

# Higher sample sizes and observer inter-calibration are needed for reliable scoring of leaf phenology in trees

Guohua Liu<sup>1,2</sup> | Isabelle Chuine<sup>3</sup> | Rémy Denéchére<sup>2,4</sup> | Frédéric Jean<sup>5</sup> |  
Eric Dufrêne<sup>2</sup> | Gaëlle Vincent<sup>2</sup> | Daniel Berveiller<sup>2</sup> | Nicolas Delpierre<sup>2,6</sup> 

<sup>1</sup>Jiangsu Key Laboratory of Agricultural Meteorology, Nanjing University of Information Science and Technology, Nanjing, China; <sup>2</sup>Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique et Evolution, Orsay, France; <sup>3</sup>Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), UMR 5554, CNRS, Montpellier, France; <sup>4</sup>Centre for Ocean Life, Technical University of Denmark, DTU Aqua, Lyngby, Denmark; <sup>5</sup>INRAE, UR629, Ecologie des Forêts Méditerranéennes (URFM), Domaine Saint Paul Site Agroparc, Avignon Cedex 9, France and <sup>6</sup>Institut Universitaire de France (IUF), Paris, France

## Correspondence

Nicolas Delpierre

Email: nicolas.delpierre@universite-paris-saclay.fr

Handling Editor: Amy Iler

## Abstract

1. Reliable phenological observations are needed to quantify the impact of climate change on tree phenology. Ground observations remain a prime source of phenological data, but their accuracy and precision have not been systematically quantified. The high subjectivity of ground phenological observations affects their accuracy, and the high within-population variability of tree phenology affects their precision. The magnitude of those effects is unknown to date.
2. We first explored the inter-observer variability in the timing of bud development and leaf senescence in trees using a unique dataset of seven observer inter-calibration sessions. Then, using tree phenological data collected in three European forests ( $n = 2,346$  observations for budburst,  $n = 539$  for leaf senescence), we quantified how the 'observer uncertainty' (accuracy of the observations) and the 'population sampling uncertainty' (precision of the observations) combine to affect the estimates of the budburst and the leaf senescence dates.
3. The median observer uncertainty was 8 days for budburst (BBCH = 7) and 15 days for leaf senescence (BBCH = 95). As expected, the population sampling uncertainty decreased with increasing sample size, and was about 6 days for budburst and 10 days for leaf senescence for a sample of 10 individuals monitored per population (corresponding to the median sample size in the phenological literature). As a whole, the overall uncertainty of phenological observations could reach up to 2 weeks for budburst and 1 month for leaf senescence.
4. *Synthesis.* This paper quantifies for the first time the accuracy and precision of ground phenological observations in forest trees and as such offers tables to estimate the uncertainty of phenological data. We show that reliable estimates of budburst and leaf senescence require three times ( $n = 30$ ) to two times ( $n = 20$ ) larger sample sizes as compared to sample sizes usually considered in phenological studies. We further call for an increased effort of observer inter-calibration, required to increase the accuracy of phenological observations. These recommendations reduce the uncertainty of phenological data, thereby improving the

estimation of phenological trends over time, the response of phenology to temperature or the inference of phenological model parameters.

#### KEYWORDS

accuracy, ground phenological observations, observer uncertainty, precision, sampling uncertainty

## 1 | INTRODUCTION

Monitoring and collecting time series of phenological events is of crucial importance to understand how ecosystems respond to climate, as the phenology of living organisms is strongly determined by environmental conditions, and highly responsive to the ongoing climate change (Menzel, 2013b; Parmesan & Yohe, 2003; Root et al., 2003). Most of the phenological data collected to date, with the earliest records having reached our epoch dating back to the 9th century (Aono & Kazui, 2008), were and are still obtained by visual observation from the ground (Chmielewski et al., 2013; Rutishauser et al., 2007). Most of the data currently used in phenological studies are of this kind, although remote sensing and phenological cameras are increasingly used to track the timing of phenological events (Ahrends et al., 2008; Graham et al., 2010; Richardson et al., 2007; Zhang et al., 2003). These ground observation data are collated in phenological databases consisting of hundreds of thousands observations (e.g. for trees >770,000 budburst and >380,000 leaf senescence observations in pep725.eu; >160,000 budburst and >150,000 leaf senescence observations in data.pheno.fr). Classically, plant phenological observations consist in dates of occurrence of key events, such as budburst, leaf unfolding, first flowering, full bloom, leaf coloration and leaf fall, which are scored visually. These events (or phenological stages) are point events in continuous sequences of phenological development or 'phenophases' (e.g. one would consider one tree to have reached budburst when 50% of its buds have opened, Supporting Information S1). Visual scoring depends on the point of view, observer's vision and the seeing of the observer (Bush et al., 2018). For example, the distance or position held from observers and size and colour of the object observed are likely to influence the visual scoring. If the phenological event is difficult to see (e.g. small buds, held high in the canopy and opening over a few days), it is likely to be scored differently by different observers. Moreover, this kind of ground observation is also sensitive to the personal experience of the observer, of the species and even the individual tree observed. Hence, the quality of the phenological observation is prone to subjective inaccuracy and depends heavily on the observational skills and effort of the observers. Thus, the ground observations are observer-biased (Kharin, 1976; Menzel, 2002; Schaber, 2002). Moreover, to boost current research that requires large amount of data obtained at large scales, that scientists cannot produce on their own in a reasonable time (Feldman et al., 2018), there is an increasing contribution of volunteers in collecting phenological observations worldwide (e.g. USA National Phenology Network, Budburst

Project, Phenoclim, Observatoire Des Saisons; Asse et al., 2018; Chandler et al., 2017; Dickinson et al., 2010; Fuccillo et al., 2015; Mayer, 2010). These phenological data are obtained by multiple observers and are thus naturally prone to observer bias. This observer bias (i.e. the *accuracy* of phenological date) has not been quantified to date, owing to the rarity of cross-observer campaigns.

Another prominent source of uncertainty in phenological observations is population sampling. Within a tree population, one usually observes a large variability in the timing of spring and autumn phenophases among individuals, due to the influences of microclimate and genetics (Cole & Sheldon, 2017; Delpierre et al., 2017; Doi et al., 2010; Rousi & Heinonen, 2007). This within-population variability of phenology is remarkable with a mean within-population amplitude (i.e. time interval between the earliest and latest observed individual) of 19 days and 26 days for spring and autumn phenology respectively (Delpierre et al., 2017), varying from year to year in relation to temperature conditions (Denéchère et al., 2021). For phenological observations at the population scale, it is usual to select a limited number of individuals to represent that population, and take the average date as representative of the phenological event. Obviously, the precision of the average date estimate depends on the sample size considered. Schwartz et al. (2013) reported that the phenological variations among individual trees in a specific microclimate can be adequately represented with a sample of 30 or more individuals. We made a literature survey of the number of individuals observed within one population in current tree phenological studies, and found the median number of sampled individual tree is around 10 (Table 1). This is much lower than 30, and the uncertainty associated with this sampling effect (i.e. the *precision* of phenological dates) remains unknown, though concerning most of the studies.

The uncertainty of phenological ground observations, resulting from the combination of low accuracy (due to inter-observer variability) and low precision (due to insufficient tree population sampling) may significantly affect the assessment of phenological responses to climate change. Indeed, to be detected, a trend signal must be higher than the noise (uncertainty) of the data, and trends in phenological dates are of a few days per decade (Menzel, 2013b). Consequently, estimating the uncertainty of phenological data would increase the robustness of our assessments of phenological changes.

In this study, we aimed at quantifying the uncertainty of phenological observation due to the variability among observers and to population sampling. To this aim, we used phenological observations coming from two different datasets: the cross-observer scoring of trees collected during seven inter-calibration sessions of observers,

**TABLE 1** Number of trees observed (i.e. sample size) per population, as reported in published papers

Phenophase	Sample size in one population	Reference
Budburst	10	Vitasse, Delzon, Dufrêne, et al. (2009)
	6–10	Hall et al. (2007)
	6–41	Vitasse, Delzon, Bresson, et al. (2009)
	27	Rousi and Heinonen (2007)
	3	Donnelly et al. (2017)
	12	Augspurger and Bartlett (2003)
	5	Fu et al. (2013)
	3	Richardson et al. (2006)
	3–10	Xie et al. (2018)
	1–45	Yu et al. (2016)
	70	Dantec et al. (2015)
	27	Delpierre et al. (2017)
	10–222	Chesnoiu et al. (2009)
	220	Scotti-Saintagne et al. (2004)
	36	Crawley and Akhteruzzaman (1988)
	38	Dongen et al. (1997)
	10	Wesołowski and Rowiński (2006)
	5	Basler and Körner (2014)
	3	Keenan and Richardson (2015)
	98	Fox et al. (1997)
	~20	Menzel (2013a)
	3–5	China Meterological Administration (1993)
	3–5	Archetti et al. (2013)
	10	Pérez-de-Lis et al. (2016)
	10	Laube et al. (2013)
	19–30	Capdevielle-Vargas et al. (2015)
	170	Kramer et al. (2008)
Leaf senescence	6–10	Hall et al. (2007)
	6–41	Vitasse, Delzon, Bresson, et al. (2009)
	3	Richardson et al. (2006)
	12	Augspurger and Bartlett (2003)
	1–45	Yu et al. (2016)
	3–5	Archetti et al. (2013)
	3	Keenan and Richardson (2015)
	27	Delpierre et al. (2017)
	30	Capdevielle-Vargas et al. (2015)
	36	Delpierre et al. (2009)
	3–5	China Meterological Administration (1993)

(Continues)

**TABLE 1** (Continued)

Phenophase	Sample size in one population	Reference
	3–10	Xie et al. (2018)
	5	Fu et al. (2014)
	4	Naschitz et al. (2014)
	3–12	Rosenthal and Camm (1996)

and observations of budburst and leaf senescence collected in large tree populations for several years. We specifically addressed four questions: (a) Does the inter-observer difference in phenological score vary for spring (continuous opening of buds in tree crowns) and autumn (continuous coloration and fall of leaves in tree crowns) phenophases? (b) How large is the uncertainty of the date of occurrence of these phenological stages due to the observer? (c) How large is the uncertainty due to sample size for tree population sampling? and (d) How large is the overall (i.e. observer plus population sampling) uncertainty in current phenological studies?

## 2 | MATERIALS AND METHODS

### 2.1 | Observer inter-calibration data (dataset #1)

This dataset was obtained during seven inter-calibration sessions of observers conducted in France during 2007–2017, three of which concerned budburst observations, and four of which concerned observation of leaf senescence (Table 2). The sessions took place in different locations each year and concerned different tree species. An average of 30 individual trees (from 12 to 49 trees) were observed by 10 to 37 observers each time following the same protocol and scoring scale (here a modified version of the BBCH scale originally published by (Meier, 1997) was used, see Table S1). The observers were scientists, technicians and doctoral students in biological sciences, and most of them participated all the sessions, which lasted one and a half day each. The inter-calibration sessions were organized as follows. At the first session, all participants were trained to the BBCH scoring scale and protocol. At all sessions, all participants first trained altogether on a set of tree individuals. Participants subsequently also scored another set of tree individuals independently. They used binoculars for observing tree crowns. Observers were instructed not to exchange information or discuss during this scoring in order to guarantee the independence of the scoring. The data obtained with this independent scoring were used for this study.

### 2.2 | Ground phenological observations in tree populations (dataset #2)

We used three sets of ground phenological observations (Table 3), conducted in different forests: (a) observations of two species (*Quercus petraea*, *Carpinus betulus*) in the Fontainebleau-Barbeau

Phenophase	Site (lat., lon.)	Year	Number of observers	Number of trees observed
Budburst	Vincennes (48.8°N, 2.4°E)	2007	12	35
	Antibes (43.6°N, 7.1°E)	2013	10	49
	Bordeaux (44.8°N, 0.6°W)	2014	22	12
Leaf senescence	Nogent-sur-Vernisson (47.8°N, 2.8°E)	2008	12	23
	Mont Ventoux (44.2°N, 5.3°E)	2011	27	44
	Nancy (48.7°N, 6.2°E)	2015	37	30
	Saint-Michel-l'Observatoire (43.9°N, 5.7°E)	2017	23	35

**TABLE 2** Characteristics of the observer inter-calibration sessions

**TABLE 3** Description of the phenological datasets acquired in tree populations

Site	Latitude (°N)	Longitude (°E)	Number of observers	Phenophase	Species	Observed year	Number of trees
Barbeau (France)	48.5	2.8	2	Budburst	<i>Carpinus betulus</i>	2013	28
					<i>Quercus petraea</i>	2013, 2015, 2016, 2017	25–66
			1	Leaf senescence	<i>Quercus petraea</i>	2015, 2016, 2017	24–31
Orsay (France)	48.7	2.2	3	Budburst	<i>Carpinus betulus</i>	2012, 2014, 2015	29–49
					<i>Castanea sativa</i>	2012, 2014, 2015	30–62
					<i>Quercus petraea</i>	2012, 2014, 2015	26–59
			1	Leaf senescence	<i>Carpinus betulus</i>	2011, 2012, 2014, 2015	30–46
					<i>Castanea sativa</i>	2012, 2014, 2015	28–56
					<i>Quercus petraea</i>	2011, 2012, 2014, 2015	31–58
Wytham Woods (UK)	51.8	1.3	5	Budburst	<i>Acer pseudoplatanus</i>	2013, 2014	109–111
					<i>Betula pendula</i>		93–96
					<i>Corylus avellana</i>		145–156
					<i>Fagus sylvatica</i>		69–71
					<i>Fraxinus excelsior</i>		194–195
					<i>Quercus robur</i>		196

forest (France) from 2013–2017 (Delpierre et al., 2016, 2017); (b) observations of three species (*Quercus petraea*, *Carpinus betulus*, *Castanea sativa*) in the Orsay forest (France) from 2012 to 2015 (Delpierre et al., 2017); (c) observations of six species (*Quercus robur*, *Fraxinus excelsior*, *Fagus sylvatica*, *Betula pendula*, *Corylus avellana* and *Acer pseudoplatanus*) in Wytham Woods (Oxfordshire, UK) in 2013 and 2014 (Cole & Sheldon, 2017). The three datasets share the particularity of documenting the whole sequence of spring budburst or leaf senescence at the individual tree scale. This means that each tree was monitored from winter state (all buds closed) to 100% budburst and from mature green summer leaves to all leaves coloured or fallen. In Fontainebleau-Barbeau and Orsay, the same tree

individuals were monitored during spring and autumn. The three sets of observations were each collected consistently by a small group (less than five, Table 3) of observers who have been trained and inter-calibrated before data collection, such that the observer variability in dataset #2 is minimal.

At the Fontainebleau-Barbeau and Orsay forests, the percentage (on a 10%-step scale) of open leaf buds (in the upper-third of the tree crown), leaf coloration and leaf fall (over the whole tree crown) for individual trees were monitored with binoculars two to three times a week from March to June and once a week from September to December. All observations on the 10%-step scale were converted to modified BBCH scores (Table S1) before

further processing. In spring, the budburst date of an individual tree was assumed to occur at the date when 50% buds were open (BBCH = 7). In autumn, the progress of leaf senescence was estimated by the combination of leaf coloration and leaf fall (Vitasse, Delzon, Bresson, et al., 2009). The leaf senescence date (BBCH = 95) occurred when 50% leaves were either coloured or fallen according to the following equation:

$$LS_t = \frac{LC_t \times (100 - LF_t)}{100} + LF_t, \quad (1)$$

where,  $LS_t$  is the percentage of leaf coloured or fallen in the selected tree at date  $t$ ,  $LC_t$  is the percentage of coloured leaf (i.e. non-green) in the selected tree at date  $t$  and  $LF_t$  is the percentage of leaf shed from the selected tree at date  $t$ . We mention here that the functional interpretation of this leaf senescence index, which is widely used in the phenological community, is not straightforward. Establishing links between phenological scores and the functioning of leaves is an active field of research (e.g. Bauerle et al., 2012; Märien et al., 2019).

At Wytham Woods, the development of buds was scored over whole tree crowns using a key of phenological stages (Hinks et al., 2015) different from BBCH, at a 3-day interval from March to May (Cole & Sheldon, 2017). A seven-stage key (stage 1 = small dormant buds, 2 = larger, slightly elongated buds, 3 = larger, loosened greenish brown buds, 4 = further elongated buds with leaves starting to erupt, 5 = leaves emerging but still tight, 6 = leaves loosening and extending outwards, 7 = leaves fully emerged and unfurled) was used for *Quercus robur* and a five-stage key (stage 1 = dormant bud, 2 = elongated/swollen buds, 3 = budburst, 4 = leaves emerging and extending outwards, 5 = leaves fully emerged and unfurled) used for the other species.

The scale used for the observation of budburst in the Fontainebleau-Barbeau and Orsay forests differed with that used in Wytham Woods. Indeed, bud observations in the Fontainebleau-Barbeau or Orsay forests started from the first signs of buds opening

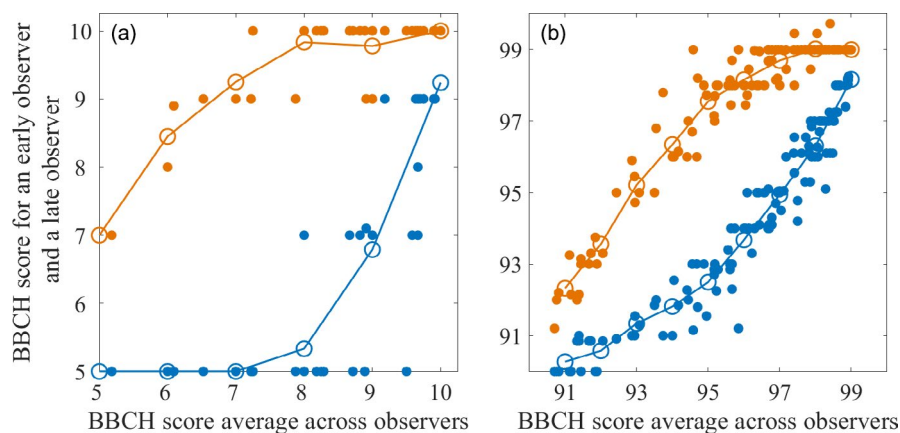
(0% buds opened), and ended when 100% buds were open, corresponding to the interval from stage 4 to stage 7 using the seven-stage key and stage 3 to stage 5 using the five-stage key in Wytham Woods. In order to homogenize protocols, we scaled the Wytham Woods observations of budburst to a 0%–100% scale similar to the one used in Fontainebleau-Barbeau and Orsay, setting stage 3 (or 4 depending on the species) of the Wytham Woods scale to 0% and setting stage 5 (or 7, depending on the species) to 100%. Thereafter all data were converted to the modified BBCH scale (Supporting Information S1) to match the scale used during the intercalibration sessions of observers.

We ensured that the tree individual phenological sequences collected at Fontainebleau-Barbeau, Orsay and Wytham woods were monotonically increasing with time. If a phenological observation record was less than the previous record, we corrected it to be equal to the previous record (affects 0.25%/7.8% of all budburst/leaf senescence data). Last, the actual phenological observations conducted 2–3 times per week (budburst) to once a week (leaf senescence) were linearly interpolated at a daily time scale, in order to construct daily phenological observation time series (Figure 2, black lines).

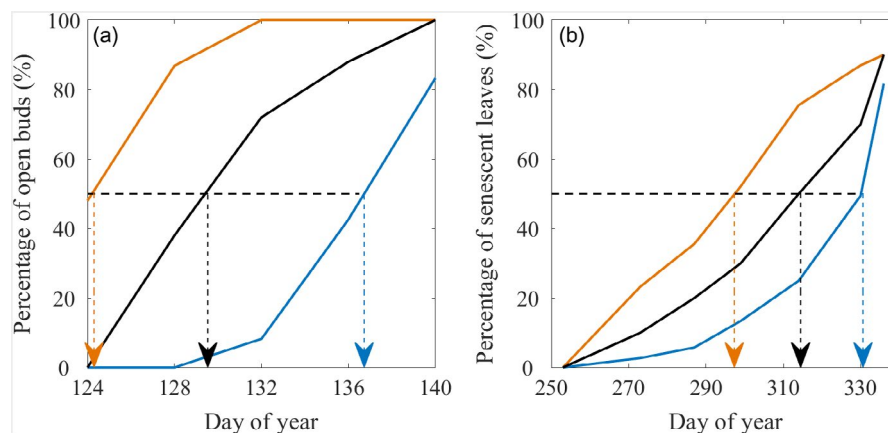
## 2.3 | Estimation of the uncertainty due to the observer

### 2.3.1 | Identifying 'early' and 'late' observers during inter-calibration sessions

During each inter-calibration session, observers estimated the phenological stage of a given number of individual tree crowns (see Table 2) using the modified BBCH scale. For each of those trees, we computed the average, 5th centile and 95th centile (Figure 1) of BBCH scores across observers. The 5th centile (blue points on Figure 1) corresponds to low BBCH scores attributed to a tree individual. Such low scores originate from observers prone to report



**FIGURE 1** Among-observer variability of phenological observations during inter-calibration sessions for budburst (a) and leaf senescence (b). The y-axis represents the 5th centile (blue, produced by a 'late observer') and 95th centile (orange, produced by an 'early observer') of the among-observer distribution of BBCH score established during the inter-calibration sessions. The x-axis represents the average BBCH scores of all observers. The orange and blue line and open circles represent the binned BBCH scores of early and late observers (binning was done per unit step on the BBCH scale)



**FIGURE 2** Estimating the date of budburst (a) and leaf senescence (b) in an individual tree from the point of view of an 'early' or a 'late' observer. The black line is the actual development process in the tree crown; the orange (blue) line is the estimation of development process in the tree crown for early (late) observers. The black arrow points to the actual observed budburst/leaf senescence date; orange and blue arrows point to the budburst/leaf senescence date observed by an 'early' and 'late' observer respectively

events late (called hereafter 'late observers'). The 95th centile (orange points on Figure 1) corresponds to high BBCH scores attributed to an individual tree. Such high scores originate from observers prone to report events early (hereafter 'early observers'). From these data, we constructed a conversion rule between the 'true' (we assumed the average of actual observations to be the true observation) phenological state (x-axis on Figure 1) and the phenological state reported by early and late observers (y-axis on Figure 1).

### 2.3.2 | Observer uncertainty: From inter-calibration sessions to scoring in tree populations

We applied the among-observer variability quantified during inter-calibration sessions (from dataset #1, see 2.3.1) on ground phenological data collected in tree populations (dataset #2) to construct artificial phenological sequences that would be produced either by a late or by an early observer (Figure 2). We then determined, for the date of budburst by an early (late) observer when the 'early (late) observer' phenological sequence reached 50% buds opened (BBCH = 7; orange (blue) arrow on Figure 2). We defined the difference, in days, between the early and the late observer dates as the maximum observer uncertainty.

We calculated the observer uncertainties for all individual trees observed in Fontainebleau-Barbeau, Orsay and Wytham Woods forests, and we estimated the density distributions of the observer uncertainty for each population-year<sup>1</sup> and all population-years using nonparametric Kernel smoothing (Bowman, 1984). In order to explore the observer uncertainty further, the budburst (BBCH = 7) and leaf senescence (BBCH = 95) speeds of each individual tree in the three forests were computed as the linear slope of the front and back two points closest to the point at BBCH = 7 or 95. The speed was measured in percentage of phenological development per day.

### 2.4 | Estimation of the population sampling uncertainty

In a given tree population of  $N$  individuals, we sampled  $n$  (from 1 to  $N$ ) individuals and calculated the sampling uncertainty in determining a phenological date associated to this sample size. To this aim, we randomly picked a sample of  $n$  individuals out of the  $N$  individuals in the population, and we calculated the mean, for example, budburst date for this sample. We repeated the sampling of  $n$  individuals 100 times. This resulted in 100 average budburst dates obtained for a sample size of  $n$ . We calculated the 95th percentile (called 'A') and 5th percentile (called 'B') of these 100 budburst dates. We defined the difference between the 95th percentile date and 5th percentile date (A-B) as the uncertainty of the budburst date related to the population sampling (i.e. this is the time interval, in days, containing 90% of the estimates of the average budburst date established over a number  $n$  of trees sampled in the population). We calculated the population sampling uncertainty for budburst in Wytham Woods (six species in 2013 and 2014) and for leaf senescence of *Carpinus betulus* (2011), *Castanea sativa* (2011, 2012) and *Quercus petraea* (2011, 2012) in Orsay, given that these populations have data for a large number of individual trees (Table 3), that is, large enough to be considered representative of the true average date over the tree population.

### 2.5 | Estimation of the overall uncertainty

We also estimated the overall uncertainty, combining the uncertainties due to the observer and the populations sampling. For this, we proceeded similarly to our estimation of the population sampling uncertainty. First, for a sample size of  $n$  trees, we computed 100 mean, for example, budburst dates based on data observed by an early or by a late observer. Then we averaged those 100 mean budburst dates for early (called 'C') and late (called 'D') observers. The difference between the averaged budburst dates for early observers

<sup>1</sup>A 'population-year' refers to one tree population being observed during 1 year.



and that for late observers (D–C) was defined as the observer uncertainty for a sample size of  $n$  trees. Since both the observer (D–C) and population sampling (A–B) uncertainties are independent (Schaber, 2002), we added them to quantify the overall uncertainty of phenological observations. The overall uncertainties were calculated at the same site-years as the sampling uncertainties.

One aim of this work was to provide a quantification of the uncertainty of phenological observations for its incorporation in the analysis of phenological data. To this aim, we calculated the 5th, 25th, 50th, 75th, 95th percentiles of the sample sizes found in the literature survey summarized in Table 1 which were used to establish populations' average dates of budburst and leaf senescence (Table 4). We estimated the sampling and total uncertainties of phenological dates for sample sizes corresponding to these percentiles.

### 3 | RESULTS

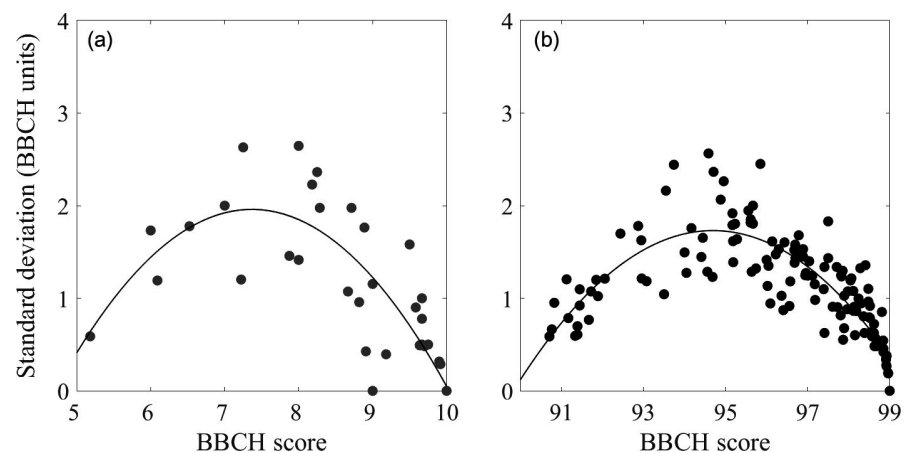
#### 3.1 | The variability of phenological observations among observers varies with phenological development stages

The uncertainties on bud development (BBCH from 5 to 10) and leaf senescence development (BBCH from 90 to 99) observed by multiple observers during inter-calibration sessions are shown in Figure 3. The mean standard deviation in the beginning (BBCH around 5 or 90) and

**TABLE 4** Percentiles of sample size for budburst and leaf senescence in current phenological studies of trees. The percentiles are derived from the sample sizes of natural tree populations appearing in Table 1. The approximate values of sample size are shown in parentheses

Percentile (%)	Sample size for budburst	Sample size for leaf senescence
5	3	3
25	5	5
50	12 (~10)	8 (~10)
75	29 (~30)	27 (~30)
95	41 (~40)	39 (~40)

**FIGURE 3** The standard deviation of budburst (a) and leaf senescence (b) scoring established among observers during inter-calibration sessions. The black line is a quadratic fitting curve based on the observations, plotted for visualization purpose



the end (BBCH around 10 or 99) of the bud development and leaf senescence sequence was low, while it was high in the middle stage (i.e. BBCH around 7 or 95; Figure 3). For budburst, the mean standard deviation could reach up to 2.0 BBCH units at BBCH = 7; but the standard deviations were near zero for trees with few (BBCH = 5) or all buds open (BBCH = 10; Figure 3a). Similarly, observers differed most in the estimation of intermediate stage of leaf senescence (BBCH = 95) with a standard deviation of scoring of 1.7 BBCH units. They tended to agree more for trees just entering senescence (BBCH = 91) or fully senescent (BBCH = 99; Figure 3b). In conclusion, the phenological observation variation among observers varied along phenophases: Observers tended to reach an agreement on early and late stages, and differed the most for intermediate stages.

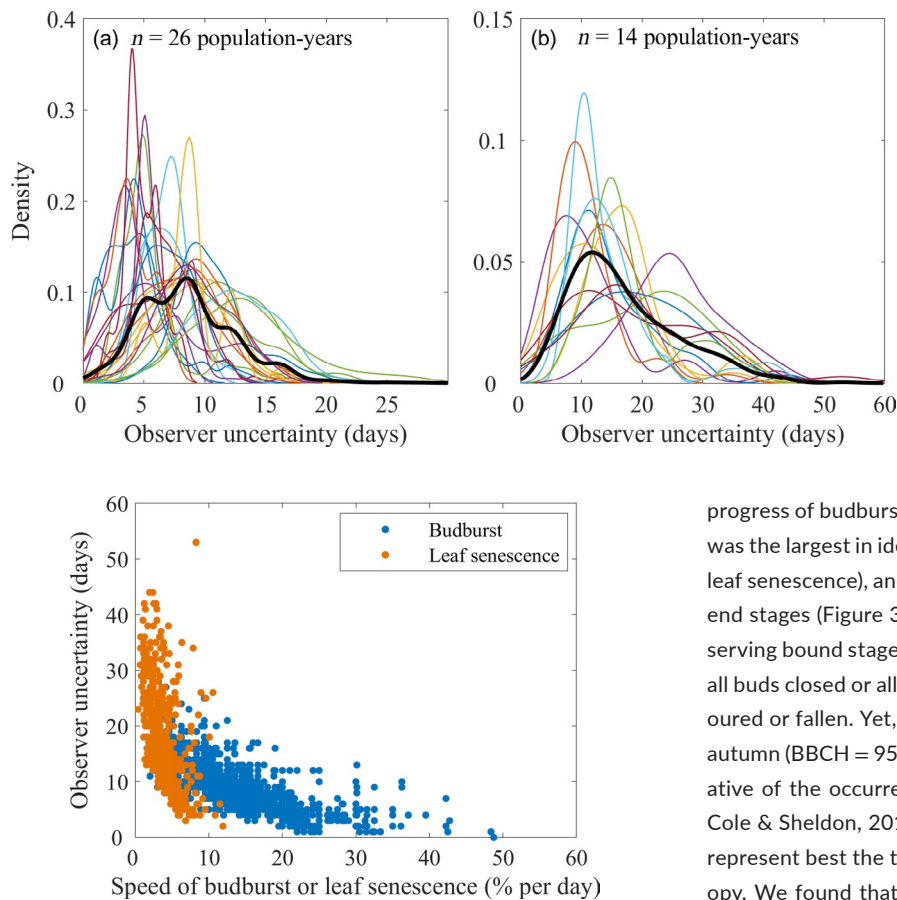
#### 3.2 | Uncertainty on budburst and leaf senescence observations due to the observer

In Figure 4, we report the densities of observer uncertainties calculated for each population-year in our dataset. The observer uncertainty varied among individual trees for a given population-year. The observer uncertainty in determining the date of budburst of a particular tree ranged from zero to 28 days, with a median of 8 days (Figure 4a). For leaf senescence, the observer uncertainty ranged from 2 to 53 days with a median of 15 days (Figure 4b).

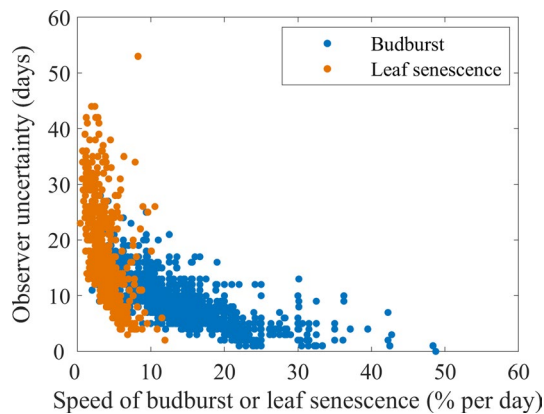
The observer uncertainties were significantly related to the speed of development of budburst and leaf senescence (Figure 5), with a rank correlation coefficient of  $-0.78$  ( $p < 0.05$ ). Thus, faster development rates of budburst or leaf senescence generated smaller observer uncertainties.

#### 3.3 | Population sampling and overall uncertainties of phenological observations

The population sampling and overall uncertainties at the population scale varied with the number of sampled trees (Figure 6). With the increasing number of sampled trees, both the population sampling and overall uncertainties of phenological observations decreased for



**FIGURE 4** Density distribution of observer uncertainty in determining budburst (BBCH 7) (a) and leaf senescence (BBCH 95) (b) in different population-years. Each coloured line represents one population-year, and the black line is the kernel density overall population-years. Please note differences on the x-axis scale of panels (a) and (b)



**FIGURE 5** Relationship between observer uncertainty and phenological development speed for budburst and leaf senescence

budburst (Figure 6a) and leaf senescence (Figure 6b). When observing three individual trees of a population, the population sampling uncertainty was about 10.5 days and 19.1 days for budburst and leaf senescence, respectively. Adding the observer uncertainty, the overall uncertainty reached up to 19.7 days (budburst) and 37.6 days (leaf senescence). But when the number of observed trees reached 40, the population sampling and overall uncertainties respectively dropped to 2.5 days and 11.6 days for budburst, and 2.5 days and 21.0 days for leaf senescence. Current tree phenological studies use a median of 10 individual trees to determine the average phenological date of a population (Table 4). According to our calculations, the sampling uncertainty for this sample size is 5.6 days for budburst, and 9.6 days for leaf senescence. Taking further into account the effect of the observer, the overall uncertainty reaches 14.8 days for budburst (Figure 6a) and 28.0 days for leaf senescence for this sample size (Figure 6b).

## 4 | DISCUSSION

### 4.1 | The observer uncertainty of phenological observations

We used data collected during seven phenological inter-calibration sessions to quantify the among-observer variability in evaluating the

progress of budburst and leaf senescence. The observer uncertainty was the largest in identifying intermediate stages (50% buds open or leaf senescence), and the lowest in identifying the beginning and the end stages (Figure 3). It is reasonable to imagine a consensus in observing bound stages of the spring and autumn phenophases, that is, all buds closed or all buds open, and all leaves green or all leaves coloured or fallen. Yet, the middle stages of the spring (BBCH = 7) and autumn (BBCH = 95) phenophases are usually selected as representative of the occurrence of these phenophases (Bruns et al., 2003; Cole & Sheldon, 2017; Jochner et al., 2013), probably because they represent best the transitions from dormant to fully developed canopy. We found that the observer uncertainties for individual trees were about 8 days for budburst at BBCH = 7 and 15 days for leaf senescence at BBCH = 95 (Figure 4).

This 'observer uncertainty' has multiple causes. First, the interpretation of phenological protocols often leaves room for subjectivity (Schaber, 2002). Second, there are real difficulties in observing the target morphological changes at the bud/leaf scale (i.e. buds opening, coloration of leaves) at a distance of several (tens of) meters in the tree crown, even with binoculars, or quantifying those changes at the scale of a tree crown (e.g. try to distinguish a 40% from a 50% opening of buds throughout a tree crown). These difficulties increase the probability of discrepancies among observers.

Interestingly, we observed lower observer uncertainties on budburst, as compared to leaf senescence. First, the spreading of budburst within a tree crown occurs more rapidly (budburst progressed at an average speed of 14% per day, Figure 5) than the coloration and shedding of leaves (leaf senescence progressed at an average rate of 4% per day), and we have found that uncertainty decreased with the speed of a phenophase (Figure 5). Second, the higher uncertainties in the evaluation of leaf senescence might stem from the higher subtlety of morphological and colour changes occurring during leaf senescence. Indeed, the observation of budburst requires distinguishing the emergence of leaf tips from the bud (form change) when the observation of leaf senescence requires quantifying both changes in leaf colour (that human eye perceives less precisely than form changes, Chuine & Régnière, 2017) and the percentage of leaves remaining in the canopy (Equation 1). Overall, this results in a lower quality of autumn phenology data, as compared to spring data,



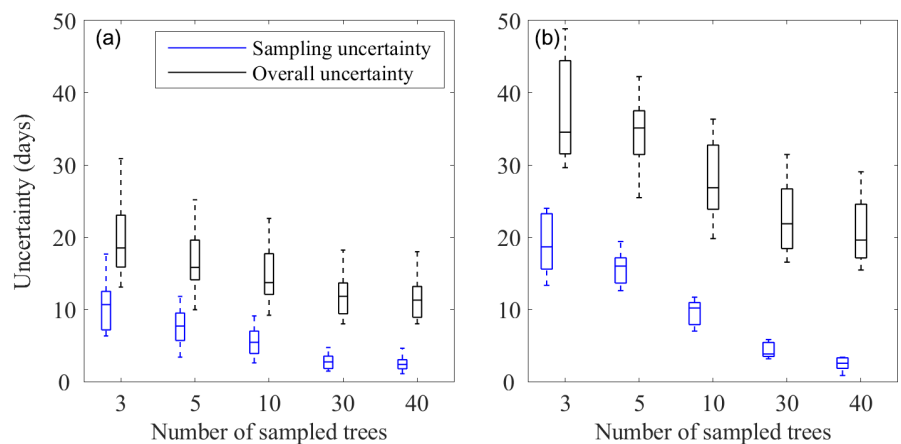
which may help to explain the less-accurate performance of autumn phenological models, when compared to spring phenological models (Delpierre et al., 2009; Vitasse et al., 2011; Yang et al., 2012).

## 4.2 | The population sampling uncertainty of phenological observation

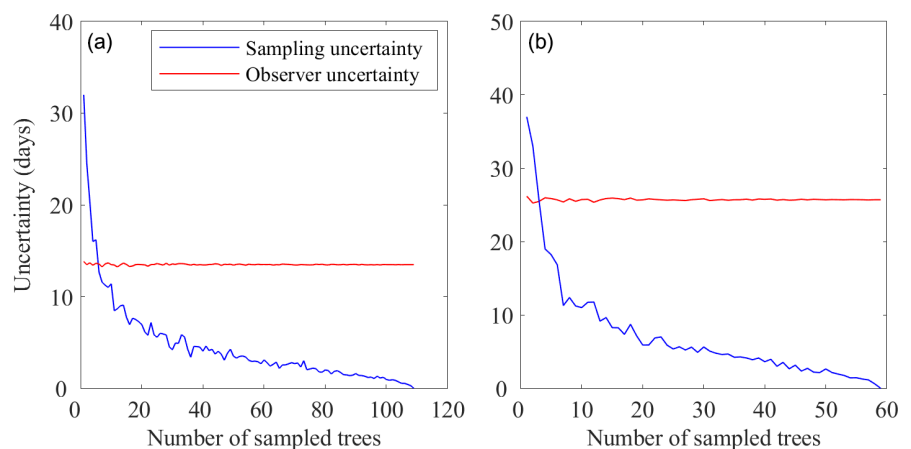
As expected from basic statistical rules, the population sampling uncertainty of phenological observations decreased with the increasing sample size (Figure 6). We found that a minimum sample size of 29 (budburst) or 18 (leaf senescence) individuals (on average for all population-years) was necessary to estimate the population average date of the budburst or leaf senescence with an uncertainty of 3 (budburst) or 7 (leaf senescence) days, corresponding to the usual time resolution of phenological observation for these phenophases. Compared with these 'ideal' sample sizes, we found that about 76% of the budburst and 81% of leaf senescence studies collated in our literature survey used lower sample sizes (Table 1), with a median sample size of 10 individuals (Table 4). We thus suggest that researchers double (from 10 to 20) for leaf senescence and triple (from 10 to 30) for budburst, the sample sizes used in phenological studies. For budburst, our recommendation is in line

with Schwartz et al. (2013) who proposed  $n = 30$  as an adequate sample size. The population sampling uncertainty for a sample size of 10 was about 5.6 (~6) days for budburst and 9.6 (~10) days for leaf senescence. Obviously, the population sampling uncertainty related to the within-population variability of the phenological events owes to micro-environmental and genetic variations among trees (Cole & Sheldon, 2017; Delpierre et al., 2017; Denéchére et al., 2021; Rousi & Heinonen, 2007).

The combination of inter-observer and sampling uncertainty yielded overall uncertainties of 2 weeks (14.8 days) for budburst and 1 month (28.0 days) for leaf senescence, considering a sample size of 10 individuals. The overall uncertainty for leaf senescence is nearly twice larger than that for budburst, implying again that spring phenological observations are more reliable than autumn phenological observations. Furthermore, and logically, the overall uncertainties of phenological observations decreased with the number of trees sampled. The observer uncertainty did not change with the sample size, contrary to the population sampling uncertainty (Figure 7). Hence, when the sample size was small, the population sampling uncertainty formed the largest part of the uncertainty; while the observer uncertainty became the main contributor to the overall uncertainty when the sample size increased (Figure 7).



**FIGURE 6** The population sampling and overall uncertainties for different tree sample sizes, established for budburst (a) and leaf senescence (b)



**FIGURE 7** Examples of observer uncertainty and population sampling uncertainty varying with the observed number of sampled trees for the budburst date of *Acer pseudoplatanus* in 2014 at Wytham Woods (a) and leaf senescence date of *Quercus petraea* in 2011 at Orsay (b)

### 4.3 | Suggestions on how to reduce the uncertainty of phenological observations

We demonstrated that ground phenological observations of tree phenology are subject to high uncertainties. However, the influence of such uncertainties on the calculation of phenological trends with time, or in establishing the response of phenological events to environmental variability remains to be quantified. We do not mean that all ground phenological data should be tarnished by an overall uncertainty of 14.8 days for budburst or 28.0 days for leaf senescence. These are maximum indicative uncertainty values that do not apply in the case of phenological data acquired over years by the same observer and over the same, be it small, number of trees, as long as they are representative of the population distribution of phenological dates (i.e. no bias to early or late individuals; Miller-Rushing et al., 2008). This is because (a) we expect an individual observer to be consistent with their own observations across years and (b) trees have a 'phenological identity' that repeats from year to year (early versus. intermediate versus. late trees; Delpierre et al., 2017). Yet, these uncertainty estimates should be considered when working with phenological data acquired over varying tree individuals across time and/or by different, not cross-checked, observers.

Other means to obtain accurate phenological data do exist beside ground observation. Phenological indexes obtained from remote sensing, digital webcam or unmanned aerial vehicles (UAVs) may help solve the problem of observation subjectivity. But they have their own caveats, such as issues with the signal processing and the establishment of correspondences with ground phenological observations. Moreover, the challenge of mixed vegetation types remains a major limitation for remote sensing (Richardson, Hufkens, Milliman, & Froking, 2018) and data derived from satellite still depend on ground-based measurements for validation (Fisher & Mustard, 2007; Hmimina et al., 2013; Soudani et al., 2008; Testa et al., 2018). The use of digital webcam and UAVs (Klosterman et al., 2018; Berra et al., 2019) to generate phenological data is still developing and data records are short as compared to ground phenological observations. Thus, to date, ground phenological observations remain central for the quantification and understanding of tree phenology. This is why the phenological community should make efforts to reduce the uncertainty of phenological observations in future observations campaigns. To this aim, we propose the following recommendations:

1. Converge to standardized and illustrated protocols in order to reduce the subjectivity of phenological observations (see e.g. Denny et al. (2014) for a detailed example). In order to minimize the inter-observer variability in the interpretation of protocols, very detailed and illustrated (with photographs) protocols should be used. Ideally, a unique phenological scale should be used in order to allow for cross-comparison across networks of observers. This is why the BBCH scale (Meier, 1997) was designed

for herbaceous and tree crops, but its adaptation to forest trees is missing to date. An adaptation of the BBCH scale to forest trees would further require addressing the question of upscaling: Indeed, scoring tree phenology requires scaling to the tree crown of observations conducted at the scale of the organ (e.g. what percentage of all buds present in the crown has reached a desired development stage).

2. Promote the combination of camera-based with ground phenological observations. The objectivity and repeatability of the processing of camera images could help to constrain the subjectivity of human eye appreciations of phenological stages. Digital cameras have been used successfully in recent years for tracking phenological events (Klosterman et al., 2014; Richardson, Hufkens, Milliman, Aubrecht, et al., 2018; Zhang et al., 2018). This technology could provide a valuable support for ground phenological observations provided increased efforts on the comparability of both approaches. One way to proceed is to systematize ground observation campaigns at sites equipped with phenological cameras (e.g. Delpierre et al., 2020; Keenan et al., 2014).
3. Train observers and inter-calibrate them. Training the observers prior to data collection is critical to reduce the observer bias, which reached 8 days for budburst and 15 days for leaf senescence, and was the main source of phenological uncertainty (Figures 6 and 7) in our dataset acquired by professional scientists. Organizing inter-calibration sessions can be challenging if a large number of observers is concerned, but it is worth the effort. This is especially important for citizen science programs where limited training can lead to poor and unreliable data quality (Foster-Smith & Evans, 2003; Kosmala et al., 2016; MacKenzie et al., 2017), yet the return on investment is large since data recorded by trained citizen scientists are usually more accurate and precise than untrained professional observers (Fuccillo et al., 2015).
4. An adequate number of individual trees must be monitored to ensure adequate sampling of a population. In order to reduce the sampling bias, future phenological observations should double or triple the number of trees observed. On the basis of our results, we recommend a minimum of about 30 trees for budburst and 20 trees for leaf senescence in order to get a precision of the same order as the usual time resolution of observations (3 days for spring, 7 days for autumn). We have not stressed here the role of the sampling frequency, but it is clear that degrading the sampling interval from twice a week in spring / once a week in autumn impacts the estimate of phenological dates (Miller-Rushing et al., 2008; Schwartz et al., 2013).

### ACKNOWLEDGEMENTS

The authors thank all contributors to the observers' intercalibration campaigns organized by the metaprogramme ACCAF Perpheclim and SOERE TEMPO (<https://tempo.pheno.fr>). We thank Eleanor Cole (Department of Zoology, University of Oxford) for sharing the Wytham woods dataset. We thank David Inouye and an anonymous

reviewer for their helpful comments on the manuscript. Guohua Liu also appreciates the CSC (China Scholarship Council) for a doctoral fellowship to do this work in Université Paris-Sud. We acknowledge support from The Startup Foundation for Introducing Talent of NUIST. The authors declare no conflict of interest.

## AUTHORS' CONTRIBUTIONS

N.D. conceived the ideas and designed the methodology; I.C. initiated and designed the inter-comparison of phenological observations among observers; N.D., E.D., G.V., and D.B. collected phenological data at Orsay and Fontainebleau-Barbeau; G.L., N.D. and R.D. analysed the data; G.L. wrote the first version of the manuscript. All authors contributed critically to interpret the results and gave final approval for publication.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/1365-2745.13656>.

## DATA AVAILABILITY STATEMENT

The data used in this paper are accessible through the Dryad Digital Repository <https://doi.org/10.5061/dryad.1g1jwstvx> (Delpierre et al., 2021).

## ORCID

Nicolas Delpierre  <https://orcid.org/0000-0003-0906-9402>

## REFERENCES

- Ahrends, H. E., Brügger, R., Stöckli, R., Schenk, J., Michna, P., Jeanneret, F., Wanner, H., & Eugster, W. (2008). Quantitative phenological observations of a mixed beech forest in northern Switzerland with digital photography. *Journal of Geophysical Research: Biogeosciences*, 113, 1–11. <https://doi.org/10.1029/2007jg000650>
- Aono, Y., & Kazui, K. (2008). Phenological data series of cherry tree flowering in Kyoto, Japan, and its application to reconstruction of springtime temperatures since the 9th century. *International Journal of Climatology*, 28, 905–914. <https://doi.org/10.1002/joc.1594>
- Archetti, M., Richardson, A. D., O'Keefe, J., & Delpierre, N. (2013). Predicting climate change impacts on the amount and duration of autumn colors in a New England forest. *PLoS ONE*, 8, e57373. <https://doi.org/10.1371/journal.pone.0057373>
- Asse, D., Chuine, I., Vitasse, Y., Yoccoz, N. G., Delpierre, N., Badeau, V., Delestrade, A., & Randin, C. F. (2018). Warmer winters reduce the advance of tree spring phenology induced by warmer springs in the Alps. *Agricultural and Forest Meteorology*, 252, 220–230. <https://doi.org/10.1016/j.agrformet.2018.01.030>
- Augsburger, C. K., & Bartlett, E. A. (2003). Differences in leaf phenology between juvenile and adult trees in a temperate deciduous forest. *Tree Physiology*, 23, 517–525. <https://doi.org/10.1093/treephys/tps/23.8.517>
- Basler, D., & Körner, C. (2014). Photoperiod and temperature responses of bud swelling and bud burst in four temperate forest tree species. *Tree Physiology*, 34, 377–388. <https://doi.org/10.1093/treephys/tpu021>
- Bauerle, W. L., Oren, R., Way, D. A., Qian, S. S., Stoy, P. C., Thornton, P. E., Bowden, J. D., Hoffman, F. M., & Reynolds, R. F. (2012). Photoperiodic regulation of the seasonal pattern of photosynthetic capacity and the implications for carbon cycling. *Proceedings of the National Academy of Sciences of the United States of America*, 109(22), 8612–8617. <https://doi.org/10.1073/pnas.1119131109>
- Berra, E. F., Gaulton, R., & Barr, S. (2019). Assessing spring phenology of a temperate woodland: A multiscale comparison of ground, unmanned aerial vehicle and Landsat satellite observations. *Remote Sensing of Environment*, 223, 229–242.
- Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71, 353–360. <https://doi.org/10.2307/2336252>
- Bruns, E., Chmielewski, F.-M., & VanVliet, A. J. H. (2003). The global phenological monitoring concept. In M. Schwartz (Ed.), *Phenology: An integrative environmental science* (pp. 93–104). Springer.
- Bush, E. R., Bunnefeld, N., Dimoto, E., Dikangadissi, J. T., Jeffery, K., Tutin, C., White, L., & Abernethy, K. A. (2018). Towards effective monitoring of tropical phenology: Maximizing returns and reducing uncertainty in long-term studies. *Biotropica*, 50, 455–464. <https://doi.org/10.1111/btp.12543>
- Capdevielle-Vargas, R., Estrella, N., & Menzel, A. (2015). Multiple-year assessment of phenological plasticity within a beech (*Fagus sylvatica* L.) stand in southern Germany. *Agricultural and Forest Meteorology*, 211, 13–22. <https://doi.org/10.1016/j.agrformet.2015.03.019>
- Chandler, M., See, L., Copas, K., Bonde, A. M. Z., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A., & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, 213, 280–294. <https://doi.org/10.1016/j.biocon.2016.09.004>
- Chesnoiu, E. N., Șofletea, N., Curtu, A. L., Toader, A., Radu, R., & Enescu, M. (2009). Bud burst and flowering phenology in a mixed oak forest from Eastern Romania. *Annals of Forest Research*, 52, 199–206. <https://doi.org/10.15287/afr.2009.136>
- China Meteorological Administration. (1993). Phenological observation (in Chinese). In *Observation criterion of agricultural meteorology (in Chinese)* (pp. 131–164). China Meteorological Press.
- Chmielewski, F.-M., Heider, S., Moryson, S., & Bruns, E. (2013). International phenological observation networks: Concept of IPG and GPM. In M. Schwartz (Ed.), *Phenology: An integrative environmental science* (pp. 137–153). Springer.
- Chuine, I., & Régnière, J. (2017). Process-based models of phenology for plants and animals. *Annual Review of Ecology, Evolution, and Systematics*, 48, 159–182. <https://doi.org/10.1146/annurev-ecolsys-110316-022706>
- Cole, E. F., & Sheldon, B. C. (2017). The shifting phenological landscape: Within- and between- species variation in leaf emergence in a mixed-deciduous woodland. *Ecology and Evolution*, 7, 1135–1147. <https://doi.org/10.1002/ece3.2718>
- Crawley, M. J., & Akhteruzzaman, M. (1988). Individual variation in the phenology of oak trees and its consequences for herbivorous insects. *Functional Ecology*, 2, 409–415. <https://doi.org/10.2307/2389414>
- Dantec, C. F., Ducasse, H., Capdevielle, X., Fabreguettes, O., Delzon, S., & Desprez-Loustau, M. L. (2015). Escape of spring frost and disease through phenological variations in oak populations along elevation gradients. *Journal of Ecology*, 103, 1044–1056. <https://doi.org/10.1111/1365-2745.12403>
- Delpierre, N., Chuine, I., & Cole, E. (2021). Tree phenology - observer intercalibration and individual tree phenological scoring [Dataset]. *Dryad*, <https://doi.org/10.5061/dryad.1g1jwstvx>
- Delpierre, N., Dufrêne, E., Soudani, K., Ulrich, E., Cecchini, S., Boe, J., & Francois, C. (2009). Modelling interannual and spatial variability of leaf senescence for three deciduous tree species in France. *Agricultural and Forest Meteorology*, 149, 938–948. <https://doi.org/10.1016/j.agrformet.2008.11.014>
- Delpierre, N., Guillemot, J., Dufrêne, E., Cecchini, S., & Nicolas, M. (2017). Tree phenological ranks repeat from year to year and correlate with growth in temperate deciduous forests. *Agricultural*

- and Forest Meteorology, 234, 1–10. <https://doi.org/10.1016/j.agrformet.2016.12.008>
- Delpierre, N., Soudani, K., Berveiller, D., Dufrêne, E., Hmimina, G., & Vincent, G. (2020). 'Green pointillism': Detecting the within-population variability of budburst in temperate deciduous trees with phenological cameras. *International Journal of Biometeorology*, 64, 663–670. <https://doi.org/10.1007/s00484-019-01855-2>
- Delpierre, N., Vitasse, Y., Chuine, I., Guillemot, J., Bazot, S., Rutishauser, T., & Rathgeber, C. B. K. (2016). Temperate and boreal forest tree phenology: From organ-scale processes to terrestrial ecosystem models. *Annals of Forest Science*, 73, 5–25. <https://doi.org/10.1007/s13595-015-0477-6>
- Denéchère, R., Delpierre, N., Apostol, E. N., Berveiller, D., Bonne, F., Cole, E., Delzon, S., Dufrêne, E., Gressler, E., Jean, F., Lebourgeois, F., Liu, G., Louvet, J.-M., Parmentier, J., Soudani, K., & Vincent, G. (2021). The within-population variability of leaf spring and autumn phenology is influenced by temperature in temperate deciduous trees. *International Journal of Biometeorology*, 65, 369–379. <https://doi.org/10.1007/s00484-019-01762-6>
- Denny, E. G., Gerst, K. L., Miller-Rushing, A. J., Tierney, G. L., Crimmins, T. M., Enquist, C. A. F., Guertin, P., Rosemartin, A. H., Schwartz, M. D., Thomas, K. A., & Weltzin, J. F. (2014). Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications. *International Journal of Biometeorology*, 58, 591–601. <https://doi.org/10.1007/s00484-014-0789-5>
- Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41, 149–172. <https://doi.org/10.1146/annurev-ecolsys-102209-144636>
- Doi, H., Takahashi, M., & Katano, I. (2010). Genetic diversity increases regional variation in phenological dates in response to climate change. *Global Change Biology*, 16, 373–379. <https://doi.org/10.1111/j.1365-2486.2009.01993.x>
- Dongen, S. V., Backeljau, T., Matthysen, E., & Dhondt, A. A. (1997). Synchronization of hatching date with budburst of individual host trees (*Quercus robur*) in the winter moth (*Operophtera brumata*) and its fitness consequences. *Journal of Animal Ecology*, 66, 113–121. <https://doi.org/10.2307/5969>
- Donnelly, A., Yu, R., Caffarra, A., Hanes, J., Liang, L., Desai, A. R., Liu, L., & Schwartz, M. D. (2017). Interspecific and interannual variation in the duration of spring phenophases in a northern mixed forest. *Agricultural and Forest Meteorology*, 243, 55–67. <https://doi.org/10.1016/j.agrformet.2017.05.007>
- Feldman, R. E., Zemaite, I., & Miller-Rushing, A. J. (2018). How training citizen scientists affects the accuracy and precision of phenological data. *International Journal of Biometeorology*, 62, 1421–1435. <https://doi.org/10.1007/s00484-018-1540-4>
- Fisher, J. I., & Mustard, J. F. (2007). Cross-scalar satellite phenology from ground, Landsat, and MODIS data. *Remote Sensing of Environment*, 109, 261–273. <https://doi.org/10.1016/j.rse.2007.01.004>
- Foster-Smith, J., & Evans, S. M. (2003). The value of marine ecological data collected by volunteers. *Biological Conservation*, 113, 199–213. [https://doi.org/10.1016/S0006-3207\(02\)00373-7](https://doi.org/10.1016/S0006-3207(02)00373-7)
- Fox, C. W., Waddell, K. J., Grockers, F. R., & Mousseau, T. (1997). Variation in budbreak phenology affects the distribution of a leafmining beetle (*Brachys tessellatus*) on turkey oak (*Quercus laevis*). *Ecoscience*, 4, 480–489. <https://doi.org/10.1080/11956860.1997.11682427>
- Fu, Y., Campioli, M., Deckmyn, G., & Janssens, I. A. (2013). Sensitivity of leaf unfolding to experimental warming in three temperate tree species. *Agricultural and Forest Meteorology*, 181, 125–132. <https://doi.org/10.1016/j.agrformet.2013.07.016>
- Fu, Y., Campioli, M., Vitasse, Y., De Boeck, H. J., Van den Berge, J., AbdElgawad, H., Asard, H., Piao, S., Deckmyn, G., & Janssens, I. A. (2014). Variation in leaf flushing date influences autumnal senescence and next year's flushing date in two temperate tree species. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 7355–7360. <https://doi.org/10.1073/pnas.1321727111>
- Fuccillo, K. K., Crimmins, T. M., de Rivera, C. E., & Elder, T. S. (2015). Assessing accuracy in citizen science-based plant phenology monitoring. *International Journal of Biometeorology*, 59, 917–926. <https://doi.org/10.1007/s00484-014-0892-7>
- Graham, E. A., Riordan, E. C., Yuen, E. M., Estrin, D., & Rundel, P. W. (2010). Public Internet-connected cameras used as a cross-continental ground-based plant phenology monitoring system. *Global Change Biology*, 16, 3014–3023. <https://doi.org/10.1111/j.1365-2486.2010.02164.x>
- Hall, D., Luquez, V., Garcia, V. M., St Onge, K. R., Jansson, S., & Ingvarsson, P. K. (2007). Adaptive population differentiation in phenology across a latitudinal gradient in European aspen (*Populus tremula*, L.): A comparison of neutral markers, candidate genes and phenotypic traits. *Evolution*, 61, 2849–2860. <https://doi.org/10.1111/j.1558-5646.2007.00230.x>
- Hinks, A. E., Cole, E. F., Daniels, K. J., Wilkin, T. A., Nakagawa, S., & Sheldon, B. C. (2015). Scale-dependent phenological synchrony between songbirds and their caterpillar food source. *The American Naturalist*, 186, 84–97. <https://doi.org/10.1086/681572>
- Hmimina, G., Dufrêne, E., Pontailleur, J.-Y., Delpierre, N., Aubinet, M., Caquet, B., de Grandcourt, A., Burban, B., Flechard, C., Granier, A., Gross, P., Heinesch, B., Longdoz, B., Moureaux, C., Ourcival, J.-M., Rambal, S., Saint André, L., & Soudani, K. (2013). Evaluation of the potential of MODIS satellite data to predict vegetation phenology in different biomes: An investigation using ground-based NDVI measurements. *Remote Sensing of Environment*, 132, 145–158. <https://doi.org/10.1016/j.rse.2013.01.010>
- Jochner, S., Alves-Eigenheer, M., Menzel, A., & Morellato, L. P. C. (2013). Using phenology to assess urban heat islands in tropical and temperate regions. *International Journal of Climatology*, 33, 3141–3151. <https://doi.org/10.1002/joc.3651>
- Keenan, T. F., Darby, B., Felts, E., Sonnentag, O., Friedl, M. A., Hufkens, K., O'Keefe, J., Klosterman, S., Munger, J. W., Toomey, M., & Richardson, A. D. (2014). Tracking forest phenology and seasonal physiology using digital repeat photography: A critical assessment. *Ecological Applications*, 24, 1478–1489. <https://doi.org/10.1890/13-0652.1>
- Keenan, T. F., & Richardson, A. D. (2015). The timing of autumn senescence is affected by the timing of spring phenology: Implications for predictive models. *Global Change Biology*, 21, 2634–2641. <https://doi.org/10.1111/gcb.12890>
- Kharin, N. G. (1976). Mathematical models in phenology. *Journal of Biogeography*, 3, 357–364. <https://doi.org/10.2307/3037977>
- Klosterman, S., Hufkens, K., Gray, J. M., Melaas, E., Sonnentag, O., Lavine, I., Mitchell, L., Norman, R., Friedl, M. A., & Richardson, A. (2014). Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery. *Biogeosciences*, 11, 4305–4320. <https://doi.org/10.5194/bg-11-4305-2014>
- Klosterman, S., Melaas, E., Wang, J. A., Martinez, A., Frederick, S., O'Keefe, J., Orwig, D. A., Wang, Z., Sun, Q., Schaaf, C., Friedl, M., & Richardson, A. D. (2018). Fine-scale perspectives on landscape phenology from unmanned aerial vehicle (UAV) photography. *Agricultural and Forest Meteorology*, 248, 397–407.
- Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14, 551–560. <https://doi.org/10.1002/fee.1436>
- Kramer, K., Buiteveld, J., Forstreuter, M., Geburek, T., Leonardi, S., Menozzi, P., Povillon, F., Schelhaas, M. J., Teissier du Cros, E., Vendramin, G. G., & Van der Werfa, D. C. (2008). Bridging the gap between ecophysiological and genetic knowledge to assess the adaptive potential of European beech. *Ecological Modelling*, 216, 333–353. <https://doi.org/10.1016/j.ecolmodel.2008.05.004>



- Laube, J., Sparks, T. H., Estrella, N., Höfler, J., Ankerst, D. P., & Menzel, A. (2013). Chilling outweighs photoperiod in preventing precocious spring development. *Global Change Biology*, 20, 170–182. <https://doi.org/10.1111/gcb.12360>
- MacKenzie, C. M., Murray, G., Primack, R., & Weihrach, D. (2017). Lessons from citizen science: Assessing volunteer-collected plant phenology data with Mountain Watch. *Biological Conservation*, 208, 121–126. <https://doi.org/10.1016/j.biocon.2016.07.027>
- Mariën, B., Balzarolo, M., Dox, I., Leys, S., Lorène, M. J., Geron, C., Portillo-Estrada, M., AbdElgawad, H., Asard, H., & Campioli, M. (2019). Detecting the onset of autumn leaf senescence in deciduous forest trees of the temperate zone. *New Phytologist*, 224(1), 166–176. <https://doi.org/10.1111/nph.15991>
- Mayer, A. (2010). Phenology and citizen science: Volunteers have documented seasonal events for more than a century, and scientific studies are benefiting from the data. *BioScience*, 60, 172–175. <https://doi.org/10.1525/bio.2010.60.3.3>
- Meier, U. (1997). *Growth stages of mono- and dicotyledonous plants*. Blackwell Wissenschafts-Verlag.
- Menzel, A. (2002). Phenology: Its importance to the global change community. *Climatic Change*, 54, 379–385. <https://doi.org/10.1023/A:101612521549>
- Menzel, A. (2013a). Europe. In M. D. Schwartz (Ed.), *Phenology: An integrative environmental science* (pp. 53–65). Springer.
- Menzel, A. (2013b). Plant phenological 'Fingerprints'. In M. D. Schwartz (Ed.), *Phenology: An integrative environmental science* (pp. 335–350). Springer.
- Miller-Rushing, A. J., Inouye, D. W., & Primack, R. B. (2008). How well do first flowering dates measure plant responses to climate change? The effects of population size and sampling frequency. *Journal of Ecology*, 96, 1289–1296. <https://doi.org/10.1111/j.1365-2745.2008.01436.x>
- Naschitz, S., Naor, A., Wolf, S., & Goldschmidt, E. E. (2014). The effects of temperature and drought on autumnal senescence and leaf shed in apple under warm, east mediterranean climate. *Trees*, 28, 879–890. <https://doi.org/10.1007/s00468-014-1001-6>
- Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37–42. <https://doi.org/10.1038/nature01286>
- Pérez-de-Lis, G., Rossi, S., Vázquez-Ruiz, R. A., Rozas, V., & García-González, I. (2016). Do changes in spring phenology affect early-wood vessels? Perspective from the xylogenesis monitoring of two sympatric ring-porous oaks. *New Phytologist*, 209, 521–530. <https://doi.org/10.1111/nph.13610>
- Richardson, A. D., Bailey, A. S., Denny, E. G., Martin, W. C., & O'Keefe, J. (2006). Phenology of a northern hardwood forest canopy. *Global Change Biology*, 12, 1174–1188. <https://doi.org/10.1111/j.1365-2486.2006.01164.x>
- Richardson, A. D., Hufkens, K., Milliman, T., Aubrecht, D. M., Chen, M., Gray, J. M., Johnston, M. R., Keenan, T. F., Klosterman, S. T., Kosmala, M., Melaas, E. K., Friedl, M. A., & Frolking, S. (2018). Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. *Scientific Data*, 5, 180028. <https://doi.org/10.1038/sdata.2018.28>
- Richardson, A. D., Hufkens, K., Milliman, T., & Frolking, S. (2018). Intercomparison of phenological transition dates derived from the PhenoCam Dataset V1.0 and MODIS satellite remote sensing. *Scientific Reports*, 8, 5679. <https://doi.org/10.1038/s41598-018-23804-6>
- Richardson, A. D., Jenkins, J. P., Braswell, B. H., Hollinger, D. Y., Ollinger, S. V., & Smith, M.-L. (2007). Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Oecologia*, 152, 323–334. <https://doi.org/10.1007/s00442-006-0657-z>
- Root, T. L., Price, J. T., Hall, K. R., Schneider, S. H., Rosenzweig, C., & Pounds, J. A. (2003). Fingerprints of global warming on wild animals and plants. *Nature*, 421, 57–60. <https://doi.org/10.1038/nature01333>
- Rosenthal, S. I., & Camm, E. L. (1996). Effects of air temperature, photoperiod and leaf age on foliar senescence of western larch (*Larix occidentalis* Nutt.) in environmentally controlled chambers. *Plant, Cell and Environment*, 19, 1057–1065. <https://doi.org/10.1111/j.1365-3040.1996.tb00212.x>
- Rousi, M., & Heinonen, J. (2007). Temperature sum accumulation effects on within-population variation and long-term trends in date of bud burst of European white birch (*Betula pendula*). *Tree Physiology*, 27, 1019–1025. <https://doi.org/10.1093/treephys/27.7.1019>
- 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: Biogeosciences*, 112, <https://doi.org/10.1029/2006JG000382>
- Schaber, J. (2002). *Phenology in Germany in the 20th century: Methods, analyses and models* (Doctoral thesis). University of Potsdam.
- Schwartz, M. D., Hanes, J. M., & Liang, L. (2013). Comparing carbon flux and high-resolution spring phenological measurements in a northern mixed forest. *Agricultural and Forest Meteorology*, 169, 136–147. <https://doi.org/10.1016/j.agrformet.2012.10.014>
- Scotti-Saintagne, C., Bodénès, C., Barreneche, T., Bertocchi, E., Plomion, C., & Kremer, A. (2004). Detection of quantitative trait loci controlling bud burst and height growth in *Quercus robur* L. *Theoretical and Applied Genetics*, 109, 1648–1659. <https://doi.org/10.1007/s00122-004-1789-3>
- Soudani, K., Le Maire, G., Dufrêne, E., François, C., Delpierre, N., Ulrich, E., & Cecchini, S. (2008). Evaluation of the onset of green-up in temperate deciduous broadleaf forests derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data. *Remote Sensing of Environment*, 112, 2643–2655. <https://doi.org/10.1016/j.rse.2007.12.004>
- Testa, S., Soudani, K., Boschetti, L., & Mondino, E. B. (2018). MODIS-derived EVI, NDVI and WDRVI time series to estimate phenological metrics in French deciduous forests. *International Journal of Applied Earth Observation and Geoinformation*, 64, 132–144. <https://doi.org/10.1016/j.jag.2017.08.006>
- Vitasse, Y., Delzon, S., Bresson, C. C., Michalet, R., & Kremer, A. (2009). Altitudinal differentiation in growth and phenology among populations of temperate-zone tree species growing in a common garden. *Canadian Journal of Forest Research*, 39, 1259–1269. <https://doi.org/10.1139/X09-054>
- Vitasse, Y., Delzon, S., Dufrêne, E., Pontailleur, J.-Y., Louvet, J.-M., Kremer, A., & Michalet, R. (2009). Leaf phenology sensitivity to temperature in European trees: Do within-species populations exhibit similar responses? *Agricultural and Forest Meteorology*, 149, 735–744. <https://doi.org/10.1016/j.agrformet.2008.10.019>
- Vitasse, Y., François, C., Delpierre, N., Dufrêne, E., Kremer, A., Chuine, I., & Delzon, S. (2011). Assessing the effects of climate change on the phenology of European temperate trees. *Agricultural and Forest Meteorology*, 151, 969–980. <https://doi.org/10.1016/j.agrformet.2011.03.003>
- Wesołowski, T., & Rowiński, P. (2006). Timing of bud burst and tree-leaf development in a multispecies temperate forest. *Forest Ecology and Management*, 237, 387–393. <https://doi.org/10.1016/j.foreco.2006.09.061>
- Xie, Y., Wang, X., Wilson, A. M., & Silander Jr., J. A. (2018). Predicting autumn phenology: How deciduous tree species respond to weather stressors. *Agricultural and Forest Meteorology*, 250, 127–137. <https://doi.org/10.1016/j.agrformet.2017.12.259>
- Yang, X., Mustard, J. F., Tang, J., & Xu, H. (2012). Regional-scale phenology modeling based on meteorological records and remote sensing observations. *Journal of Geophysical Research: Biogeosciences*, 117, 1–18. <https://doi.org/10.1029/2012jg001977>



- Yu, R., Schwartz, M. D., Donnelly, A., & Liang, L. (2016). An observation-based progression modeling approach to spring and autumn deciduous tree phenology. *International Journal of Biometeorology*, 60, 335–349. <https://doi.org/10.1007/s00484-015-1031-9>
- Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84, 471–475. [https://doi.org/10.1016/S0034-4257\(02\)00135-9](https://doi.org/10.1016/S0034-4257(02)00135-9)
- Zhang, X., Jayavelu, S., Liu, L., Friedl, M. A., Henebry, G. M., Liu, Y., Schaaf, C. B., Richardson, A. D., & Gray, J. (2018). Evaluation of land surface phenology from VIIRS data using time series of PhenoCam imagery. *Agricultural and Forest Meteorology*, 256, 137–149. <https://doi.org/10.1016/j.agrformet.2018.03.003>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Liu G, Chuine I, Denéchére R, et al.

Higher sample sizes and observer inter-calibration are needed for reliable scoring of leaf phenology in trees. *J Ecol*.

2021;109:2461–2474. <https://doi.org/10.1111/1365-2745.13656>