em></td>
<td>Ecological Centre Paua, University of Tartu</td>
</tr>
<tr>
<td>Russia</td>
<td>89</td>
<td>1951–2004</td>
<td><em>Betula</em></td>
<td>Moscow State University</td>
</tr>
<tr>
<td>Slovakia</td>
<td>4</td>
<td>1986–2004</td>
<td><em>Betula pendula</em></td>
<td>Slovak Hydrometeorological Institute</td>
</tr>
<tr>
<td>Spain</td>
<td>4</td>
<td>2002–03</td>
<td><em>Betula pendula,</em></td>
<td>Galician Aerobiological Network</td>
</tr>
<tr>
<td>Switzerland</td>
<td>138</td>
<td>1996–2004</td>
<td><em>Betula pendula</em></td>
<td>MeteoSwiss</td>
</tr>
<tr>
<td>Ukraine</td>
<td>5</td>
<td>1951–98</td>
<td><em>Betula</em></td>
<td>Moscow State University</td>
</tr>
<tr>
<td>UK</td>
<td>3414</td>
<td>1999–2004, station specific</td>
<td><em>Betula pendula</em></td>
<td>UK Phenological Network</td>
</tr>
<tr>
<td>Total</td>
<td>6595</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

represent the two tree-like species, silver birch (*Betula pendula* Roth.) and downy birch (*Betula pubescens* Ehr.), both common and with natural distributions extending from the mountainous regions of southern Europe to northernmost Fennoscandia, and through Siberia to the east coast of Asia (Atkinson, 1992).

The data base includes observations of three plant parameters: date of bud burst, date of leaf unfolding and the first flowering day. The total number of the data points (all stations, all years) considered for the study was: 6215 for bud burst, 58,755 for leaf unfolding and 27,519 for the first flowering day. Most of the observations were made after 1980, but some date back to the middle of the 19th century, e.g. the data set of the Finnish Society of Science and Letters covers more than 150 years. The longest single-site time series in the analysis covers the period 1970–2005 (36 years) but some stations in the PhenoData reported for even longer.

The methodologies of the observations and definitions of the phenological phases vary somewhat from country to country (Tables 1 & 2) (Elagin & Lobanov, 1979; Kubin et al., 2004). Also the number of trees monitored and the frequency of monitoring varies, depending on the country. In several countries the birch types are not reported, which has to be treated as an internal uncertainty of the particular subsets. This is consistent with the absence of quantitative information about the geographical distribution of the birch taxa mentioned by Sofiev et al. (2006), who compiled a map of ‘general birch’ without a split into individual species.

The definition of phenological phases based on BBCH methodology (Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie, after Meier, 1997) is widely used within the agricultural sector; all participants in the COST 725 action have therefore adopted the use of BBCH methodology (http://www.cost725.org), and the European Phenological Network (EPN) recommend, that (new) phenological networks should adopt it as the basis for setting up or updating their monitoring programmes (Bruns & van Vliet, 2003). However, until now, not all national networks have adopted the Europe-wide unified definitions of the phenological phases, nor a single procedure of observations describing the number of trees to be looked at, minimal area covered, etc.

Organizational details can also affect the data. Professional networks monitor objects regularly at least twice a week. For amateur observers the frequency may differ, although during transition phases it is at least the same as for professionals. The densest networks in the UK and Germany are operated by amateurs, as well as the newly established network in the Netherlands (not in the data base). Therefore, the phenological observations by non-professionals should not be underestimated, as they provide the bulk of the currently available information in Europe.

While compiling the data base, a pre-screening was made to ensure its self-consistency, but this was reduced to a minimum in order not to disturb the data by excessive filtering. The pre-screening was done to all data, independent of the potential national practices of same type performed to data before delivery. The requirements for the final data set were: (1) it should be internally consistent, i.e. the discrepancies between the national sub-sets should be smaller than their internal variability; and (2) it should be free of crude errors due to misprints or misprocessing of the data. The pre-screening therefore included the following:
From Table 2, it can be seen that the largest discrepancy in the definitions of the phenological phases is 10% versus 50% of unfolded leaves as a criterion for leaf unfolding. Its impact can be estimated using the results of Rousi & Heinonen (2007), who studied the temporal distribution of leaf unfolding of a mixed birch stand (B. pendula and B. pubescens) at Punkaharju, Central Finland (61.8°N, 29.3°E). They found the variance in timing between individual plants to be 2.24 days. Assuming a normal distribution, we estimated that the systematic bias between the 10% and 50% criteria is about 1.9 days. This bias was already smaller than the uncertainties deriving from the limited frequency of observations (Table 2). Taking into account the other internal uncertainties in each subset (e.g. mixed birch species, unknown microclimate of each station, individual plant characteristics, etc.) the overall impact of this bias was believed to

Figure 1 (a) Locations of the phenological stations. Colours show the length of the time series for each station (years). (b) Betula leaf unfolding in 1999 (Julian days). The grid shown has a resolution of 2.250°.
Table 2 Characteristics of the monitoring system and criteria for a phenological phenomenon to take place at an observation site.

<table>
<thead>
<tr>
<th>Type of network</th>
<th>No. of trees</th>
<th>Frequency of monitoring</th>
<th>BBCH used</th>
<th>Leaf unfolding (BBCH 11)</th>
<th>Start of flowering (BBCH 60); when takes place at one site?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belarus 1 (Aeroteam)</td>
<td>Professional</td>
<td>1</td>
<td>Every day/second day during transition phases</td>
<td>No</td>
<td>Leaves at several sites of the object have unfolded</td>
</tr>
<tr>
<td>Belarus 2</td>
<td>Professional</td>
<td>1</td>
<td>2 to 3 times per week</td>
<td>No</td>
<td>10% of leaves unfolded</td>
</tr>
<tr>
<td>Estonia</td>
<td>Professional</td>
<td>n/a</td>
<td>2 to 3 times per week</td>
<td>No</td>
<td>10% of leaves unfolded</td>
</tr>
<tr>
<td>Finland 1</td>
<td>Professional</td>
<td>5</td>
<td>2 to 3 times per week</td>
<td>No</td>
<td>50% of leaves of 5 trees unfolded</td>
</tr>
<tr>
<td>Finland 2</td>
<td>Amateur</td>
<td>Daily</td>
<td>No</td>
<td>No information</td>
<td>No data</td>
</tr>
<tr>
<td>Germany</td>
<td>Amateur</td>
<td>1</td>
<td>Own consideration, but frequently during transition phases</td>
<td>Yes</td>
<td>Leaves have unfolded at least at three sites of the tree</td>
</tr>
<tr>
<td>Latvia</td>
<td>Professional</td>
<td>5</td>
<td>2 to 3 times per week</td>
<td>No</td>
<td>10% of leaves unfolded</td>
</tr>
<tr>
<td>Lithuania</td>
<td>Professional</td>
<td>5</td>
<td>2 to 3 times per week</td>
<td>No</td>
<td>10% of leaves unfolded</td>
</tr>
<tr>
<td>Russia 1</td>
<td>Professional</td>
<td>5</td>
<td>2 to 3 times per week</td>
<td>No</td>
<td>10% of leaves unfolded</td>
</tr>
<tr>
<td>Slovakia</td>
<td>Professional</td>
<td>1</td>
<td>Several times a week during transition phases</td>
<td>Yes</td>
<td>Leaves have unfolded at least at three sites of the tree</td>
</tr>
<tr>
<td>Spain</td>
<td>Professional</td>
<td>1</td>
<td>Twice a week</td>
<td>Yes</td>
<td>Leaves have unfolded at least at three sites of the tree</td>
</tr>
<tr>
<td>Switzerland</td>
<td>Amateur</td>
<td>1</td>
<td>2 to 3 times per week</td>
<td>Partly</td>
<td>50% of leaves have unfolded</td>
</tr>
<tr>
<td>Czech Rep.</td>
<td>Professional</td>
<td>3 to 5</td>
<td>Every second day</td>
<td>No, but similar</td>
<td>Mode of days the phase has appeared in 10% of leaves of individual trees</td>
</tr>
<tr>
<td>UK</td>
<td>Amateur</td>
<td>1</td>
<td>Own consideration, but frequently during transition phases</td>
<td>Yes</td>
<td>Leaves have unfolded at least at three sites of the tree</td>
</tr>
</tbody>
</table>

BBCH number refers to BBCH methodology number (Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie, after Meier, 1997).
be small, for which reason we did not apply any bias correction in the current study.

We analysed the widest possible range of spatial scales. The density of stations in several countries allowed for consideration of a wide range of spatial scales down to a few kilometres (Fig. 1a). There was also one data set from Spain with tree-specific information given for several trees at a single station, which corresponded to a spatial scale of a few tens of metres and reflects the variability in the behaviour of a single tree. The largest-scale consideration covered the whole of Europe.

Factors influencing the representativeness of a single phenological station

There are three main sources of objective uncertainty in the determination of a phenological phase at a station and its extrapolation to the surrounding region: local microclimate, meteorological variability and plant-specific variability. Non-ideal observations also contribute to the overall uncertainty: a non-daily monitoring frequency, a mixture of different taxa, subjective inaccuracies in determining the phases, etc. However, such irregularities in the data gathering and processing, being subjective, are essentially irreducible and thus have to be treated as an underlying basic source of uncertainties common to all the data.

The first contributor is the local microclimate. The immediate surroundings of the site, such as the southward slope of a hill or the proximity of a lake to the forest, play an important role in the phenological stages. The local microclimate has a very small characteristic spatial scale – well below 1 km. Its impact at larger scales can be considered random in space, but the bias in the dates of the phenological stages observed at the station is stable in comparison with the grid averages. Indeed, the effect of, for example, the northern slope of a valley is stable from year to year. Its influence can be somewhat reduced if observations at a station cover at least a few hundred metres (e.g. a botanical garden or a forest). It can also be treated during data preparation, for example, as by Hakkinen et al. (1995) who have combined several point-wise phenological observations into a single regional time series by adjusting observations with a station-specific, temporally fixed bias correction.

The second contributor is the meteorological variability. Presumably, it should be more important at larger scales, because plant processes have a considerable 'memory' of past weather conditions and thus are not sensitive to small-scale short-term events, such as a single rainfall event (Sarvas, 1955, p. 21). In contrast to microclimate, meteorological variability does not create a temporally stable bias for a specific station but is rather seen as random fluctuations in the dates of the phenological phases from year to year. These fluctuations should be spatially synchronized over synoptic-scale areas (at least a few tens of kilometres) and will thus disappear when the data are treated at a higher spatial resolution. In the analysis, that should be seen as a reduction of the variability with increasing resolution.

The contribution of the third source of uncertainty – the plant-specific variability (Rousi & Pusenius, 2005) – is practically indistinguishable from the impact of observational errors, differences in methodology and practice (at both country and station levels), etc. The contribution of all these factors to the overall uncertainty should be largely random (except for some specific aspects of methodology, which, as shown above, are believed to be small) in both time and space. In a few cases (e.g. frequency of observations) some part of it can be quantified, but in most cases only an overall estimate can be obtained via the procedures described below.

Importantly, the meteorological factors play both synchronization and de-synchronization roles depending on the spatial scale. They are one of the sources of uncertainties at large scales when the synoptic structures (cyclones, fronts, etc.) are not resolved. For smaller scales the Moran effect starts to dominate; plants that are under the same meteorological stress respond in a similar way, their phenological stages thus becoming closer to each other.

Quantitative characterization of the uncertainty and delineation of its sources

The methodology described below is largely based on grids with varying resolutions defined for the domain, onto which grids the stations were projected directly, without any interpolation between the sites or other gridding techniques (see Appendix S1 in Supplementary Material for the details).

In many cases (especially at a high spatial resolution) the number of stations falling into a single grid cell was small. As a protective measure against outliers, we always used statistically robust measures, such as the median and percentiles, instead of sensitive parameters, such as the arithmetic average and variance. We also required at least five stations to fall into a grid cell for it to be included in the analysis. For the analysis of temporal variability, an additional requirement of at least 5 years of common reporting period was made.

Three types of statistical measures were selected for the analysis:

Grid-based estimate of spatial uncertainty

The spatial variability \( V \) of the date of a phenological phase within a grid cell was defined as: \( V = P_{16} - P_{84} \) where \( P_{16} \) and \( P_{84} \) are the 84th and 16th percentiles, respectively, of the single observations made by the stations located within this grid cell. The variability \( V \) was thus a robust estimate of the double-standard-deviation \( 2\sigma \) uncertainty interval, which contains about 64% of the observations. For example, for seven observations in a grid cell, \( V \) was the difference between the second earliest and the second latest observed dates. To reveal the scale dependence of the variability, we considered several grids with resolution varying from a few tens of metres up to a few thousand kilometres.

Grid-based estimate of temporal uncertainty

The temporal stability of the data of a single station was estimated as the probability for the station to report dates earlier/close-to/ later than the median over the corresponding grid cell. A single observation was classified as ‘early’ (or ‘late’) if its deviation from
the grid-cell median date exceeded 2 days. This threshold was selected to be close to a typical synoptic time-scale of 3 days. For each station, we computed in how many cases the station reported the date within the median ± 2 days, later than 2 days or earlier than 2 days from the median. Finally, we marked the station as ‘early/late’ if more than 70% of its observations were ‘early’ or ‘late’, respectively. The sites with more than 70% of dates reported within 2 days of the median were labelled as ‘representative’. Stations not counted in any of the above groups were considered as ‘random’ as they did not demonstrate any regularity in year-to-year behaviour regarding the grid-cell median. For example, if the station was within 2 days of the median in 50% of years, ‘late’ in 25% and ‘early’ in 25% of cases, it was called ‘random’.

Grid-free spatial uncertainty analysis via structure functions

To support and double-check the grid-based analysis, the method of structure functions was brought into use as a complementary tool. It is an independent method for the analysis of space-distributed stochastic fields, which is neither based on any imposed regular grids nor involves any other regularization techniques. It is widely used for the analysis of meteorological fields and observations. Its formal definition and main features are explained in Appendix S2. From a physical point of view, the structure function \( S_f(r_1, r_2) \) of a field \( f \) quantifies the decorrelation of the subregions of the field \( r_1 \) and \( r_2 \) depending on the distance between these sub-regions \( r_1 - r_2 \). The higher \( S_f(r_1, r_2) \), the more independent these regions are. Vice versa, if \( S_f = 0 \) then the processes in these regions perfectly coincide. In our case, the \( S_f \) of the phenological phase \( p \) showed decorrelation of the observing stations depending on the distance between them. The ‘distance’ here also included the direction from one station to another, because decorrelation along longitude was evidently different from that along latitude.

RESULTS OF THE REPRESENTATIVENESS ANALYSIS AND DISCUSSION

General characteristics of spatial variability

The key result of the study (Fig. 2) was that the variability \( V \) for practically all resolutions was between 7 and 16 days (i.e. the standard deviation \( \sigma \) varied from 3 to 8 days) and behaved quite similarly for all considered phenological parameters. The variability tended to decrease at higher resolutions but this trend was not strong: (1) for very high-resolution grids with a cell size from < 1 km to 10–20 km, the increasing resolution reduced the uncertainties in the first flowering day but seemingly did not affect the leaf unfolding; (2) for a wide range of meso-to-regional (20–500 km) scales no clear dependency of variability on resolution existed; and (3) extra-coarse resolution – over thousands of kilometres – lead to a clear increase in the variability.

Analysis via the structure functions for leaf unfolding (Fig. 3) did not include the subkilometre range because of an insufficient number of data. The rest of the spectrum was very well seen, providing more insight into the above observation and refining some of the conclusions. As the separation distance (a direct analogy of grid resolution) approached 2 km (the minimum computed distance), the unified European structure function \( S_f \) (built using the whole data set) approached the limit of 5–6 days. With increasing scale, it grew to 6.5 days for ~50 km and then continued to grow slowly to reach 7 days for a resolution of about 500 km. At this scale – but not earlier – the difference between north–south and east–west directions became visible, so that the decorrelation of stations located far from each other in a north–south direction grew faster (an evident footprint of a different climate). A further scale increase resulted in widening the variability, which grew fastest across climatic zones, i.e. in a north–south direction.

Structure functions built for individual regions revealed further peculiarities of site representativeness in each part of Europe. Thus, the maximum decorrelation between sites at all distances was observed in the UK. There, even co-located sites (at a distance of ~2–4 km) showed a variability between 6.5 and 7.5 days. For larger scales, it grew faster than in other regions and reached 9 days for a scale of 500 km. After that, the analysis became inaccurate due to the limited size of the UK.

The decorrelation of German sites largely resembled that of the whole of Europe, but was slightly lower – by about 1 day.
Structure functions built from Finnish sites were noisier due to the limited number of stations, but still revealed the main feature of the region; the sites were correlated much more strongly than in the other considered areas. Closely located stations tended to deviate by less than 4 days with a very slow growth towards 5 days for a scale of 300 km. Larger scales became too noisy due to the limited size of the domain; the only conclusion that could be drawn was that the north–south separation of the sites already plays an important role for distances of about 200–300 km.

In all regions, the finest scales partially resolved the local microclimate structures. The main irreducible variability for such scales was therefore the individual fluctuations of the plants and observation-specific noise. The seemingly different behaviour of leaf-unfolding data (Fig. 2a showed no reduction of uncertainty for subkilometre scales) may have been artificial because the variability estimate for the highest resolution was computed from the data for just four stations in Spain, which provided information for single trees. With all the specifics of birch in Spain, this four-site data set was clearly insufficient for deriving general conclusions on the small-scale variability of birch phenophases. Conclusions based on structure functions seemed to be more reliable – and clearly showed a decrease of variability with increasing resolution.

Averaging over larger spatial scales (20–500 km) smoothed out the small-scale fluctuations and individual specifics of the stations. For all these scales, the microclimate was unresolved, while the meteorological processes did not contribute to the variability because their scales are larger. From the opposite point of view, the Moran effect should have been in full force here, eventually synchronizing the dates from different stations. As a result, the dependence of variability and resolution was broken and the variability became nearly constant (10–12 days for leaf unfolding and 14–16 days for the first flowering day) up to a grid cell size of ~500 km.

Very coarse grids with averaging over more than 1000 km almost always covered several climatic zones with a single grid cell, a fact that contributes to the subgrid variability. For structure functions (Fig. 3), this range corresponds to the right-hand part of the curves where the variability depended on direction, showing the strongest growth in the cross-climatic north–south direction. The variability of the first flowering day (Fig. 2b) showed a moderate increase because of the relatively high contribution of short-term meteorological fluctuations to the underlying phenological processes. The leaf unfolding (Fig. 2a) variability, or the contrary, increased strongly to ~15 days for a grid cell size of ~1000 km. This corresponded to the typical synoptic (or large) scale in meteorology (500–5000 km and 1–5 days).

Figure 3 Structure functions for leaf unfolding in Europe, the UK, Germany and Finland. The series represents the sectors with regard to longitudinal direction (0º is along a west–east parallel). Structure function reflects the variability in data between two points located at given distance along given direction. The strongest growth is in the cross-climate north–south direction. See Appendix S2 for methodological details.
The above results look very discouraging due to the very large variability between the sites at all scales and in almost all regions. In some years the difference in bud burst, leaf unfolding or first flowering date between the earliest and the latest birch individuals standing in the same garden could be as large as 1 month — as was observed at one station in Spain (42.3° N, 7.5° W).

Influence of local micro-climate and topography

The second part of the study was dedicated to a separation of the influence of local microclimate and topography from that of meteorological processes. This analysis required long time series and a dense network, and was thus almost solely based on German observations. The working hypothesis was that the local microclimate should create a spatially random but temporally synchronized bias for the specific station whereas the influence of meteorological processes would be much more synchronized in space and random in time. Indeed, a microclimate caused by local geographical specifics, such as a hilly surface with northern–southern slope differences or a freezing lake that takes long to thaw in spring, would affect the trees every year in a similar way. Meteorological effects, such as a late, rainy spring, would affect large regions synchronously, but would vary strongly from year to year. The microscale noise from the tree specifics and imperfect observations were independent additional components to the above, and were assumed to have similar features for all stations within a specific grid cell.

The above two factors could be separated by an analysis of the ability of single stations to follow the regional averages year by year (see Appendix S1 for details). For areas having a sufficient network density and a long observational history, it appeared possible to label each station in accordance with its behaviour – ‘early’, ‘representative’, ‘late’ or ‘random’ – and to perform this labelling for all grid resolutions. The expectation was that the fraction of ‘representative’ stations increases for a finer resolution. This trend would reveal the impact of microclimate. For example, if some station was ‘late’ at coarse resolutions but became ‘representative’ for a sufficiently small grid cell size Δx, this station was affected by a local microclimate with a characteristic scale Δx.

The result of the analysis – the fraction of stations in each category in relation to the grid resolution – is shown in Fig. 4(a,b). As expected, the smaller grid cell sizes corresponded to a higher fraction of ‘representative’ stations because the corresponding microclimate processes were resolved by these grids. The number of ‘early’ and ‘late’ stations decreased proportionally, which also pointed to the local microclimate specifics as the primary reason for such systematic bias. The number of ‘random’ stations did not change, indicating that their uncertainties were not connected with spatially regular phenomena. Such stations tended to deviate randomly from any neighbouring site, regardless of the distance between them.

The numbers of ‘early’ and ‘late’ stations were nearly equal to each other for all resolutions, with a slightly larger fraction of ‘late’ sites. This additionally illustrates an asymmetry in the temporal distribution of phenological phases: the probability of long delays is higher than the chance of starting a phase very early.

Figure 4 Fractions of ‘early’, ‘late’, ‘representative’ and ‘random’ stations in relation to the grid cell mean as a function of resolution: (a) leaf unfolding, (b) first flowering day.

To evaluate the sensitivity of the analysis to the above thresholds in the station classification, we repeated the analysis with 3 days being taken as the threshold for a ‘large’ deviation of an observation from the grid median (rather than 2 days as explained in Appendix S1 and used above). For the small grid sizes, the number of ‘representative’ stations increased from under 30% (for ± 2 days) to over 40% (for ± 3 days) and even at a grid size of 11.25° there were still over 10% of stations that could be designated as ‘representative’ (compared with 2% for a threshold of ± 2 days). The fractions of ‘late’ and ‘early’ stations did not differ markedly although were decreasing. However, the fraction of ‘random’ sites dropped to 50% for high-resolution grids (vs. 62% for ± 2 days threshold). When the grid size was larger than 2.25°, the differences were negligible.

The other sensitivity test was to call a station ‘representative’ with just 50% of the observations falling close (± 2 days) to the grid cell median — instead of the 70% used above. The fraction of ‘random’ stations then dropped to 25%, and tended to decrease towards large grid cell sizes. The fractions of the other classes changed accordingly — the ‘representative’ class increased while the ‘early’ and ‘late’ classes became thinner. For example, for a high-resolution grid cell size of 0.187° all four fractions were.
close to 25%. Their dependence on resolution was similar to that of the base case.

**Regional peculiarities of the variability**

Analysing structure functions, we have already seen that the largest uncertainty was in the UK, while stations in the northern regions were the most representative. As an additional illustration of the regional specifics and spatial trends of the variability, Figure 5 presents examples of maps of median dates of leaf unfolding and their variability for the reference grid having a 1.125° resolution. A comparison of Fig. 5 with the structure functions in Fig. 3 confirms the tendency towards a much higher uncertainty in the south-west of the studied region – in addition to the evident

**Figure 5** Maps of the median date (Julian days) of *Betula* leaf unfolding and its variability (days) for the ERA-40 grid, (1.125° × 1.125° resolution) in 1999, 2003 and mean 1970–2003. Variability is computed for grid cells with more than three stations.
The mechanisms of synchronization of the phenophases in the northern regions are probably somewhat different from those in the continental climate of Russia, and have a stronger connection with the shorter vegetation period and risk of late-spring and early-autumn frost damage in addition to a faster rise of temperature in spring.

**Influence of uncertainties in the phenological data on their applications**

The analysis of spatial representativeness of the phenological sites presented here follows the general procedure of pre-evaluation of observational data used in numerical modelling. The key assumption imposed on the data when using them in such grid-based systems is that the variations within the grid cells (the non-resolved part of a phenomenon) are much smaller than the simulated part of the signal. Similar assumptions are generally true for all of applications; the noise in the data must be small compared with the signal.

The uncertainties discussed above, however, are quite large. Even for a grid size of 30 km, the variability is from 6–14 days, depending on the region, with a significant reduction seen only in continental and northern climates. These correspond to a standard deviation of from 3–7 days. Depending on the application and strength of the signal, such variability can be acceptable or not. In the latter case, certain measures are needed before the data become usable.

A primary goal of the current study was to support the POLLEN project by evaluating the possibility and suggesting a methodology for using the phenological data for the pollen release model within the scope of an atmospheric dispersion model. According to experience with similar model types, the subgrid variability should not exceed the synoptic time scale of ~3 days. This means that the uncertainty in the starting date of pollen emission should not exceed 2 days (the above threshold for ‘representative’ observations). The large objective uncertainties in the phenological data therefore require special measures before these observations can be used for pollen forecasting (Estrella et al., 2006). For another popular use of phenological data – climatological studies (e.g., Chuine et al., 2000; Sparks et al., 2000; Ahas et al. 2002; Menzel et al., 2003; Studer et al., 2005) – the variability also seems to be high, because the expected signal is just a few (1–3) days per decade, which is again comparable with the variability itself, and makes the conclusions vulnerable to the limited representativeness of the station data, especially for ‘random’ stations.

Seemingly the simplest way is to filter out the ‘random’ stations and to correct the bias in the ‘early’ and ‘late’ ones. However, this leads to a substantial reduction of the volume of the data set and strong changes in its features. It may be acceptable in some
applications and entirely wrong in others. In our view, the variability discovered above is objective and is mainly natural in origin rather than originating from observational artefacts, which certainly contribute but do not dominate. If this is true, the filtered data set will no longer reproduce reality; it will become nice-looking, but in many senses useless.

An alternative approach might be to switch from a deterministic approach, which requires a minimum of internal uncertainty in the input data, to a probabilistic one, which takes this uncertainty into account. This can be comparatively straightforward in pollen dispersion modelling but more difficult in some other applications. Numerical weather predictions and air-quality forecasts to an increasing extent already use probabilistic ensemble forecasts (e.g. Molteni et al., 1996; Galmarini et al., 2004a, b) and it would be interesting to apply a similar treatment to phenological time series.

One of the important problems arising from the high uncertainty of phenological observations is that their treatment requires a large abundance of data to provide reliable results. Therefore high-density networks with the longest possible time series become a matter of the utmost importance. In most cases, such networks can be maintained only with the help of amateur observers, which in turn raises the problem of the quality assurance of the data, training for the observers, etc.

CONCLUSIONS

A unique phenological data base has been compiled for birch species, bringing together long-term observations from 15 countries over the European continent. A crude pre-screening of the data base has been performed and has confirmed the usability of the data set after a minor reduction. The data were used for quantifying the spatial representativeness of phenological observations.

The analysis highlighted the problem of the limited spatial and temporal representativeness of the phenological data, and showed that representativeness may strongly affect their usage and require treatment appropriate for the application in hand.

The variability of a single observation expressed as the difference between the 16th and 84th percentiles (a robust estimate, corresponding to double the standard deviation for a normal distribution) with regard to the regional median date appeared to range from 6 to 16 days depending on the size of the region and location. The uncertainty was generally smaller for regions of smaller size (corresponding to a higher resolution.).

The representativeness of the eastern and northern sites appeared to be substantially higher than that of those in the western part of the domain. Sites located in mountains normally had poor representativeness.

A reasonable explanation for the observed dependence of the representativeness on the geographical location relies on the Moran effect due to meteorological forcing. Shorter springs in the continental and northern climates and shorter vegetation periods with a high frost risk would generally force the plants to act more synchronously, following the cause of the weather developments. However, this matter requires further investigation of the biological mechanisms behind it and cannot be confirmed within the scope of the current study.

Station-specific analysis showed that only 10–30% of sites (depending on the grid resolution) stably report dates close to the regional averages (within 2 or 3 days). Another 20–40% of the sites are either stably ‘early’ or ‘late’ in comparison with the regional median. Practically regardless of the spatial resolution, a large fraction of the sites (~60% for a threshold of 2 days and 25% for 3 days) fluctuate widely and randomly around the regional mean dates.

The study showed that direct utilization of the phenological observations for constructing a pollen atmospheric dispersion model is not feasible. Special measures should be implemented in the emission module to reflect the objective variability of the dates of the phenological phases.

These findings also call for appropriate treatment of the phenological data in other studies relying on the spatial and temporal representativeness of the data, most of all in climate research.

Methods of pre-processing the data depend on the specific application, but could include filtration of noisy parts of the data set (a potentially dangerous action disturbing the data features), or a switch to probabilistic description of the phenomena.

ACKNOWLEDGEMENTS

The current study is part of the Academy of Finland POLLEN project on ‘Evaluation and forecasting of atmospheric concentrations of allergenic pollen in Europe’ (http://pollen.fmi.fi).

The phenological data were provided by: Belarus Aerobiology team, Meteo Swiss (Switzerland), Czech Hydrometeorological Institute, German Weather Service DWD, Estonian Meteorological and Hydrometeorological Institute EMHI, Tartu University (Estonia), Moscow State University (Russia), Galician Aerobiological Network (Spain), Finnish Forest Research Institute, The Finnish Society of Sciences and Letters, UK Phenological Network, Centre for Ecology and Hydrology (UK), Plante Forsk (Norway), Institute of Meteorology and Water Management (Poland) and the Slovak Hydrometeorological Institute.

REFERENCES


P. Siljamo et al.


SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Appendix S1 Details of grid definition and processing.

Appendix S2 Definition, main features of structure functions and their application to phenological analysis.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1466-8238.2008.00383.x

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

BIOSKETCHES

Pilvi Siljamo is a researcher (meteorology) at the Finnish Meteorological Institute. Specific areas of interest are atmospheric dispersion models, modelling pollen long-range transport and dispersion calculation during emergency-type short-term releases of various air pollutants.

Mikhail Sofiev is a physicist and adjunct professor at the Finnish Meteorological Institute. Specific areas of interest are mathematical modelling of atmospheric pollution by various compounds, model verification and comparison with measurements, statistical methodology for model evaluation, analysis of the measurement data, inverse and adjoint dispersion modelling.

Hanna Ranta is a lecturer and aerobiologist at the University of Turku. Specific areas of interest are pollen research, development, evaluation and application of pollen long-range transport modelling and official pollen forecasts in Finland.

Tapio Linkosalo is a phenologist at the University of Helsinki. Specific areas of interest are studying the annual rhythm of boreal forests and the impact of climatic warming on it, phenology of boreal forests and applied mathematics.

Editor: Matt McGlone