SHORT COMMUNICATION

The plant phenological online database (PPODB): an online database for long-term phenological data

Jonas Dierenbach · Franz-W. Badeck · Jörg Schaber

Abstract We present an online database that provides unrestricted and free access to over 16 million plant phenological observations from over 8,000 stations in Central Europe between the years 1880 and 2009. Unique features are (1) a flexible and unrestricted access to a full-fledged database, allowing for a wide range of individual queries and data retrieval, (2) historical data for Germany before 1951 ranging back to 1880, and (3) more than 480 curated long-term time series covering more than 100 years for individual phenological phases and plants combined over Natural Regions in Germany. Time series for single stations or Natural Regions can be accessed through a user-friendly graphical geo-referenced interface. The joint databases made available with the plant phenological database *PPODB* render accessible an important data source for further analyses of long-term changes in phenology. The database can be accessed via www.ppodb.de.

Keywords Phenology database · Long-term data · Phenology data · Combined time series · Outliers · Plant phenology · Online access

Introduction

Plant phenological data, even though being low-technology subjective observations often made by volunteers, are a prerequisite

J. Dierenbach · J. Schaber (✉)
Institute for Experimental Internal Medicine, Medical Faculty,
Otto von Guericke University, Leipziger Str. 44,
39120 Magdeburg, Germany
e-mail: schaber@med.ovgu.de

F.-W. Badeck
Potsdam Institute for Climate Impact Research (PIK), PF 60 12 03,
14412 Potsdam, Germany

of studies within several modern research fields, like, e.g., climate change (Badeck et al. 2004; Menzel and Fabian 1999; Wolkovich et al. 2012; Penuelas et al. 2009; Fridley 2012; Chuine et al. 2004; Menzel et al. 2006; Rosenzweig et al. 2007) and crop yields (Tao et al. 2006), and are of interest to the agro- and pharma-industries. As flexible and free access to phenological data is limited, we built an online database, which provides a flexible, free and unrestricted access to plant phenological observations. We compiled data from various sources and made it accessible in a consistent format. In addition, we digitized historical data, thereby, for the first time, providing phenological observations for 130 years (1880–2009) for a large geographical region, i.e. Germany, in a consistent format. The data can either be retrieved as time series with respect to a plant, phase and station by a simple geo-referenced interface, or by a full-fledged SQL database access allowing for a wide range of individual queries (examples given below). Moreover, we provide combined time series for Natural Regions, which are corrected for outliers.

Data and methods

In the following, we refer to 'observation' as the reported day of the year in which a certain phenological phase, e.g. blossoming, of a certain plant was observed at a certain location, i.e. station. The data in *PPODB* is compiled from three distinct databases (see Table 1 for summary statistics):

- Phenological observations collected by the Deutscher Wetterdienst (German meteorological service, DWD) from 1951 to 2009 (database DWD in Table 1).
- The historical phenological database provided by the DWD, which is a collection of phenological observations from Central Europe, mainly Germany, covering

Table 1 PPODB overview. Data are given as numbers of stations, phases, observations and observation periods in the different combined databases including plant varieties

Database	Plant type	Number of species	Number of phases	Number of stations	Number of observations	Observation period
DWD	Wild	37	75	6,514	5,897,274	1951–2009
Lat: 47.4′–55′, Lon: 6′–15′	Agro	25	140	6,410	5,981,960	1951–2009
	Fruit	24	67	6,433	3,534,316	1951–2009
	Vine	4	27	1,176	100,609	1951–2009
	All	90	309	6,544	15,514,159	1951–2009
HIS	Wild	26	48	1,195	128,387	1921–1955
Lat: 47.4′–55.1′, Lon: 6.1′–22.5′	Agro	40	157	1,262	50,630	1921–1955
	Fruit	20	64	1,191	45,924	1921–1955
	All	86	269	1,664	224,941	1921–1955
HPDB	Wild	28	58	1,099	109,782	1880–1941
Lat: 18.8′–63.1′, Lon:–76.6′–112.2′	Agro	11	24	944	16,955	1880–1941
	Fruit	18	42	1,078	52,943	1880–1941
	NA[a]	290	1,398	1,161	122,015	1880–1941
	All	293	1,565	1,184	284,235	1880–1941
Combined data[b]	Wild	37	75	7,952	6,142,068	1880–2009
Lat: 38.71′–63.1′, Lon:–76.6′–112.2′	Agro	47	249	7,807	6,064,664	1880–2009
	Fruit	31	99	7,859	3,623,751	1880–2009
	Vine	4	27	1,182	101,383	1951–2009
	NA[a]	277	1,344	1,073	116,185	1880–1944
	All	352	1,961	8,333	16,065,517	1880–2009

For a short description of the individuals databases (DWD, HIS, HPDB) and their sources refer to the Data and Methods section

[a] Unspecified plant types refer to plant–phase combination that are not in the actual DWD database

[b] The Combined Data refers to the database where stations and corresponding observations from all three databases have been combined and merged

the years 1880 until 1941 compiled from various sources (database HPDB in Table 1).

- To fill the gap between the two aforementioned databases, we digitized phenological data that were available only in printed form. These data were collected by the volunteer network of the Deutscher Reichswetterdienst, and were published after World War II. These observations cover the years 1922 until 1944. Additionally, we digitized phenological data that were published between 1951 and 1961 in the meteorological yearbooks of the DWD (DWD 1951, 1953, 1960, 1961, 1991). Taken together, these historical data cover the years 1921–1955 and is made publicly available here for the first time (database HIS in Table 1).

In Table 1 we also provide the geographical area covered in each of the databases, respectively. We provide a table of the countries, number of observations and observational time range per country in Supplementary Table S1. Please refer to the description of the online database at www.ppodb.de for more details and a full description of the database.

Results

There are three ways to access the database, i.e. time series access for single station or Natural Regions, or full-fledged SQL-access. We shortly illustrate these three main features.

When accessing the database via www.ppodb.de the user first encounters a page where two perspectives can be chosen, i.e. single stations or Natural Regions, for a certain group of plants (Fig. 1).

For clarity, we grouped the different plant types into agricultural plants, fruits, wild growing plants and vines.

Single stations

In the single station perspective plants, corresponding phases, and stations can be selected via drop-down menus (Fig. 2). Stations can additionally be selected by clicking on respective markers in the map. Initially,
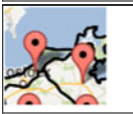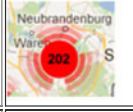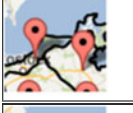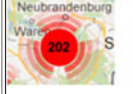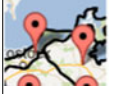
Welcome to PPODB,

this **P**lant-**P**henological **O**nline **D**atabase offers access to plant-phenological observations in Central Europe mainly Germany, from about 9000 stations, covering 130 years (1880-2009), comprising over 16 millions observations.

The PPODB offers a unique data source, where, for the first time, plant-phenological observations from 129 years are brought together in one consistent database. Moreover, sophisticated outlier detection methods and quality checks are applied.

There are two ways to access the data:

1) In the SQL-interface we provide a full-fledged access to the complete database, allowing for highly flexible data retrieval via SQL-queries.

2) Below, we provide a visualization and access to phenological time series via a geo-referenced interface, using google-maps. Here you can visualize and retrieve phenological time series for a single station or a Natural Region, depending on the chosen perspective.

In order to access phenological time series, please select a combination of **plant type** and **perspective** from the table below by clicking on the corresponding image:

plant type	perspective	
	single stations	Natural Regions
agriculture		
fruits		
wild growing plants		
vines		

At the moment available perspetives are:

single stations: Here you can select single stations to analyze data.
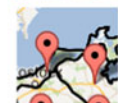
natural regions: Here you can select Natural Regions to analyze your data.

PPODB is optimized for Firefox and Google Chrome. You might have problems with Internet Explorer.

Fig. 1 *PPODB* start interface (screenshot from browser)

a map of Central Europe is displayed where most of the stations are grouped into clusters indicated by coloured circles for a better overview. The numbers on the coloured circles indicate the number of stations which are represented by this cluster. Clicking on cluster symbols zooms into the map, where the location of single stations becomes visible. Single stations are marked by red balloons, which contain some general information about the station, like station name, longitude, latitude, altitude, number and range of years in which observations were made (Fig. 2).

Once plant, phase and station are selected, the corresponding time series can be displayed either in a graph (menu 'plot only'), table (menu 'data only') or both (menu 'data and plot'). Optionally, a trend line is provided with the calculated trend and corresponding *P*-value (Fig. 3).

In case the station is present in different databases, the respective observations are colour coded (Fig. 3). Note that observations from the same station and year might have been reported in different databases with differing values. We kept all reported observations in the databases, even though in these cases the day of observation of the respective phase is ambiguous.

In the Supplementary Material we provide an additional summary table with all species–phase combinations in the combined database, which are still being observed by the German Weather Service, with their number of stations and

Fig. 2 Single station perspective. Example for the phase 'beginning of flowering' of horse chestnut at the station 'Geisenheim(DWD)' (screenshot). *Balloons* indicate phenological stations. By clicking on the *balloons* some meta-information about the station is displayed

observations, and average length of time series per plant, phase and station.
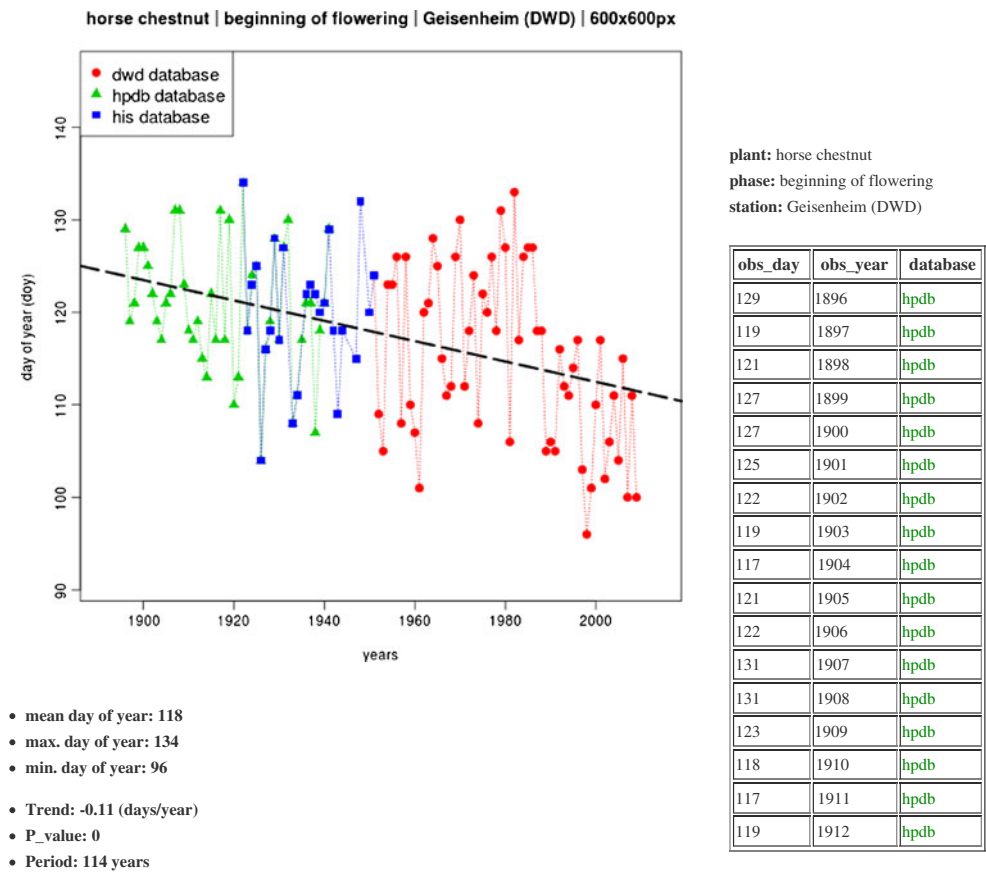
Natural regions

One of the main reasons to construct this database and to merge stations from different databases was to enable the construction of long phenological time series, so-called combined time series, in order to study the effect of climate change on plant phenology (Schaber 2002; Schaber and Badeck 2002, 2003, 2005; Schaber et al. 2010). A combined times series is a sophisticated average over many time series

that corrects for artefacts introduced by simple averages due to the unequal distribution of observations in time and space (Schaber et al. 2010). In Fig. 4 we show histograms of the number of time series of a certain length for single stations and Natural Regions, respectively. For Natural Regions, there is a substantial increase of long time series at the expense of short time series. There are more than 480 combined time series for certain phenological phases for certain Natural Regions covering more than 100 years.

Selecting the Natural Regions perspective in the start menu (Fig. 1), the user is presented an interface, where plant, phase and Natural Region can be selected (Fig. 5).

Fig. 3 Plot of the time series for the phase 'beginning of flowering' of horse chestnut at the station 'Geisenheim(DWD)'. The colour code of the data points indicates the source of the corresponding observation (screenshot), which is also displayed in the corresponding table

horse chestnut | beginning of flowering | Geisenheim (DWD) | 600x600px

- **mean day of year: 118**
- **max. day of year: 134**
- **min. day of year: 96**

- **Trend: -0.11 (days/year)**
- **P_value: 0**
- **Period: 114 years**

plant: horse chestnut

phase: beginning of flowering

station: Geisenheim (DWD)

obs_day	obs_year	database
129	1896	hpdb
119	1897	hpdb
121	1898	hpdb
127	1899	hpdb
127	1900	hpdb
125	1901	hpdb
122	1902	hpdb
119	1903	hpdb
117	1904	hpdb
121	1905	hpdb
122	1906	hpdb
131	1907	hpdb
131	1908	hpdb
123	1909	hpdb
118	1910	hpdb
117	1911	hpdb
119	1912	hpdb

Again plant, phase and Natural Regions can be selected by drop down menus. Natural Regions can also be selected by clicking on the map. From this perspective, combined time series with error bars can be
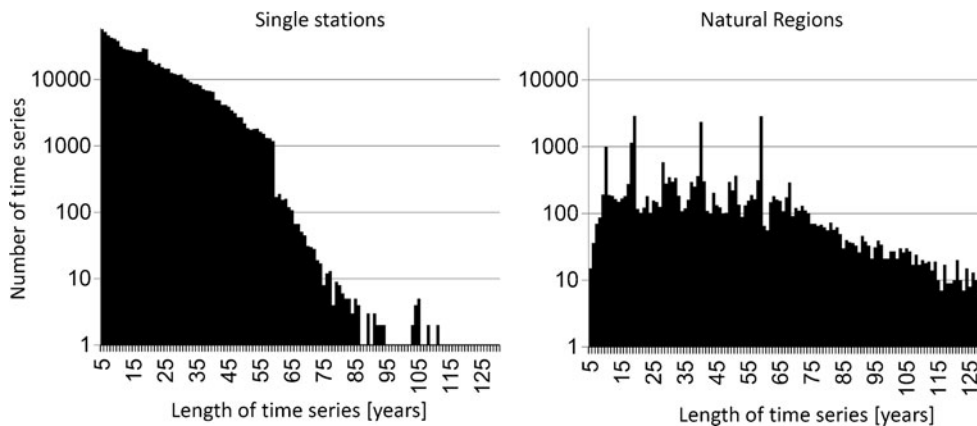

Fig. 4 Histograms of the number of time series per length of time series over all species, phases and single stations (*left panel*) and Natural Regions (*right panel*), respectively. The corresponding data can be retrieved from the database with the following SQL-queries (*left panel*): "select c, count(c) from (select stat_id, phase_id, count(distinct obs_year) as c from all_pheno_obs where phase_id ! = 0 group by stat_id, phase_id) as sq group by c", and (*right panel*): "select c, count(c) from (select naturraumgruppen_id, phase_id, count(distinct obs_year) as c from pheno_nr_ts group by naturraumgruppen_id, phase_id) as sq group by c"
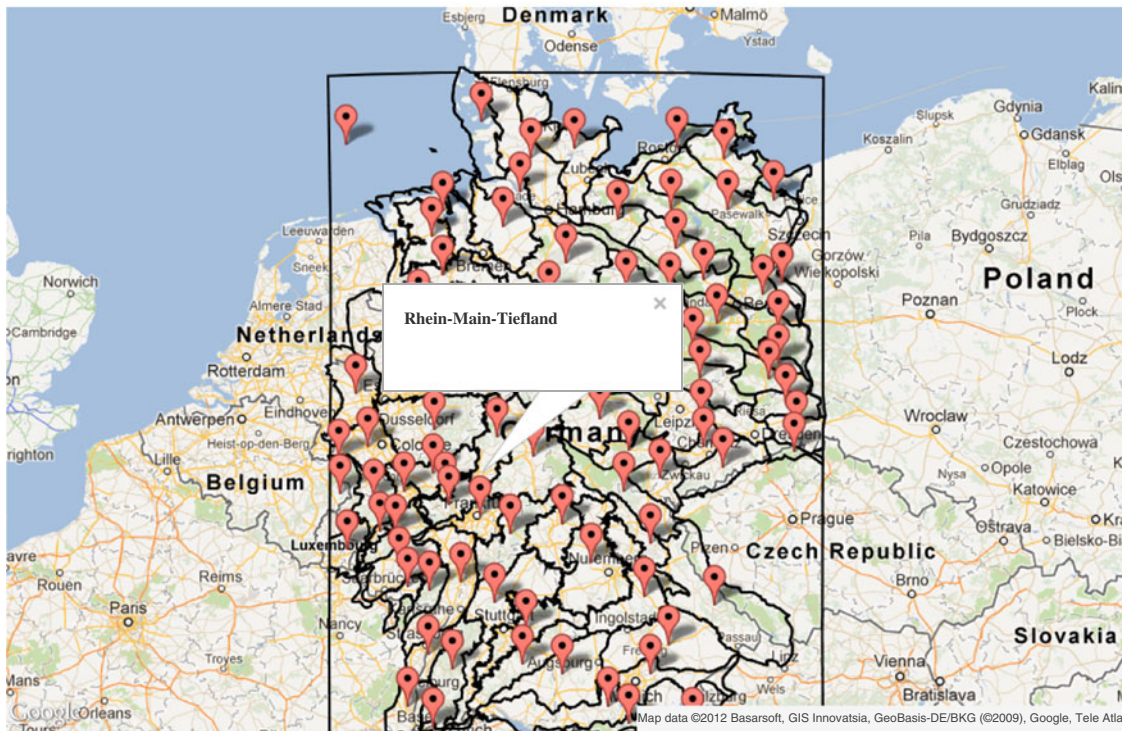
Fig. 5 Natural Regions perspective. Example for the phase 'bud burst of beech in the Natural Region Rhein-Main Tiefland' (screenshot). *Balloons* indicate the geographical centre of the respective Natural Region. By clicking balloons the name of Natural Region is displayed

displayed (Fig. 6). Again, a trend can be optionally displayed.

The origin of the combined data is colour-coded as above with the extension that an estimated combined data point can come from more than one database. In the corresponding table the number of observations for each combined data point is also displayed.

SQL access

Through the 'SQL access'-tab (see Fig. 1), the database can be accessed via SQL statements that allow all kinds of individual queries. The data for Fig. 4, e.g., can be extracted

by one single SQL statement (see Fig. 4). For the summary Tables S1–S3 we also provide the respective SQL-statements as an example of the flexibility and range of queries.

Discussion

The joint databases made available with *PPODB* render accessible an important data source for further analyses of long-term changes in phenology.

The database is unique in as far as it covers more than a century of observations for a large geographical region and

Fig. 6 Plot of the time series for the phase bud burst of beech in the Natural Region 'Rhein-Main Tiefland'. The colour code of the data points indicates the source of the corresponding observation (screenshot). The *error bars* are the lower and upper 95 % confidence levels of the estimated mean day of year (*dots*), which are displayed in the corresponding table (L95CL and L95UL, respectively). The number of observations per year (n_obs) in the table indicate the number of observations/stations the combined mean per year was calculated from
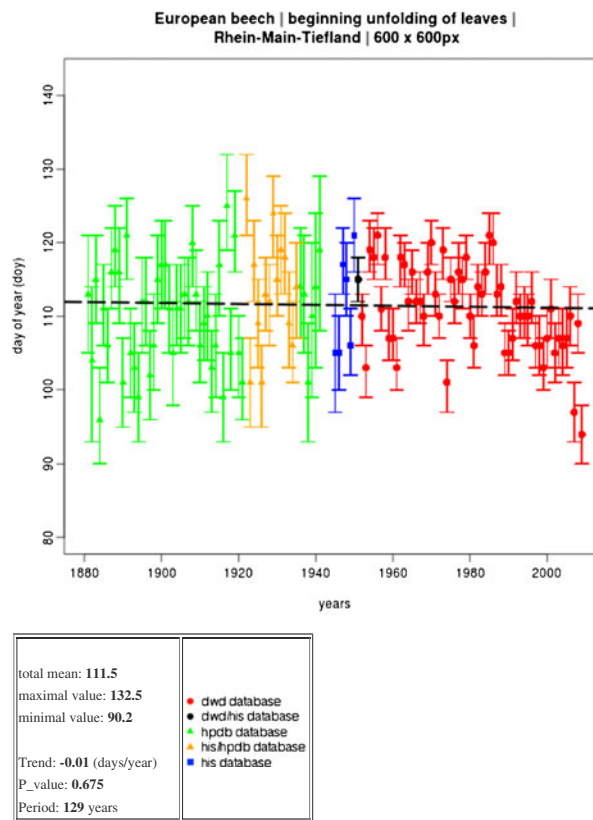
plant: European beech
phase: beginning unfolding of leaves
natural region: Rhein-Main-Tiefland

obs_year	mean doy	L95Cl	U95Cl	n_obs	database
1881	113.7	105.69	121.8	2	hpdb
1882	104.1	93.38	114.82	1	hpdb
1883	115.4	109.65	121.31	5	hpdb
1884	97	90.93	103	5	hpdb
1885	111.7	106.4	117.03	7	hpdb
1886	106.4	101.11	111.75	7	hpdb
1887	116.8	111.37	122.29	6	hpdb
1888	119.9	114.07	125.65	5	hpdb
1889	116.2	109.47	122.92	3	hpdb
1890	101.4	95.28	107.57	4	hpdb
1891	121.2	115.5	126.94	5	hpdb
1892	105.8	99.65	111.85	4	hpdb
1893	103.8	97.65	109.85	4	hpdb
1894	99.8	93.65	105.85	4	hpdb
1895	112	105.9	118.1	4	hpdb
1896	112.5	106.9	118.04	5	hpdb
1897	102.9	96.93	108.9	4	hpdb
1898	106.8	100.16	113.3	3	hpdb
1899	115.7	109.77	121.71	4	hpdb
1900	117.5	111.52	123.46	4	hpdb

total mean: **111.5**
maximal value: **132.5**
minimal value: **90.2**

Trend: **-0.01** (days/year)
P_value: **0.675**
Period: **129** years

- dwd database
- dwd/his database
- hpdb database
- his/hpdb database
- his database

at the same time a substantial number of species as well as many observation stations. It complements another phenological online database, the paneuropean phenology database *PEP725* (Koch et al. 2009) (www.pep725.eu), which is also unique in the sense that it partly covers other countries than *PPODB* for which it has more contemporary data. *PEP725* observations start as early as 1868, but only for the relatively small region of the Netherlands. *PEP725* provides data retrieval with downloadable species-related observations per country, where observations, station description and phase description are provided in separate files. With *PPODB* we provide an instrument including an SQL interface to the complete database that greatly facilitates data retrieved for all kind of summary information or very specific and focussed information and can potentially be used to improve the access to the *PEP725* data base. In addition, the data of the HIS and HPDB databases can be used to construct long-term combined time series with data provided by *PEP725* for other countries, especially Poland. Moreover, the combined time series for the Natural Regions of Germany provide a unique data source of reliable long-term phenological time series for a range of species and phenological phases.

References

Badeck FW, Bondeau A, Bottcher K, Doktor D, Lucht W, Schaber J, Sitch S (2004) Responses of spring phenology to climate change. New Phytol 162(2):295–309. doi:10.1111/j.1469-8137.2004.01059.x

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(7015):289–290. doi:10.1038/432289a

DWD (1951) Deutsches Meteorologisches Jahrbuch der US-Zone. Meteorologisches Jahrbuch. Deutscher Wetterdienst, Offenbach

DWD (1953) Deutsches meteorologisches Jahrbuch (Britische Zone). Meteorologisches Jahrbuch. Deutscher Wetterdienst, Offenbach

DWD (1960) Deutsches meteorologisches Jahrbuch, Gebiet der ehemaligen Französischen Besatzungszone. Meteorologisches Jahrbuch. Deutscher Wetterdienst, Offenbach

DWD (1961) Deutsches meteorologisches Jahrbuch. Saarland. 1945–1955. Meteorologisches Jahrbuch. Deutscher Wetterdienst, Offenbach

DWD (1991) Anleitung für die phänologischen Beobachter des Deutschen Wetterdienstes, vol 17. Vorschriften und Betriebsunterlagen. Deutscher Wetterdienst, Offenbach am Main

Fridley JD (2012) Extended leaf phenology and the autumn niche in deciduous forest invasions. Nature 485(7398):359–362. doi:10.1038/nature11056

Koch E, Dittmann E, Lipa W, Menzel A, Nekovar J, Sparks TH, van Vliet AJH (2009) COST725—establishing a European phenological data platform for climatological applications: major results. Adv Sci Res 3:119–122. doi:10.5194/asr-3-119-2009

Menzel A, Fabian P (1999) Growing season extended in Europe. Nature 397:659

Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, Alm-Kuebler K, Bissolli P, Braslavska O, Briede A, Chmielewski FM, Crepinsek Z, Curnel Y, Dahl A, Defila C, Donnelly A, Filella Y, Jatcza K, Mage F, Mestre A, Nordli O, Penuelas J, Pirinen P, Remisova V, Scheifinger H, Striz M, Susnik A, Van Vliet AJH, Wielgolaski F-E, Zach S, Zust A (2006) European phenological response to climate change matches the warming pattern. Glob Chang Biol 12(10):1969–1976. doi:10.1111/j.1365-2486.2006.01193.x

Penuelas J, Rutishauser T, Filella I (2009) Ecology. Phenology feedbacks on climate change. Science 324(5929):887–888. doi:10.1126/science.1173004

Rosenzweig C, Casassa G, Karoly DJ, Imeson A, Liu C, Menzel A, Rawlins S, Root TL, Seguin B, Tryjanowski P (2007) Assessment of observed changes and responses in natural and managed systems. In: Parry ML, Canziani OF, Palutikof JP, Linden PJvd, Hanson CE (eds) Climate change 2007: impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp 79–131

Schaber J (2002) Phenology in Germany in the 20th century: methods, analyses and models. PhD, University of Potsdam, Potsdam

Schaber J, Badeck FW (2002) Evaluation of methods for the combination of phenological time series and outlier detection. Tree Physiol 22(14):973–982

Schaber J, Badeck FW (2003) Physiology-based phenology models for forest tree species in Germany. Int J Biometeorol 47(4):193–201. doi:10.1007/s00484-003-0171-5

Schaber J, Badeck F-W (2005) Plant phenology in Germany over the 20th century. Reg Environ Chang 5(1):37–46. doi:10.1007/s10113-004-0094-7

Schaber J, Badeck F, Doktor D, von Bloh W (2010) Combining messy phenological time series phenological research. In: Hudson IL, Keatley MR (eds) Springer, Netherlands, pp 147–158. doi:10.1007/978-90-481-3335-2_7

Tao F, Yokozawa M, Xu Y, Hayashi Y, Zhang Z (2006) Climate changes and trends in phenology and yields of field crops in China, 1981–2000. Agric For Meteorol 138(1–4):82–92. doi:10.1016/j.agrformet.2006.03.014

Wolkovich EM, Cook BI, Allen JM, Crimmins TM, Betancourt JL, Travers SE, Pau S, Regetz J, Davies TJ, Kraft NJ, Ault TR, Bolmgren K, Mazer SJ, McCabe GJ, McGill BJ, Parmesan C, Salamin N, Schwartz MD, Cleland EE (2012) Warming experiments underpredict plant phenological responses to climate change. Nature 485(7399):494–497. doi:10.1038/nature11014