TITLE PAGE

Digitizing the plant phenological dataset (1750-1875) from collections of Professor Adolf Moberg: Towards the development of historical climate records

Jari Holopainen[a], Samuli Helama[b]* and Henry Väre[c]

[a]Department of Geographical and Historical Studies, University of Eastern Finland, 80100 Joensuu Finland; +358 (0)294 45 1111; jari.holopainen@uef.fi

[b]Natural Resources Institute Finland, Rovaniemi, 96300 Finland; +358 (0)29 532 6000; samuli.helama@luke.fi

[c]Botanical Museum, Finnish Museum of Natural History, 00014 University of Helsinki, Finland; +358 (0) 2941 911; henry.vare@helsinki.fi

*Corresponding author: samuli.helama@luke.fi

ABSTRACT

Long records of phenological observations are commonly used as data in global change and palaeoclimate research and to analyse plants' responses to climatic changes. Here we delve into the historical archives of plant phenological observations (1750-1875) compiled and published previously by Professor Adolf Moberg (Imperial Alexander University of Finland). The digitized dataset represents 44487 observations of 450 different plant species for their 15 different phenological phases made in 193 sites across Finland, and results in 662 different phenological variables. The five most frequently observed variables are the blooming of rye, the sowing of barley, the blooming of bird cherry, the leaf outbreak of birch, and the sowing of oat. The spring and summer observations demonstrate positive relationships between the onset date and the site latitude, this relationship becoming negative for observations made in the autumn. This latitudinal effect is evident in the raw data as demonstrated by the temporal correlations between the unadjusted mean phenological records and the mean latitude of the sites. After the latitudinal effect is removed from the original data such correlations are much reduced and the new set of phenological records based on the adjusted dates can be computed. The resulting mean phenological records correlate negatively and statistically significantly with the mean temperatures from April through July. Linear trends indicate (i) summer onsets having become delayed by more than one week over the full period and (ii) shortening of the growing seasons since 1846. The data is made available in an open repository.

Key words: biogeography, historical climatology, indications of spring, phenology, plant–climate interactions, temperature

1. Introduction

Phenological datasets consist of observations made on natural seasonal events. The changes in the annual cycle of plants are closely linked to the seasonal course of temperature and water availability and the phenological data constitute first-hand evidence of plants' responses to these changes. First, these data contribute strongly to our understanding of global change biology (Menzel 2002). Importantly, the plant phenological data have demonstrated an average advance of spring/summer 2.5 days per decade within European countries, in accordance with instrumentally observed warming since 1970s over the same region (Menzel et al. 2006). Yet, the long records help to identify the ways the enhanced warming may alter the climatic drivers of phenological phases by changing the relationships between the seasonal courses of temperature, moisture, and plant development (Cook & Wolkovich 2016). Second, the long phenological data provide the palaeoclimate reconstructions with written records of past climate variability (Craddock 1974; Kington 1974). Phenological data of various plant species have indeed been used to reconstruct temperature variations over the past centuries in order to assess the character of recent climate regime (Chuine et al. 2004; Rutishauser et al. 2007; Holopainen et al. 2009). Compared to other natural proxy archives such as pollen and sedimentary records, phenological records benefit from being time-series directly comparable with meteorological records without chronological uncertainties.

Longest of the phenological records originate from notes and diaries of early naturalists and enthusiastic volunteers (Margary 1925; Lappalainen & Heikinheimo 1992; Holopainen et al. 2012). These data make it possible extending the phenological records over the 19th and 18th centuries and evaluation of their trends and climatic signals on decadal to centennial time intervals and scales (Sparks & Carey 1995; Holopainen et al. 2006, 2013; Rutishauser et al. 2009). Such documents may have survived over decades and centuries in institutional or personal archives (Margary 1925; Chuine et al. 2004; Holopainen et al. 2013) or they may have been published as yearbooks or larger collections of data by contemporary scholars soon after the initial observations had been recorded. In Finland, the collection of plant phenological observations started already in 1750s, following the general recommendations set by Carl von Linné (Terhivuo et al. 2009), and has continued since then as a more or less coherent national effort by a number of universities and societies (Holopainen et al. 2012). In practice, the phenological observations have been made by hundreds of volunteers collectively contributing to the effort by returning their formal notebooks to the organizations that coordinate the process of data collection and maintain the repositories of such materials. During the early days of this process, an enormous effort of his own was made by Professor Adolf Moberg (1813–1895), the Imperial Alexander University of Finland, who compiled a large amount of original 18th and 19th century notes into what became four volumes of books published in Swedish. These volumes present the phenological data collected in Finland until then in well-organized format, as tabular lists of various natural events and their seasonal timing, the data originally obtained from a number of individual contributors across the country (Moberg 1857, 1860, 1885, 1894).

Here we present this data after having digitized the information from the books of Moberg (1857, 1860, 1885, 1894). We concentrate on plant phenological observations and demonstrate the characteristics of this data as available between the years 1750 and 1875. This data has not been largely studied. The dates of rye harvests from three sites were previously used, along with much larger collection of Estonian proxy data, to reconstruct spring and summer temperature variability over the past centuries (Tarand & Kuiv 1994; Tarand & Nordli 2001). The flowering dates of two tree species, the rowan and the bird cherry were analysed since 1750s (Terhivuo et al. 2009). Other analyses have used the dates of flowering and leaf bud burst of birch to statistically demonstrate their usefulness to detect the signals of climate variability and warming in the region since 1846 (Linkosalo et al. 2009; Hari et al. 2017). Moreover, a limited portion of this data has been used for indicating (Holopainen et al. 2006, 2013) and reconstructing past variability in spring temperatures, along with other proxy data, in south-west Finland since 1750s (Holopainen et al. 2009). However, the dataset has not yet been studied in full. Our analysis classifies the data by plant species and their phenological phases recorded, and makes a geographical presentation of the available sites and thus of a spatiotemporal coverage of the dataset. Although not covering the 20th and 21st century era of warming climate, the data has value to climate and plant scientists for assessing the variability in plant responses to climatic variations over the pre-industrial era. Moreover, the interval of available data (1750-1875) overlaps with the period of climate anomalies attributable to the Little Ice Age conditions, inferred as a climatic cooling between 1570 and 1900 from many palaeoclimate records around the Northern Hemisphere and especially the North Atlantic/European sector (Bradley and Jones 1993; Matthews & Briffa 2005). In terms of palaeoclimatology, this data (Moberg 1857, 1860, 1885, 1894) will contribute to high-resolution assessments of seasonal climate variability over the period when the regular observations of meteorological phenomena were only scarcely made in the region and over which the proxy data as indicators of climate variability are therefore needed (Holopainen 2006).

2. Material and methods

Plant phenological data was digitized from the books of Moberg (1857, 1860, 1885, 1894) by manually typing the information into electronic format and saved in Microsoft Excel. Each observation was characterized by the plant species given in Latin, the phenological phase (e.g. budburst, flowering), the year, month, and the day of the month, as well as the site name and its geographical coordinates. In the case of agrophenological observations, the phase was often related to seasonal human activity (e.g. sowing, haymaking). These are the data of which collection was coordinated by the Royal Academy of Turku, the Finnish Economic Society, the Pro Natura Society and the Finnish Society of Sciences and Letters. This coordination included more or less the manufacturing and sending of the official cards, on which the phenological observations were formally written down by the volunteers, as well as the reception of the completed cards and their administration thereafter. Below, the issues that need to be addressed to create a homogenized database are shortly described.

2.1 The two calendars

The Gregorian calendar we use today was preceded by the Julian calendar that was, however, subject to the vernal equinoctial drift in the calendar (Dutka 1988). The former calendar was introduced in 1582 but it was not widely used until later time. In this study region, the transition took place in 1753 when February 17th (Julian) was followed by March 1st (Gregorian). Another part of the revision dealt with centurial years not divisible by 400 that became ordinary years of 365 days. As a consequence, the phenological observations made before the date of transition need to be realigned to Gregorian calendar. Accurate conversion can be done using the available tables (Kerzhner 1984) or using the equations tailored for the purpose (Hatcher 1984).

2.2. A new meridian

It was not until 1884 when the International Meridian Conference established the Greenwich meridian as the initial meridian for longitude (Sadler 1978). Before that the Ferro Meridian (18°W) was commonly used (referring to the westernmost point of the once-known world, El Hierro in Canary Islands) as the prime meridian in many countries in continental Europe as it was also employed in the three volumes of Moberg (1857, 1860, 1885). The new Greenwich meridian was adopted not until the last book of Moberg (1894). This Ferro-Greenwich shift (17°40’) was done consistently for the longitudinal coordinates published in the earlier volumes (Moberg 1857, 1860, 1885).

2.3 Site coordinates

The approach of reporting the site coordinates (whether given respective to the Ferro or Greenwich meridian) appears to vary. That is, the coordinates may be expressed either as degrees and minutes or as a fraction of degree, even within the same book. For the site of Kemi, as an example, such reporting would either result as 65°49’N and 24°32’E, or, 65¾°N and 24½°E (Moberg 1894). Of note, the coordinates reported as degrees and minutes are more accurate and preferred hereafter. For calculatory purposes, we further transformed this information into decimal coordinates.

2.4 Plant species

All the scientific names of species included in the original data (Moberg 1857, 1860, 1885, 1894) were consistently checked for spelling and changed when the names used in data did not match with the modern nomenclature. The scientific names of species were used where possible but in some cases it was not possible to ascertain which species was actually recorded, as there may be changes in the names since the time of original publications. For example, the rejected name *Betula alba* (recorded 1317 times in the original data) does not differentiate between *B. pubescens* and *B. pendula*. Other families for which the name of the species could not in every case be determined included *Crataegus*, *Galium*, *Pisum*, *Rhinanthus* and *Ulmus*. This work followed the nomenclature given in Hämet-Ahti et al. (1998). With these regards, the species were also classified as weeds, annual/biennial, archaeophytic, fruit-garden, ornamental, agricultural and woody plants.

2.5 Phenological phase

The descriptions of the phenological phases, originally given by Moberg (1857, 1860, 1885, 1894) in Swedish, have been translated here into English. The phenological phase of ‘början’, the start, was attributed to ’ängslåttern’, the haymaking, with no scientific names of species. We note that this event almost certainly refers to *Poa pratensis* and use this species in the following analyses. This species was not otherwise observed within the dataset.

2.6 The vernal equinox

An important detail of recording the timing of phenological events, especially in the case of long series of data, relates to the mismatch between the length of the solar year and the slightly longer average year on the Gregorian calendar, as originally noted by (Sagarin & Micheli 2001). In phenological data, this bias causes an overestimation of trends toward earlier spring signals but can be corrected by adopting the dates of phenological observations in relation to beginning of astronomical spring (vernal equinox) rather than by calendar day (Sagarin 2001, 2009). Here we overcome this potential bias by reporting the dates of phenological events using both the Gregorian calendar dates (in practice, the number of days elapsed since March 1st) and as the number of days elapsed since the vernal equinox i.e. the date the sun crosses the celestial equator from the austral to the boreal hemisphere that have varied between March 19th and 21st (Gregorian).

2.7 Mean phenological records

The phenological data of Moberg (1857, 1860, 1885, 1894) originates from a high number of sites across the country. Finland is located approximately between 60°N and 70°N and the data comes with a latitudinal gradient of more than one thousand kilometres. To assess the latitudinal effect in the observations, the observed dates of each phenological variable were correlated and plotted with the site latitudes. Linear regression was used to determine the slope of change (s ; days per latitudinal degree) in the dates as a function of latitude. Having determined this relationship, the latitudinal effect could be removed from the data of original observations by adding or subtracting a number of days needed to adjust the original date (D) into a pre-determined level as follows

$$D' = D - s(L - \overline{L}) \tag{1}$$

where D' is the new, adjusted date observed at the site with latitude L, and $\overline{L}$ is the mean latitude of all of the observations (here, 62.01°N). Once the latitudinal effect was removed from the original dates, the mean phenological records could be calculated by averaging all available values of D' for each year. This was done for monthly stratified data i.e. the most frequent calendar month during which the onset of that variable was originally observed was determined and the phenological data (P) representing each month (April = 4, May = 5 … October = 10) was averaged into seven different mean phenological records P_4, P_5 … P_{10}. This approach follows that of Menzel et al. (2005) who divided their phenological variables based on nine seasons before comparing the resulting subsets of data with climatic factors. In addition to visual comparisons, we have used Pearson correlations for statistically examining the resulting mean phenological records. We estimated the trends in P_4 through P_{10} records by fitting a linear regression model to their data over the full (1750-1875) and late (1846-1875) periods.

2.8 Climatic comparisons

Mean phenological records were correlated to mean monthly temperature series from Tornedalen (Klingbjer & Moberg 2003), St. Petersburg (Jones & Lister 2002), Uppsala and Stockholm (Moberg & Bergström 1996; Moberg et al. 2003). We have adopted these data from their publications after the original calculation and homogenization of the monthly values. To avoid spurious correlations, the series were linearly detrended. Moreover, the monthly series were transformed into series of z-scores with mean of zero and standard deviation of one. Because of their proximity to each other, the series from Uppsala and Stockholm were averaged. Next, the monthly series of z-scores from Tornedalen, St. Petersburg and the mean of Uppsala and Stockholm were averaged into monthly mean temperatures series and the mean series of all of the four meteorological sites compared with the mean phenological records using Pearson correlations. Comparisons were made over the early (1805-1845) and late periods (1846-1875). These periods were decided based on the availability in both types of data. First, the temperature series from Tornedalen starts from 1802 (Klingbjer & Moberg 2003). The series from other sites start in the 18th century but the available series from St. Petersburg (Jones & Lister 2002) do not have values between the years 1801 and 1804. Moreover, there is a notable change in the availability of the phenological data from 1845 to 1846. With these properties in mind, these analyses were run separately over the early (1805-1845) and late periods (1846-1875). The seven mean phenological records (P_4 … P_{10}) were related to mean temperature series over the late period, whereas a limited set of records (P_5 … P_8) was employed over the early period. This limitation was due to lower availability of early spring and autumn observations over the earlier years.

3. Results

3.1 Spatiotemporal distribution

The plant phenological data of Moberg (1857, 1860, 1885, 1894) contains altogether 44487 observations. Plant phenological observations have been made in 193 sites across Finland (Fig. 1). Mean latitude and longitude of all of these observations are 62.01°N and 24.39°E, respectively. A small number of sites remain on the eastern side of the national border and represent the localities in historical Finland (Karelia). The northernmost, southernmost, westernmost and easternmost sites are those of Utsjoki (69.85°N and 26.95°E), Busö (59.87°N and 23.60°E), Eckerö (60.20°N and 19.55°E), and Ilomantsi (62.67°N and 30.90°E), respectively.

The region with highest density of sites is located in the southwestern part of the country. However, the coverage remains relatively high over the region south of 66°N. The northernmost part of the country (Lapland) is obviously less well represented in the dataset. Moreover, the region east of 27°E appears less frequently covered by the data than the areas in the west. Most likely, these characteristics reflect the spatial distribution of population across the country, the regions with higher site density being those with highest density of population.

As for the temporal characteristics of the data, there appear considerable fluctuations in the data availability over the full period 1750-1875 (Fig. 2). Overall, the pre-1846 years contain less data than the years after that date. The period 1750-1845 yield, on an average, 99.8 observations per year, whereas the mean value for 1846-1875 was notably higher, 1163.7 observations per year. Most likely, the increasing number of data over this period was largely due to the intensification of phenological activities as organized by the Finnish Society of Sciences and Letters (Holopainen et al. 2012). Apparently, the number and enthusiasm of the volunteers to collect the data did not remain at the same level but decreased progressively towards the end of the study period.

These changes in temporal availability of the data are directly reflected in the spatial representativeness of the sites across the country. Overall, the full collection of sites is well spread over the country (Fig. 1a). The number of sites is more limited when viewed over any year and especially from early (Fig. 1b) to late years (Fig. 1c).

3.2 Species and their phenological phases

The digitized data represent 450 different species. The species represent 15 different phenological phases (see Table 1; Table S1). Combined, these data result in 662 different phenological variables (i.e. the combinations of species and their phases). The observations can be classified as weeds

(2.2%), annual/biennial (27.5%), archaeophytic (4.2%), fruit-garden (3.2%), ornamental (3.8%), agricultural (32.4%) and woody plants (42.9%). Eleven of the species are each represented by more than one thousand observations. Altogether, these species comprise approximately two fifths (40.4 %) of the observations. The agricultural or woody plants such as rye (*Secale cereale*), barley (*Hordeum vulgare*), potato (*Solanum tuberosum*), rowan (*Sorbus aucuparia*) and bird cherry (*Prunus padus*) are the five most frequently observed species. Moreover, there are 23 species with more than 500 observations, these observations representing nearly three fifths (59.3 %) of the full dataset. The blooming, leaf outbreak, sowing time, leafing and berry ripening are the five most frequently recorded phenological phases. Among the phenological variables, the five most frequently observed are the blooming of rye (*S. cereale*), the sowing of barley (*H. vulgare*), the blooming of bird cherry (*P. padus*), the planting of potato (*S. tuberosum*) the leaf outbreak of birch (*B. pendula* and *B. pubescens*). The data of these variables constitute virtually one tenth (10.3 %) of the observations. There are thirteen different variables with more than 500 observations, these data representing more than one fifth (21.3 %) of the observations.

3.3 Seasonal occurrence

Most of the observations are indicative of phenological phases recorded in May and June (Fig. 3a). Altogether 65 % of all data were collected during these two months. Moreover, the remaining summer months, July and August, are also covered but to a lesser degree (between 9 and 17 %). The months surrounding this late spring-summer season, April, September and October, are each represented by 2-4 % of observations. The spring (March through May), summer (June through August) and autumn (September through November) seasons contain 31.5 %, 60.4 % and 8.1 % of observations, respectively. As much as three fifths (60.9 %) of the observations are made before the midsummer (June 24th). Moreover, the phenological variables (n = 662), that combine the species with observed phase, can be classified according to the month of their most typical (i.e. the mode) observation date. With these regards, the variables are mostly associated with May, June or July representing 85.0 % of the total sum of variables, the remaining months being represented variably by 1-7 % of all variables (Fig. 3a). Of note, the sowing dates of rye and wheat can be divided into different groups according to their seasonal occurrence either in spring/early-summer or late-summer/autumn (Fig. 4). We refer to these as spring and winter rye/wheat, the known varieties of these crop plants from the history of agriculture (Wartiainen 1892).

3.4 Spatial gradients

Correlating the site latitudes with the mean dates of observations in each month showed that the variables observed in spring and summer are positively related to latitude whereas those made in autumn were negatively related to latitude (Fig. 3b). These biogeographical patterns quantify the relative lateness (earliness) of observations of the same phenological variables made in more northern sites in spring (autumn). We further analyse these relationships using a subset of 188 phenological variables with at least 30 observations with latitudinal spread of at least five degrees (their observations represent 82.9 % of the full dataset). The data of observed dates are regressed against the data of their latitudes for each phenological variable, the slope obtained from each such regression expressing the number of days the onset of that variable is shifted per change in degrees of latitude (Fig. 5). Such relationships are exemplified by positive and negative slopes for two common phenological variables, the blooming of wood anemone (*Anemone nemorosa*) (Fig. 5a) most commonly observed in May and the leading of blackcurrant (*Ribes nigrum*) (Fig. 5b) typically occurring in this data in October.

Plotting the slopes of all these variables against the mean dates obtained for each variable quantifies the tendency of spring variables that have earlier dates in the south to have their onset relatively late in the north, and those autumn variables to be considerably late in the south to be relatively early in the north (Fig. 3c). These findings are generally very similar to those obtained for the monthly stratified data (Fig. 3b) but more detailed in spatial and temporal view. Here, the slopes vary between 9.00 days/degrees and -5.27 days/degrees of latitude depending on the variable (Fig. 3c), these uttermost values representing the blooming of coltsfoot (*Tussilago farfara*) and the leafing of rose (*Rosa* sp.), respectively. Among 188 phenological variables, there are only 24 with negative slopes. That is, the great part of the variables comes with their earlier observations in more southern sites, thus representing the spring and summer variables. Again, this finding agrees well with the high number of observations made during the spring and summer rather than autumn (Fig. 3a).

3.5 Temporal variability

Relying on monthly stratified data (Fig. 3a), the phenological variables are averaged according to their predominant observed month of observation. The resulting mean records (P_4 through P_{10}) exhibit notably variable phenological conditions over the 18th and 19th centuries (Fig. 6a). Compared to these simple averages, the latitudinal effect of observation site is removed from the data of individual observations and a new set of mean phenological records is calculated. These records are based on the subset of 188 phenological variables for which the slope (days per latitude) is assessed (Fig. 3c).

Compared temporally, the adjustment factors and the mean latitudes of the sites demonstrate positive relationships i.e. the need for subtracting (adding) the adjustment factor from the dates observed north (south) of the mean latitude of the sites between April and August (P_4 through P_8); for observations made in September (P_9) and October (P_{10}) the signs of the adjustments become inverted (Fig. S1). Before any adjustment, the P_4, P_5, P_6, P_7 and P_8 records in particular correlate positively with mean latitude of the sites used to calculate the mean record; after the observations are adjusted to these variations in the latitude, the correlations are markedly reduced (Fig. 7).

Moreover, there is a notable negative relationship between the mean latitude and the P_{10} record prior to adjustment but virtually no correlation between the two variables after the adjustment is carried out. These changes demonstrate the value of this adjustment in reducing the dependence of the mean value on the latitude of the sites used to calculate that mean (i.e. latitudinal bias). The mean phenological records do not invariably cover the full period (1750-1875). The records constructed from observations made in May (P_5), June (P_6), July (P_7) and August (P_8) are less sporadic than the other records. Overall, these are the four months with most of the observations in the full dataset (Fig. 6b).

Correlations between the mean phenological records are predominantly positive (Table 1). As expected, the highest correlations are obtained for records constructed from data of adjacent months, for example, the highest pair-wise correlation, as high as nearly 0.9, is found between P_6 and P_7. Statistically significant correlations may be found for even longer sub-seasonal connections, for example, the mean records of P_7 and P_9 associate with correlation coefficient of 0.470. Correlations beyond three months are all non-significant. Statistically speaking, these variations reflect the patterns in seasonal autocorrelation as evident in our phenological records.

The mean phenological records demonstrate variable linear trends through the shorter (1846-1875) and longer (1750-1875) periods). Between 1845 and 1875, the trends of early-spring (P_4) and late-summer/autumn (P_8 through P_{10}) indications are negative, whereas the records in-between exhibit positive trends (Table 2a). However, only the trends estimated for P_6 and P_{10} are statistically significant ($p < 0.05$) indicating the early-summer and late-autumn onsets becoming delayed and earlier, hence, the shortening of the growing-season. Over the full period, the only trends being statistically significant ($p < 0.05$) were those of P_6 and P_8 (Table 2b). The slopes of these trends suggest the onset days of these summer months having become delayed by more than one week (9.7 days for P_6 and 9.0 days for P_8) over the 125-year period.

3.6 Climatic signals

Phenological records are mostly negatively associated with the temperature variables (Fig. 8). Clearly, the strongest connections are found when compared with April through June/July temperatures. That is, the warmer the spring and summer season, the earlier the phenological variables are observed. Moreover, the phenological records representing the spring and early-summer observations (P_4, P_5 and P_6) correlate strongest with April and May temperatures, whereas the records representing observations made in later part of the growing season (P_7 and P_8) exhibit strongest correlations with June temperatures. Overall, these patterns are similar over both the late (Fig. 8a) and early periods (Fig. 8b). Albeit statistically significant, the correlations calculated over the early period are, however, relatively lower than those calculated over the late period, these changes possibly reflecting the general uncertainties inherent to data of the early period with lower amount of phenological data (Fig. 2). The phenological records constructed from observations made in autumn (P_9 and P_{10}) are less clearly correlated with temperatures (Fig. 8a). Also these records exhibit statistically significant correlations with a number of temperature variables but their strength appears somewhat diluted in comparison to other phenological records. In fact, the P_9 and P_{10} records are seen to correlate markedly well with the temperature variables representing the climate considerably early in the growing season, in May and March.

4. Discussion

Here we have illustrated the plant phenological data, initially collected by the 18th and 19th contemporaries, after having digitized the information from the original publications of Moberg (1857, 1860, 1885, 1894). The numbers of observations (44487), plant species (450), phenological phases (15), variables (662) and sites (193) reveal the voluminousness of this dataset. Compared to these values, the classical Marsham phenological record from Norfolk (UK) have data from 1736-1925 for 27 phenological events of 17 plant and 10 animal species (Sparks & Carey 1995). Moreover, Rutishauser et al. (2007) analysed historical plant phenological data from Switzerland. Focussing on their data from 1702-1881, these data comprise 551 phenological observations of spring events from 15 sites representative for the Swiss Plateau region (Table 1 in Rutishauser et al. 2007). These figures do not reach the quantities of the Finnish data even over its relatively poorly replicated pre-1846 period when the data was on average covered by nearly one hundred observations per year (see Fig. 2).

Surprisingly, the pre-1846 years of the Finnish data have remained poorly acknowledged. In their review of phenological recordings from historical Finland, Kubin et al. (2008) mention the existence of such data by one sentence only. This data is neither mentioned in the history of international phenology networks, apart from the short-lived phenological campaign of three years (1750-1752) in Sweden and Finland (Koch et al. 2008), after the importance of such data had been advocated by Carl von Linné who also initiated the station network (Dahl & Langvall 2008). After all, it appears that Linné’s recommendation to create phenological datasets was taken seriously only in Finland. In fact, the phenological dataset from Finland ought to be considered as a national effort of its kind and the work by Professor Moberg, who originally compiled this data together, as a career-long personal achievement. Even so, the temporal extent of the dataset remains considerably below those plant phenological records acquired from viticulture archives since 1600 in France and Switzerland (Cook & Wolkovich 2016) and since 1370 in Burgundy (France) (Chuine et al. 2004), not to mention the cherry blossom phenological series from Kyoto, Japan, that have been recorded even since the ninth century (Aono & Kazui 2008; Aono & Saito 2010). We concur with Sparks et al. (2000) that any of such historical data are a resource that should not be left in obscurity after being collected at considerable effort in time and money.

As a national dataset, the geographical spread of the Finnish sites covers a relatively large area and spans nearly ten latitudinal degrees (Fig. 1). This means that the indications of spring and autumn are present in the data as a time-transgressive process in which the earliest spring onsets are observed in the southernmost sites with northward delay (Fig. 5a) while those of autumn onsets are first observed in the north with delay towards the south (Fig. 5b). Quantifications of such gradients for different phenological variables (Fig. 3c) were found essential in order to remove (Eq. 1) the corresponding latitudinal effects from their original observations. In actual fact, the use of adjusted dates is common in phenological literature and there appear several different types of methods for eliminating the site-dependent offsets present in the original dates. Previously, Häkkinen et al. (1995) compared four such statistical methods, one of them being a simple averaging of unadjusted dates, and concluded that the three other methods (based on the simple differences between the individual and a reference series, or the differences estimated by an optimization procedure, or based on the linear mixed mode procedure) resulted in similarly improved composite records of their bud burst dates. The use of such reference series (e.g. Chuine 2004) may become problematic when the data is more or less fragmentary. In the case of our data, it may be even more problematic to define a reference series separately to all different types of phenological variables. Alternatively, these problems may be avoided by transforming the data into z-scores and so to calculate phenological index series all of them possessing a mean of zero and standard deviation of one (Holopainen et al. 2006, 2013). A pitfall of such approach is that the variations at long wavelengths may not be preserved in the resulting index series. As a consequence, the method can be recommended when these low-frequency variations need not to be dealt with (e.g. Holopainen et al. 2009).

In our approach, the data need not be considered as time-series but the adjustment is done based on the site latitude. This method resembles that of Rutishauser et al. (2007) who adjusted their dates of the flowering of cherry and apple tree and beech budburst for the varying site altitudes in the Swiss Plateau region. Similar to our result, the slope of their linear regressions (hence, days per altitude) varied between the phenological variables (Rutishauser et al 2007). Given the purpose of this paper, it was decided not to detect outliers in the data (Schaber & Badeck 2002). We note, however, that their extraction could result in up to a 12% decrease in the confidence levels around the mean dates (Linkosalo et al. 1996). These uncertainties notwithstanding, the phenological mean records portrayed considerable temporal variations (Fig. 6a) that were to large extent attributable to spring and summer temperatures (Fig. 8). These correlations reproduce the previous findings by demonstrating the linkage between the phenological records and seasonal temperatures in the study region (Holopainen et al. 2006, 2013; Linkosalo et al. 2009) and identify those monthly temperature factors that most profoundly drive the plant phenological phases in our boreal setting. Unlike the phenological variables observed in spring and early-summer, those of late summer and autumn (P_8, P_9 and P_{10}) did not correlate strongly with the temperatures of their respective months. That is, the P_9 record exhibited negative and statistically significant ($p < 0.05$) correlations with May and July temperatures, whereas the P_{10} record showed significant correlations with only positive coefficients (Fig. 8a). Possibly, these linkages demonstrate the potential of warm late but not early autumn to prolong the growing season.

Climatic correlations exceeding the level of -0.8 (Fig. 8a) translate into R^2 of nearly 0.7 and demonstrate that a single climatic factor may explain approx. 70 % of the variance in the mean phenological record. More generally, the correlations lie between -0.6 and -0.8 for several mean records, over multiple months and both periods (Fig. 8). These associations demonstrate the twofold role of phenological data as consistent indicators of biological responses to climatic perturbations (Menzel et al. 2005, 2006; Linkosalo et al. 2009; Rutishauser et al. 2009; Cook & Wolkovich 2016) on one hand and as reliable constituents of proxy data for palaeoclimate reconstructions (Craddock 1974; Kington 1974; Chuine 2004; Rutishauser et al. 2007; Holopainen et al. 2009) on the other. It was recently suggested that the plant phenological records could provide proxy data for spring

(March, April and May) temperature variability, to be possibly adopted a component of an annual temperature signal for further reconstructions of past climate variability in northern Europe (Hari et al. 2017). We note that the plant phenological data have indeed been used for such purpose in the same region (Holopainen et al. 2009) and the correlations of this study agreed with similar, strong signal of spring temperatures in plant phenological data. Yet, our results highlight the more comprehensive value of phenological observations for any models for reconstructing our past. The monthly stratified mean phenological records (Fig. 6a) pinpoint the seasonal course of plants' functioning that are drastically shaped by the respective temperature variations. In fact, the digitized data represent pre-industrial climate conditions in the course of the Little Ice Age (Bradley and Jones 1993; Matthews & Briffa 2005). The phenological observations the data presents over this period may be regarded as the evidence of environmental history in the making, as they were once perceived in their natural or culturally shaped environment.

Many of the digitized observations involve agricultural plant species. This finding reinforced the view obtained from previous studies of historical plant phenology in the region almost certainly reflecting the importance of agrarian activities to volunteer observers and their society (Holopainen et al. 2006). These indications may also indicate the actual landscapes where the observations were made. When the observers are mentioned by name (Moberg 1860, 1885, 1894), their titles include occupations such as vicars, priests, professors, doctors, rural police chiefs, lieutenants, doctors and students. They are the naturalists whose field observations and meticulous notes have initially made this work possible. It may be fair to state that many of them have in fact represented the most well-educated persons among their societies. Moreover, they have likely been persons who have observed the seasonal change in their surrounding nature from the sphere of the vicarages, arable and village lands and other similar environments of their time. We also note the high number of recorded plant species. That such a diversity of species was in the first place classified implies that at least some of the observers (e.g. vicars, priests) must have been early naturalist whose academic studies may well have comprised lessons in natural science, even taxonomy. These features not only add credibility to the data but demonstrate the importance of scientific education in making the phenological observations to such taxonomic detail and finally creating the datasets such as that compiled by Professor Moberg. We thus concur with the spirit of Terhivuo et al. (2009) which emphasized the importance of Linné as the initiator of the phenological activities in Finland, but we also note the role of his magnum opus, *Systema Naturae*, making it possible to firmly classify the vast number of species, the development of which through the mid-18th century in fact notably coincides with the timing of the early phase recording the plant phenological observations in the same region.

Acknowledgements

Tapio Linkosalo is thanked for proving us with many ideas of phenological data analyses. J.H. was funded by the Ella and Georg Ehrnrooth Foundation and S.H. by the Academy of Finland (Grant no. 288267).

References

Bradley, R.S., Jones, P.D. (1993). ‘Little Ice Age’ summer temperature variations: Their nature and relevance to recent global warming trends. *The Holocene*, 3, 367-376.

Chuine, I., Yiou, P., Viovy, N., Seguin, B., Daux, V., Le Roy Ladurie, E. (2004). Historical phenology: grape ripening as a past climate indicator. *Nature*, 432, 289-290.

Cook, B.I., Wolkovich, E.M. (2016). Climate change decouples drought from early wine grape harvests in France. *Nature Climate Change*, 6, 715-719.

Craddock, J.M. (1974). Phenological indicators and past climates. *Weather*, 29, 332-344.

Dahl, Å., Langvall, O. (2008). Observations on phenology in Sweden – past and present. In: Nekovář, J. (eds), *The history and current status of plant phenology in Europe*. COST Action 725, Vammala. 161-165.

Dutka, J. (1988). On the Gregorian Revision of the Julian Calendar. *The Mathematical Intelligencer*, 10, 56-64.

Häkkinen, R., Linkosalo, T., Hari, P. (1995). Methods for combining phenological time series: application to bud burst in birch (*Betula pendula*) in Central Finland for the period 1896–1955. *Tree Physiology*, 15, 721-726.

Hämet-Ahti, L., Suominen, J., Ulvinen, T., Uotila, P. (1998). *Retkeilykasvio (Field Flora of Finland)*, Fourth ed. Finnish Museum of Natural History, Botanical Museum, Helsinki. 656 p.

Hari, P., Aakala, T., Hilasvuori, E., Häkkinen, R., Korhola, A., Korpela, M., Linkosalo, T., Mäkinen, H., Nikinmaa, E., Nöjd, P., Seppä, H., Sulkava, M., Terhivuo, J., Tuomenvirta, H., Weckström, J., Hollmen, J. (2017). Reliability of temperature signal in various climate indicators from northern Europe. *PLoS ONE* 12(6): e0180042. https://doi.org/10.1371/journal.pone.0180042.

Hatcher, D.A. (1984). Simple formulae for Julian day numbers and calendar dates. *Quarterly Journal of the Royal Astronomical Society*, 25, 53-55.

Holopainen, J. (2006). Reconstructions of past climates from documentary and natural sources in Finland since the 18th century. *Publications of the Department of Geology*, D9, 1-33.

Holopainen, J., Gregow, H., Helama, S., Kubin, E., Lummaa, V., Terhivuo, J. (2012). Suomen kasvifenologisista havainnoista 1700-luvun puolivälistä nykypäivään. Summary in English: History of Finnish plant phenological observations since the 1750s. *Sorbifolia*, 43, 51-66.

Holopainen, J., Helama, S., Kajander, J.M., Korhonen, J., Launiainen, J., Nevanlinna, H., Reissell, A., Salonen, V.-P. (2009). A multiproxy reconstruction of spring temperatures in south-west Finland since 1750. *Climatic Change*, 92, 213-233.

Holopainen, J., Helama, S., Lappalainen, H., Gregow, H. (2013). Plant phenological records in northern Finland since the 18th century as retrieved from databases, archives and diaries for biometeorological research. *International Journal of Biometeorology*, 57, 423-435.

Holopainen, J., Helama, S., Timonen, M. (2006). Plant phenological data and tree-rings as palaeoclimate indicators since AD 1750 in SW Finland. *International Journal of Biometeorology*, 51, 61-72.

Jones, P.D., Lister, D.H. (2002). The daily temperature record for St. Petersburg (1743–1996). *Climatic Change*, 53, 253-267

Kerzhner, I.M. (1984). Converting Dates from the Julian (Old Style) or French Republican (Revolutionary) Calendars to the Gregorian (New Style) Calendar. *Taxon*, 33, 410-412.

Kington, J.A. (1974). An application of phenological data to historical climatology. *Weather*, 29, 320-328.

Klingbjer, P., Moberg, A. (2003). A composite monthly temperature record from Tornedalen in northern Sweden, 1802–2002. *International Journal of Climatology*, 23, 1465-1494.

Koch, E., Demarée, G., Lipa, W., Zach, S., Zimmermann, K. (2008). History of international phenology networks. In: Nekovář, J. (eds), *The history and current status of plant phenology in Europe*. COST Action 725, Vammala. 12-16.

Kubin, E., Pudas, E. Venäläinen, A., Terhivuo, J. (2008). Phenological recording in Finland. In: Nekovář, J. (eds), *The history and current status of plant phenology in Europe*. COST Action 725, Vammala. 68-75.

Lappalainen, H., Heikinheimo, M. (1992). Relations between climatological and plant phenological observations. Vol. 1. Survey of plant phenological observations in Finland from 1896 to 1965. *Meteorological Publications* 20: 1-74.

Linkosalo, T., Häkkinen, R., Hari, P. (1996). Improving the reliability of a combined phenological time series by analyzing observation quality. *Tree Physiology*, 16, 661-664.

Linkosalo, T., Häkkinen, R., Terhivuo, J., Tuomenvirta H., Hari, P. (2009). The time series of flowering and leaf bud burst of boreal trees (1846–2005) support the direct temperature observations of climatic warming. *Agricultural and Forest Meteorology*, 149, 453-461.

Margary, I.D. (1925). The Marsham phenological record in Norfolk, 1736-1925, and some others. *Quaterly Journal of the Royal Meteorological Society*, 52, 27-54.

Matthews, J.A., Briffa, K.R. (2005). The ‘Little Ice Age’: Re-evaluation of an evolving concept. *Geografiska Annaler*, 87A, 17-36.

Menzel, A. (2002). Phenology: Its Importance to the Global Change Community. *Climatic Change*, 54, 379-385.

Menzel, A., Sparks, T., Estrella, N., Eckhardt, S. (2005). ‘SSW to NNE’ – North Atlantic Oscillation affects the progress of seasons across Europe. *Global Change Biology*, 11, 909-918.

Menzel, A., Sparks, T., Estrella, N., Koch, E., Aasa, A., Ahas, R., Alm-Kübler, K., Bissolli, P., Braslavska, O., Briede, A., Chmielewski, F.M., Crepinsek, Z., Curnel, Y., Dahl, A., Defila, C., Donnelly, A., Filella, Y., Jatczak, K., Mage, F., Mestre, A., Nordli, Ø., Penuelas, J., Pirinen, P., Remisova, V., Scheifinger, H., Striz, M., Susnik, A., Vliet, A.J.H. van, Wielgolaski, F.E., Zach, S., Zust, A. (2006). European phenological response to climate change matches the warming pattern. *Global Change Biology*, 12, 1969-1976.

Moberg, A. (1857). Naturhistoriska dag-anteckningar gjorda i Finland, Åren 1750-1845. Notiser ur Sällskapets pro Fauna & Flora fennica förhandlingar. *Bihang till Acta Societatis Scientarium Fennicæ*, 3, 95-250.

Moberg, A. (1860). Klimatologiska iakttagelser i Finland, föranstaltade och utgifna af Finska Vetenskaps-Societeten. Första delen. År 1846-1855. I. *Bidrag till Finlands naturkännedom, etnografi och statistic*, 7, 1-361.

Moberg, A. (1885). Klimatologiska iakttagelser i Finland, föranstaltade och utgifna af Finska Vetenskaps-Societen. Andra delen. År 1856-1875. I. Fenologiska anteckningar. *Bidrag till kännedom af Finlands natur och folk*, 41, 1-321.

Moberg, A. (1894). *Fenologiska iakttagelser i Finland åren 1750-1845*. Utgifna af Finska Vetenskaps-Societeten. Finska Litteratur-Sällskapets tryckeri, Helsingfors. *Bidrag till kännedom af Finlands Natur och Folk*, 55, 1-165.

Moberg, A., Bergström, H. (1996). Homogenization of Swedish temperature data. Part III: the long temperature records from Uppsala and Stockholm. *International Journal of Climatology*, 17, 667-699.

Moberg, A., Alexandersson, H., Bergström, H., Jones, P.D. (2003). Were Southern Swedish summer temperatures before 1860 as warm as measured? *International Journal of Climatology*, 23, 1495-1521.

Rutishauser, T., Luterbacher, J., Jeanneret, F., Pfister, C., Wanner, H. (2007). A phenology-based reconstruction of interannual changes in past spring seasons. *Journal of Geophysical Research*, 112, G04016, DOI: 10.1029/2006JG000382.

Rutishauser, T., Schleip, C., Sparks, T.H., Nordli, Ø., Menzel, A., Wanner, H., Jeanneret, F., Luterbacher, J. (2009). Temperature sensitivity of Swiss and British plant phenology from 1753 to 1958. *Climate Research*, 39, 179-190.

Sadler, D.H. (1978). Mean Solar Time on the Meridian of Greenwich. *Quarterly Journal of the Royal Astronomical Society*, 19, 290-309.

Sagarin, R. (2001). False estimates of the advance of spring. *Nature*, 414, 600.

Sagarin, R. (2009). Using nature's clock to measure phenology. *Frontiers in Ecology and the Environment*, 7, 296.

Sagarin, R., Micheli, F. (2001). Climate change in nontraditional data sets. *Science*, 294, 811.

Schaber, J.A., Badeck, F.-W. (2002). Evaluation of methods for the combination of phenological time series and outlier detection. *Tree Physiology*, 22, 973-982.

Sparks, T.H., Carey, P.D. (1995). The responses of species to climate over two centuries: an analysis of the Marsham phenological record, 1736-1947. *Journal of Ecology*, 83, 321-329.

Sparks, T.H., Jeffree, E.P., Jeffree, C.E. (2000). An examination of the relationship between flowering times and temperature at the national scale using long-term phenological records from the UK. *International Journal of Biometeorology*, 44, 82-87.

Tarand, A., Kuiv, P. (1994). The beginning of the rye harvest – a proxy indicator of summer climate in the Baltic Area. *Paläoklimaforschung*, 13, 61-72.

Tarand, A., Nordli, P.Ø. (2001). The Tallinn temperature series reconstructed back half a millennium by use of proxy data. *Climatic Change*, 48, 189-199.

Terhivuo, J, Kubin, E, Karhu, J. (2009). Phenological observations since the days of Linné in Finland. *Italian Journal of Agrometeorology*, 14, 45-49.

Wartiainen, A. (1892). *Käytännöllinen maanviljelys-oppi. Prof. J. Arrheniuksen tekemän ruotsinkielisen „Landtbrukspraktika" nimisen oppikirjan nojalla ja sitä suomentaen, mukaellen ja lisäten*. Werner Söderström, Porvoo.

TABLES

Table 1. Lists of ten most frequency observed plant species and its phenological phases in the digitized data with their number of observations (n). See Table S1 for a full list of species names.

Species	n	Phase	n
Secale cereale	3455	blooming	22659
Hordeum vulgare	2439	leaf outbreak	8068
Solanum tuberosum	1775	sowing	3047
Sorbus aucuparia	1575	leafing	2641
Prunus padus	1541	berry ripening	2269
Avena sativa	1536	harvest	1432
Betula pendula / B. pubescens	1317	branch growth	1252
Ribes uva-crispa	1156	grain maturation	1045
Ribes spicatum	1112	planting	839
Ribes nigrum	1048	new crop	580

Table 2. Pearson correlations between the mean phenological records constructed from observations made predominantly in April (P_4), May (P_5), June (P_6), July (P_7), August (P_8), September (P_9) and October (P_{10}). Statistically significant ($p < 0.05$) correlations are denoted by asterisk (*).

	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
P_4	1.000						
P_5	0.579*	1.000					
P_6	0.371*	0.810*	1.000				
P_7	0.327	0.647*	0.886*	1.000			
P_8	0.164	0.468*	0.686*	0.847*	1.000		
P_9	-0.100	0.261	0.443*	0.470*	0.536*	1.000	
P_{10}	-0.097	-0.079	0.110	0.031	-0.062	0.052	1.000

Table 3. Linear trends through the late (1846-1875) (a) and full (1750-1875) (b) periods quantified using the trend slope (change in days per year), Pearson correlation (r) and significance value (p). No estimation was carried out for P9 over the full period because of lack of data (see Fig. 6).

		P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
(a)	slope	-0.304	0.154	0.216	0.076	-0.050	-0.172	-0.188
	r	-0.260	0.239	0.376	0.147	-0.076	-0.277	-0.443
	p	0.174	0.203	0.041	0.439	0.69	0.139	0.014
(b)	slope	0.058	0.023	0.078	0.036	0.072	---	-0.013
	r	0.151	0.119	0.379	0.167	0.358	---	-0.058
	p	0.271	0.190	<0.001	0.091	<0.001	---	0.628

FIGURE CAPTIONS

Fig. 1. A map of Finland with phenological observation sites over the full period (1750-1875) (a), and over the years of 1800 (b) and 1850 (c).

Fig. 2. Temporal availability of the phenological data digitized from the books of Moberg (1857, 1860, 1885, 1894).

Fig. 3. Monthly percentages of phenological observations and variables (a), Pearson correlations between the latitude of the site and the date observed in different months (b), and the latitudinal gradient in the dates of observations as a function of the mean dates of difference phenological variables.

Fig. 4. Frequency counts of rye (a) and barley (b) according to their sowing dates.

Fig. 5. Obtaining the dependence of phenological date (y) on site latitude (x) exemplified for the blooming of wood anemone (a) and the leading of blackcurrant (b). The slope of the regression line is quantified by the parameterised equation.

Fig. 6. Mean phenological records (P_4, P_5 … P_{10}) over the study period (1750-1875) (a) and their number of observations (b). The dates mentioned in the text are denoted as calendar years.

Fig. 7. Correlations between the mean latitude (see Fig. S1) and the phenological mean records (P_4, P_5 … P_{10}) before and after the adjustment for the varying latitude is carried out.

Fig. 8. Correlations of phenological records (P_4, P_5 … P_{10}) with the monthly mean temperatures calculated over the late (a) and early period (b). The correlations with highest coefficient obtained for each phenological record is denoted. The level of statistical significance ($p < 0.05$) is marked as horizontal dashed line.