The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2015, **115**, 543–554. With 5 figures.

Unbiased inference of plant flowering phenology from biological recording data

DANIEL S. CHAPMAN[1]*, SANDRA BELL[2], STEPHAN HELFER[3] and DAVID B. ROY[4]

[1]*Centre for Ecology & Hydrology, Bush Estate, Edinburgh EH26 0QB, UK*
[2]*Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK*
[3]*Royal Botanic Gardens, Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, UK*
[4]*Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford OX10 8BB, UK*

Received 9 December 2014; revised 21 January 2015; accepted for publication 21 January 2015

Phenology is a key indicator and mediator of the ecological impacts of climate change. However, studies monitoring the phenology of individual species are moderate in number, taxonomically and geographically restricted, and mainly focused on spring events. As such, attention is being given to nonstandard sources of phenology data, such as the dates of species' biological records. Here, we present a conceptual framework for deriving phenological metrics from biological recording data, while accounting for seasonal variation in the level of activity by recorders. We develop a new Bayesian statistical model to infer the seasonal pattern of plant 'recordability'. The modelled dates of maximum recordability are strongly indicative of the flowering peaks of 29 insect-pollinated species monitored in two botanic gardens in Great Britain. Conversely, not accounting for the seasonality in recording activity results in biased estimates of the observed flowering peaks. However, observed first and last flowering dates were less reliably explained by the model, which probably reflects greater interspecific variation in levels of recording before and after flowering. We conclude that our method provides new potential for gaining useful insights into large-scale variation in peak phenology across a much broader range of plant species than have previously been studied. © 2015 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2015, **115**, 543–554.

ADDITIONAL KEYWORDS: Bayesian model – citizen science – climate change – discrete Fourier transform – growing degree days – phenology model – recorder effort.

INTRODUCTION

Among the most clearly documented effects of recent climate change are shifts in the phenology of species and ecosystems (Menzel *et al*., 2006; Sherry *et al*., 2007; Thackeray *et al*., 2010; Chapman, 2013; Settele *et al*., 2014). Indeed, phenology is increasingly recognized as a determinant of ecological responses to climate change. Phenological change has been implicated in causing mismatch of trophic interactions (Thackeray *et al*., 2010), explaining species' demographic and population dynamic performance (Miller-Rushing *et al*., 2010; Cleland *et al*., 2012), and in mediating range shifts driven by climate change (Chapman *et al*., 2014). The potential for phenological shifts to affect ecological interactions, processes, and functions has therefore motivated a great deal of interest in understanding patterns of phenological variation across species, space, and time (Stenseth & Mysterud, 2002; Phillimore *et al*., 2010; Thackeray *et al*., 2010; Hodgson *et al*., 2011).

An important limiting factor for the study of phenology is the availability of long-term and spatially replicated data. Studies monitoring phenology of individual species are moderate in number and taxonomically and geographically restricted (Chambers *et al*., 2013). They are also mainly focused on spring events, and there has been concern that many commonly used spring event indicators may

---

*Corresponding author. E-mail: dcha@ceh.ac.uk exhibit serious biases (Van Strien *et al*., 2008; Moussus, Julliard & Jiguet, 2010; Clark & Thompson, 2011). Finally, there are few data sets able to examine spatial as well as temporal variation in phenology (Chmielewski & Rötzer, 2001; Collinson & Sparks, 2008), despite recent studies indicating divergent trends in space and time (Phillimore *et al*., 2010) and variation in phenological trends along environmental gradients (Chapman, 2013). Consequently, sole reliance on the existing long-term monitoring data will give an incomplete picture of the phenological responses to climate change.

Because of this, some attention is being given to nonstandard sources of phenology data, including biological recording schemes. Biological records typically contain the date when a species was observed, as well as its location. As the probability that a species is recorded will vary over the season, according to changes in its abundance or apparency to recorders, biological recording dates may contain valuable information on phenology. A small number of previous studies have explored this possibility. Roy & Asher (2003) used mean sighting dates for butterflies to analyse spatial trends in phenology. Hassall *et al*. (2007) demonstrated an advance in the first quartile of Odonata recording dates over recent decades. By contrast, fungal-distribution records indicated a later end of fruiting (Kauserud *et al*., 2012). More recently, Bishop *et al*. (2013) compared butterfly-distribution records with population monitoring data to show that mean biological recording dates gave a good indication of directly observed peak flight dates.

However, a common feature of such studies is that raw biological recording dates have been used to estimate phenology metrics. A problem with this is that the temporal distribution of recording dates is very likely to be affected by a confounding factor not accounted for in simple analyses. Biological recording is often, although not always, a form of opportunistic citizen science in which observations are collected without standardized sampling protocols. As a result, there is typically seasonal variation in recorder activity or effort. The potential effect of this is illustrated in Figure 1, where the observed distribution of record dates arises from the product of the seasonal distributions of recording effort and the 'recordability' of the species. As shown in Figure 1, the result is that the observed distribution of recording dates lies between the peak (mode) of the unobserved distributions of recording effort and species recordability. Raw biological recording data will therefore give biased estimates of species phenology. We expect this bias to be strongest when the species is recordable for long periods of the year, when recording effort is very unevenly distributed, and when the peaks of recording effort and recordability are far apart (as in Fig. 1D).

Based on the above arguments, the use of biological recording data for phenological research requires two new developments, which form the aims of this study. First, we require a method for teasing apart the unobserved temporal distributions of recording effort and recordability from biological recording dates (e.g. Blockeel *et al*., 2014). Second, it is necessary to show that the estimated recordability of a species provides an unbiased indicator of phenology, such as first, peak, or last flowering days. For insect-pollinated plants, one might hypothesize that the day of peak recordability corresponds to the midpoint of the flowering period, when the presence of flowers means that the species will generally be most conspicuous and identifiable to recorders. Towards these aims, we developed a Bayesian modelling framework, based on Figure 1, to estimate recordability over time statistically, and fitted the model to biological recording data for British plants. To assess the reliability of the modelling, we evaluated its correspondence to phenology data collected on 29 plant species in two botanic gardens. Plant nomenclature follows Stace (2010) throughout.

MATERIAL AND METHODS

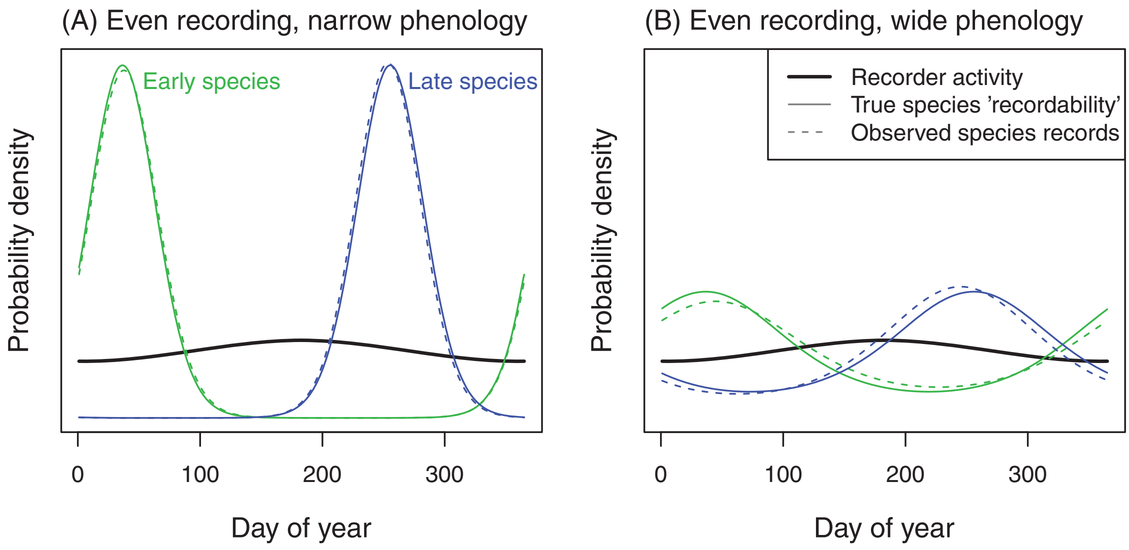
MODELS FOR THE SEASONAL DISTRIBUTION OF BIOLOGICAL RECORDS

We developed two alternative models to characterize the relationship between the day of the year and probability of recording a focal species. First, we developed a naive model that ignores recording effort and assumes that the seasonal pattern in recording is representative of the plant phenology. Based on Figure 1, we expected this model to give a biased representation of phenology. Second, we developed a recording-corrected model, explicitly accounting for the effect of recording in generating the observed data. In this second model, the seasonal pattern of recordability is assumed to represent aspects of the underlying phenology of the species.

Both models are based on the discrete Fourier transform (DFT), a highly flexible function for variation over a wrapped annual (365-day) cycle (Briggs & Henson, 1995; Moody & Johnson, 2001). We used the first and second DFT harmonics for day d, as follows:

$$h(d,\boldsymbol{\theta}) = \boldsymbol{\theta}_1 \sin\left(\frac{2\pi d}{365}\right) + \boldsymbol{\theta}_2 \cos\left(\frac{2\pi d}{365}\right) + \boldsymbol{\theta}_3 \sin\left(2\frac{2\pi d}{365}\right) + \boldsymbol{\theta}_4 \cos\left(2\frac{2\pi d}{365}\right).$$

The parameter vector $\boldsymbol{\theta}$ controls the amplitude of each phase and the resulting form of the cycle h. Additional harmonic phases increase the flexibility of h, but we wished to use the simplest possible model.



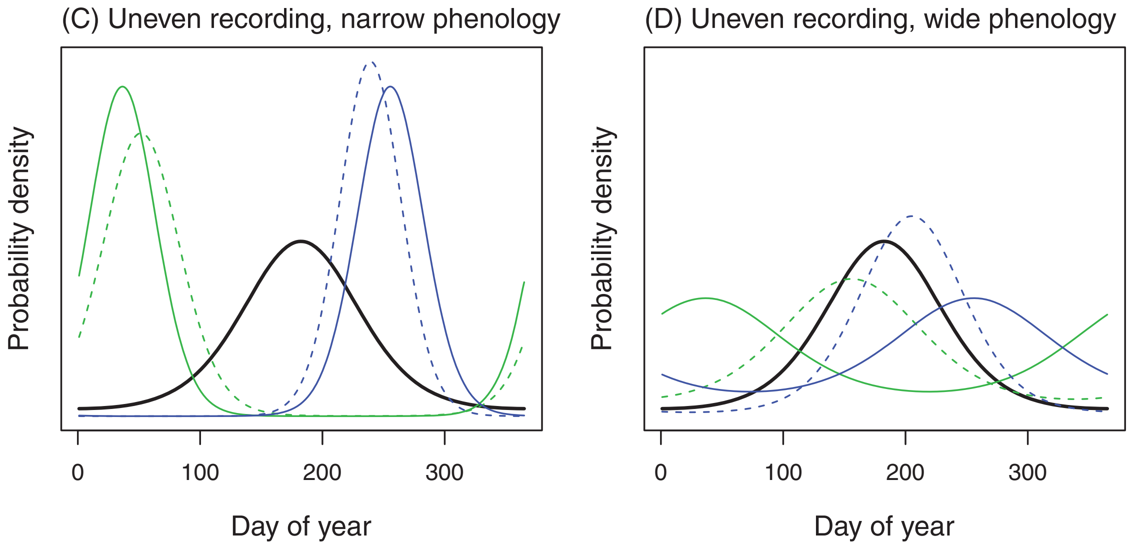


Figure 1. Seasonal variation in recorder activity is expected to bias the temporal distribution of biological records. Thick solid lines show the pattern of recording; this was very even in A and B and highly seasonal in C and D. Thin solid lines show the 'recordability' of early and late species, which may relate to their phenology. Dashed lines are the expected distribution of biological record dates, accounting for the recording pattern. It can be seen that seasonality in recording biases the distribution of recording dates towards the peak in recording activity. This is most pronounced when recording is highly seasonal and the species is recordable over a long period (D).

In the naive model, where there is no accounting for seasonal recording bias, the probability of obtaining a record on day d rather than any other day of the year (P_1) is written as:

$$P_1(d, \boldsymbol{\theta}) = \frac{e^{h(d,\boldsymbol{\theta})}}{\sum_{j=1}^{365} e^{h(j,\boldsymbol{\theta})}}.$$

Taking exponents of h ensures that $P_1 > 0$ for every day, while the denominator summation ensures that P_1 sums to 1 across the year.

The recording-corrected model for daily recording probabilities, P_2, is based on the product of DFTs representing the temporal variation in recording effort E and the recordability of the species R:

$$\begin{aligned} E(d, \boldsymbol{\theta_E}) &= e^{h(d,\boldsymbol{\theta_E})} \\ R(d, \boldsymbol{\theta_R}) &= e^{h(d,\boldsymbol{\theta_R})} \\ P_2(d, \boldsymbol{\theta_E}, \boldsymbol{\theta_R}) &= \frac{E(d, \boldsymbol{\theta_E})\, R(d, \boldsymbol{\theta_R})}{\sum_{j=1}^{365} E(d, \boldsymbol{\theta_E})\, R(d, \boldsymbol{\theta_R})} \end{aligned}.$$

In this model, the main interest is in R and its relationship to the species' phenology.

Biological recording data used in the model

The models were fitted to recording dates of selected plant species from the databases of the Botanical Society of Britain and Ireland (BSBI) accessed through the National Biodiversity Network (NBN) Gateway (https://data.nbn.org.uk/). The BSBI databases used were the Vascular Plants Database, Vascular Plants Database Additions Since 2000, Vascular Plant Records for Scottish Vice-counties, Vascular Plant Records of Perthshire and Angus (Preston, Pearman & Dines, 2002), Changing Flora of Glasgow 1982–2000 (Dickson *et al.*, 2000), Plant Life of Edinburgh and the Lothians – Vascular Plant Data (Smith, Dixon & Cochrane, 2002), The Flora of Rum: Vascular plant data (Pearman *et al.*, 2008), RISC Botanical Non-Native Species Records, and SNH Site Condition Monitoring – Vascular plants (2000–2006).

From these databases we extracted records from Great Britain (England, Wales, and Scotland), georeferenced to the 10 km × 10 km Ordnance Survey British National Grid, and with a single unique recording date. Dates were converted to the number of days since 1 January of that year, excluding 31 December during leap years. The most recent year with data was 2009. Records before 1970 were sparse and so were excluded.

Bayesian approach to fitting the models

To fit the models, we calculated the likelihood that the observed daily distributions are multinomially distributed with probabilities $P_1(d,\boldsymbol{\theta})$ (naive model) or $P_2(d,\boldsymbol{\theta_E},\boldsymbol{\theta_R})$ (recorder model). The likelihood function for the naive model is:

$$L(\boldsymbol{x_s}|\boldsymbol{\theta}) = M(\boldsymbol{x_s}, n = \Sigma\boldsymbol{x_s}, \boldsymbol{P} = P_1(d, \boldsymbol{\theta})),$$

where M is the multinomial probability density for the observed number of records per day ($\boldsymbol{x_s}$), accounting for the total number of records (n) and the daily recording probabilities ($\boldsymbol{P}$). The vector $\boldsymbol{d}$ contains all days of the year.

Fitting the recording-corrected model is more complex because recording effort E and species recordability R are confounded, causing an identifiability problem (i.e. some alternative parameterizations are exactly equivalent in likelihood). Our solution is to obtain an independent estimate of E from a group of 'benchmark species' representing native, naturalized, or widely planted evergreen trees and shrubs. We assumed that these should have approximately equal recordability through the year and so their recording dates should provide a fair representation of the seasonal pattern of recording effort.

The benchmark taxa we selected were *Abies*, *Arbutus*, *Buxus*, *Chamaecyparis*, *Cupressus*, *Euonymus japonicus*, *Hedera*, *Ilex*, *Juniperus*, *Picea*, *Pinus*, *Pseudotsuga*, *Quercus ilex*, *Rhododendron ponticum*, *Ruscus*, *Taxus*, *Thuja*, *Tsuga*, and *Ulex*. These exhibited some variation in their overall seasonal recording patterns, which may be caused by variations in sample size or spatial or temporal distribution trends. However, our assumption was that by averaging across these taxa we would obtain the most representative picture of plant-recording activity. This assumption was tested in the model validation.

Subscripting the combined records of the benchmark species with the letter b, the likelihood function for the recording-corrected model is:

$$\begin{aligned} L(\boldsymbol{x_s}, \boldsymbol{x_b}|\boldsymbol{\theta_R}, \boldsymbol{\theta_A}) = {} & M(\boldsymbol{x_s}, n = \Sigma\boldsymbol{x_s}, \boldsymbol{P} = P_2(\boldsymbol{d}, \boldsymbol{\theta_E}, \boldsymbol{\theta_R})) \\ & M(\boldsymbol{x_b}, n = \Sigma\boldsymbol{x_b}, \boldsymbol{P} = P_2(\boldsymbol{d}, \boldsymbol{\theta_E}, \boldsymbol{\theta}_0)). \end{aligned}$$

For the benchmark species, $\boldsymbol{\theta_0} = [0,0,0,0]$, yielding equal recordability on every day and ensuring that they are only used for estimation of recording effort.

Model fitting was implemented through Metropolis sampling of the Markov Chain Monte Carlo (MCMC) algorithm on the model log-likelihoods, as implemented by the MCMCpack R library (Martin, Quinn & Park, 2011). Flat, improper priors were used for all parameters. MCMC chains included a burn-in of 10^4 iterations, after which posterior distributions were sampled through 2×10^5 MCMC iterations that were thinned every 20 iterations. Metropolis sampling tuning parameters were set to ensure an MCMC acceptance rate of ~25% (values of 1.4 for the naive model and 0.9 for the recording-corrected model). Raftery and Lewis's diagnostics (Raftery & Lewis, 1992) were used to ensure that the chain lengths and thinning intervals were adequate. A diagnostic for stationarity and convergence of the chain was also applied (Heidelberger & Welch, 1983).

In practice, not all the data were used in model fitting. First, we observed that the recording data always contained an excess of records for 1 January, probably reflecting inaccuracies in the database. Therefore, the models were fitted to data from day 2 to day 365. Second, for the recorder model we subsampled the benchmark records to ensure approximately equivalent spatial and temporal coverage to the focal species and to account for potential spatiotemporal variation in recording effort. The subsampling algorithm randomly sampled benchmark records from each year in proportion to their frequency in the focal species' records. Following this, the benchmark records were sampled in proportion to the spatial distribution of the focal species' records, resolved at 100 km × 100 km grid level. A constraint was placed on the sampling such that if the benchmark data contained fewer records than the focal species for any sampling unit, then all benchmark records for that sampling unit were retained. The alternative would be to sub-sample the records of the focal species, but we did not want to reduce the volume of data on the species of interest.

Validation of the fitted models

To test whether recordability estimates made from the biological recording data correspond to species' phenology, we compared the modelled recording and recordability patterns with actual data on flowering times. These data come from phenological monitoring at the Royal Botanic Gardens (RBG) at Edinburgh and at Kew (London), Great Britain. At both gardens, the start and end dates of flowering for individual plants or small self-sustaining populations are recorded, from which we calculated the median flowering period midpoint for each species.

RBG Edinburgh data spans the period 2006–2012, and RBG Kew data spans the period 2001–2013. For comparison, the naive and recording-corrected models were fitted to recording data of the same species over the period 2000–2009. Although not perfectly overlapping with the monitoring data, we assumed that most species would have similar phenology over the different periods because there was little trend in annual temperature over these periods.

Robust Major Axis (MA) regression was used to test relationships between the observed phenological data and estimates made from the model because of sampling error in both the dependent and independent variables (Taskinen & Warton, 2011). Observed flowering period midpoints were regressed against the estimated peak days for recording (naive model) and recordability (recording-corrected model). Observed first and last flowering dates were regressed against various percentiles of the modelled recording and recordability distributions. There is no *a priori* reason to select a particular percentile reflecting the start and end of flowering, as species may be recorded outside the flowering period. Therefore, we found percentiles in which the models most closely matched the observed dates, using paired *t*-tests to quantify the deviation.

The analysis excluded benchmark species and was also restricted to insect-pollinated species. We expected conspicuous flowers to be generally more important to recordability than would be the case for wind-pollinated flowers (although many wind-pollinated species are also more easily identified when in flower). Species with a recorded flowering midpoint outside the general flowering period of the species, as described in a standard floral guide (Rose *et al*., 2006), were also excluded as they may behave differently in the garden than in the wild. We also excluded species with fewer than 100 biological records. Applying these criteria, 20 species from RBG Edinburgh and nine from RBG Kew were included in the analysis. The full list was *Aesculus hippocastanum*, *Anemone nemorosa*, *Anthriscus sylvestris*, *Armeria maritima*, *Calluna vulgaris*, *Digitalis purpurea*, *Dryas octopetala*, *Erica cinerea*, *Erica tetralix*, *Fuchsia magellanica*, *Galanthus nivalis*, *Hyacinthoides non-scripta*, *Iris pseudacorus*, *Lamium album*, *Leucanthemum vulgare*, *Lonicera periclymenum*, *Mahonia aquifolium*, *Malva moschata*, *Muscari armeniacum*, *Narcissus pseudonarcissus*, *Oxalis acetosella*, *Primula vulgaris*, *Prunus laurocerasus*, *Ribes sanguineum*, *Rosa canina* agg., *Sambucus nigra*, *Sorbus aucuparia*, and *Stachys officinalis*.

RESULTS

Illustration of the model

Recording of the benchmark taxa exhibited very strong seasonality, with more observations during summer than winter (Fig. 2). As we expect these species to exhibit little seasonal variation in

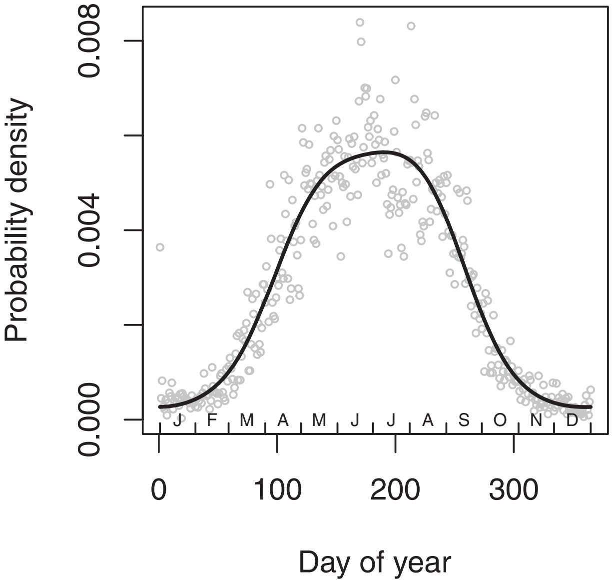


Figure 2. The seasonal pattern in record dates for the benchmark species (evergreen trees and shrubs), used to estimate the seasonality in recording activity. Points show the density of records on each day, nationally from 1970–2009. The line is the posterior mean model fitted to the data (see the equation for P_1), excluding records for the first day of the year, which can be seen to be unreliable. The posterior mean parameters are $\boldsymbol{\theta} = [0.074, -1.436, 0.063, -0.285]$. Internal ticks divide months.

recordability, this suggests that recording effort was highly seasonal and would introduce significant bias into naive estimates of phenology made from recording data.

Figure 3 illustrates the fitted naive and recording-corrected models for six common species. All had highly seasonal distributions of recording dates, so that the naive model indicated peaks in recording during summer. However, when the corrected model was implemented, the estimated seasonality in recordability deviated strongly from the raw distribution of recording dates, consistent with Figure 1. As expected, the estimated recordability of two evergreen species did not vary strongly across the year (Fig. 3A, B). Conversely, the estimated recordability for species characteristic of winter, spring, summer, and autumn flowering showed strong seasonality that was broadly consistent with their flowering periods (Fig. 3C–F). The most striking example was for the winter heliotrope *Petasites fragrans*, for which the naive model estimated a recordability peak in summer, whereas the corrected model suggested a peak in winter.

MODEL VALIDATION

Both models were fitted to the 29 species monitored in the two botanic gardens. As expected, the naive model produced later recordability peaks for early species than did the corrected model, and earlier estimates for late species (Fig. 4A). After exclusion of four species that were clear outliers from the general relationship (*F. magellanica*, *M. aquifolium*, *P. laurocerasus*, and *R. sanguineum*), there was a strong association between peak recording days estimated from both models, (Spearman's $\rho = 0.948$, $P < 0.001$; Fig. 4A). The outliers tended to have smaller numbers of records than the non-outliers (Wilcoxon signed-rank test, $P = 0.007$; Fig. 4B), which probably compromised the model fitting.

Model results from the 25 non-outlying species were compared with phenology metrics observed in the botanic gardens (Fig. 5). Estimated first, peak, and last flowering dates were strongly correlated to the observations (Spearman's $\rho > 0.59$ and $P < 0.001$ in all cases). However, the recording-corrected model always resulted in closer and less-biased relationships between observed and model-inferred phenology, as shown by MA regression fits being closer to a one-to-one relationship (Fig. 5). The timing of flowering peaks were estimated most precisely and with the least bias, whereas flowering end dates were least well explained (Fig. 5). However, it can be seen from Figure 5E that the latter result was mainly caused by poor performance for species extending their flowering period late into the year.

DISCUSSION

We hypothesized that dates from biological recording data, which are not collected for monitoring phenology but rather for the primary purpose of documenting species' distributions, should nevertheless contain useful information on phenology. However, as outlined in Figure 1, we expected that the observed recording dates would result from seasonal variations in the recordability of a species and in recording effort. As a result, previous studies that used raw biological recording dates in assessments of phenology (Roy & Asher, 2003; Hassall *et al*., 2007; Kauserud *et al*., 2012; Bishop *et al*., 2013) may be subject to substantial bias. Furthermore, we reasoned that the magnitude and nature of this bias would be affected by the seasonality and congruency of recording effort and recordability. Specifically, the raw recording dates should be most misleading about the true phenology if the species is recordable for long periods of the year, when recording effort is very unevenly distributed and when the periods of high recording effort and recordability are strongly mismatched in time.

As this recording bias will vary across species, time, and space, we strongly caution against using raw biological recording dates for phenological studies, even where these biases are expected to be minimal. For example, Robbirt *et al*. (2011) showed that herbarium specimen collection dates of the orchid *Ophrys sphegodes* responded to temperature very similarly to directly monitored flowering. The collected samples can be confirmed as flowering, and *O. sphegodes* has a very short flowering period, so we would expect minimal bias. Likewise, Kauserud *et al*. (2012) restricted their analyses to distribution records from fungi with short fruiting periods. However, even in these cases, changes in recorder behaviour over time could bias estimates of long-term phenological change. Indeed, our analysis of the benchmark taxa indicates that plant recording has started earlier in recent decades than in the past (analysis not shown) which may be caused by warmer conditions as well as availability of better resources for vegetative identification. Failing to account for earlier recording would indicate a spurious phenological advance in the records.

Given the motivation for obtaining new sources of information for phenological research, we developed a statistical model to estimate the seasonality in recording effort and species recordability from biological recording dates. Furthermore, we showed that the estimated day of peak recordability is a useful phenology metric, exhibiting a strong and unbiased correlation to peak flowering times monitored in two botanic gardens. However, the model produced less

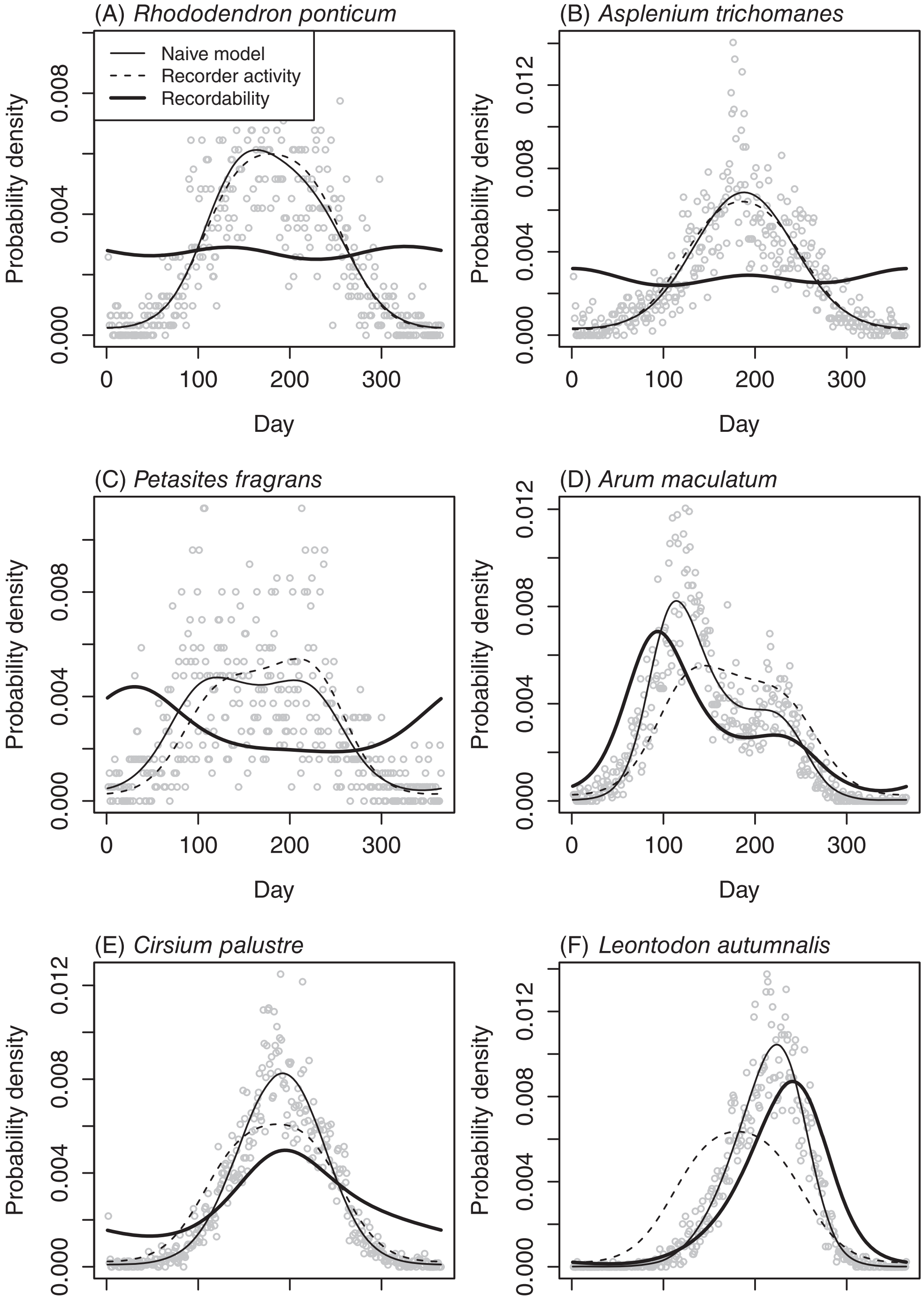


Figure 3. Examples of fitting both models to distribution records from 1970–2009. Points show the proportions of records from each day. Lines show posterior phenology curves estimated from the naive model (no accounting for temporal variation in recording) and the corrected model (temporal variation in recording activity and species recordability are modelled). Variations in the spatial and annual distributions of records among species cause differences in the estimates of the recording activity pattern. The hypothesis of this study is that temporal variation in recordability corresponds to phenology, so that recordability can be viewed as 'corrected phenology'. A, B, two evergreen species are shown, one from the benchmark group of species (A) and one from outside that group (B). C–F, nonbenchmark species with strong phenological signals for winter (C), spring (D), summer (E), and autumn (F) flowering plants.

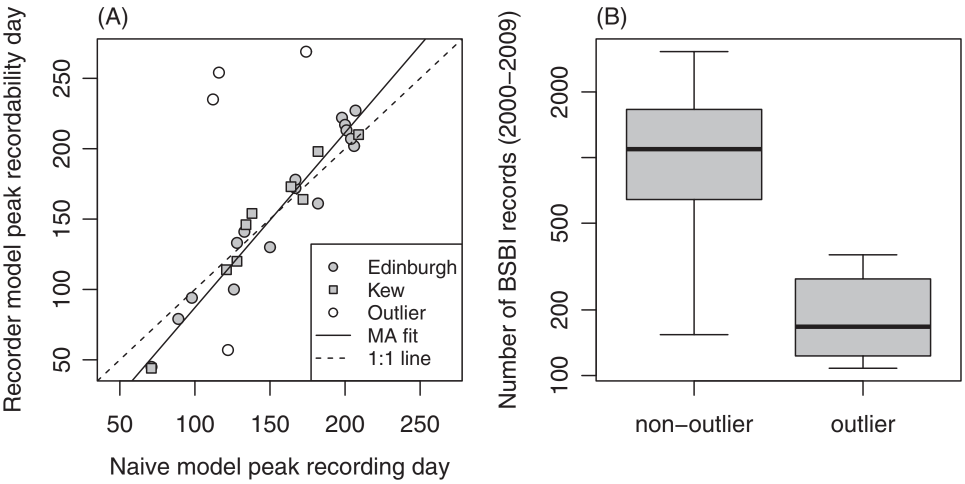


Figure 4. A, relationship between peak recording estimated from the naive model and peak recordability from the recording-corrected model. Excluding four clear outliers, species fall closely on a line just steeper than 1:1 [major axis (MA) regression fit, $y = -37.2 + 1.24x$, $R^2 = 0.946$, $P < 0.001$). B, as shown in the boxplots, the outliers had relatively few records.

accurate and more biased estimates of first and last flowering dates, suggesting that biological recording data are more suited to estimating peak phenology rather than the extremes of flowering periods, across many taxa. The likely explanation is that plant species will differ strongly in their propensity to be recorded before and after their flowering period because of variation in their ease of identification from vegetative characteristics. This will cause interspecific variation in the recordability percentile corresponding to their first and last flowering (as in Fig. 5) and means that a single percentile cannot be used across many species. By contrast, the majority of species should be most conspicuous to recorders at the peak of their flowering, so the recordability peak should apply similarly across most species.

A key feature of our model was the use of 'benchmark' species to infer recording activity. These species were prominent evergreen trees and shrubs that should therefore be more or less equally recordable through the year. Hence, we assumed that their recording dates would be informative about recording effort. Importantly, we subsampled the benchmark records to match the spatial and annual distribution of records for each species to gain the most precise indication of recording effort for each species. Nevertheless, as shown in Figure 4, four of 29 species to which the model was applied resulted in outliers from the clear and expected relationship between the naive and recorder-informed model.

Part of the reason for this was sample size, as all the outliers had fewer than 400 records. Additionally, two of the outliers were evergreen shrubs not included in the benchmarks (*M. aquifolium* and *P. laurocerasus*) but which had weak seasonality in recordability. One of the other outliers was the red-flowering currant, *R. sanguineum*, for which both models estimated a bimodal recordability curve peaking in spring and autumn. The naive model had the spring peak as the major one, whereas the recorder-informed model had the late autumn peak as the major one, resulting in the outlier. This was presumably caused by high recordability of the species for both the flowers in early spring and conspicuous fruits in late summer. The same effect can also been seen in the bimodal pattern for *Arum maculatum* in Figure 3, a species with similarly distinctive flowers and fruits (Sowter, 1949).

This illustrates the care that must be taken when interpreting the biological significance of peak recordability. In most cases we expect that it will relate to flowering, as flowers generally make plants more conspicuous and easier to identify. However, some species are likely to deviate from this general pattern. For example, some species will be equally recordable when in flower as in leaf (e.g. wind-pollinated *Corylus avellana*) so the peak may indicate the midpoint of the period in leaf. In some species, recordability peaks may correspond to conspicuous phenophases other than flowering (e.g. many *Rumex* spp. are identified from their fruits, whereas *Crassula tillaea* develops conspicuous orange–red foliage independently of flowering). Furthermore, some species will vary in recordability according to changes in the surrounding vegetation rather than their own phenology (e.g. *Viscum album* is more noticeable when its host trees have shed their leaves for winter). Therefore, careful selection of appropriate study species for wider appli-

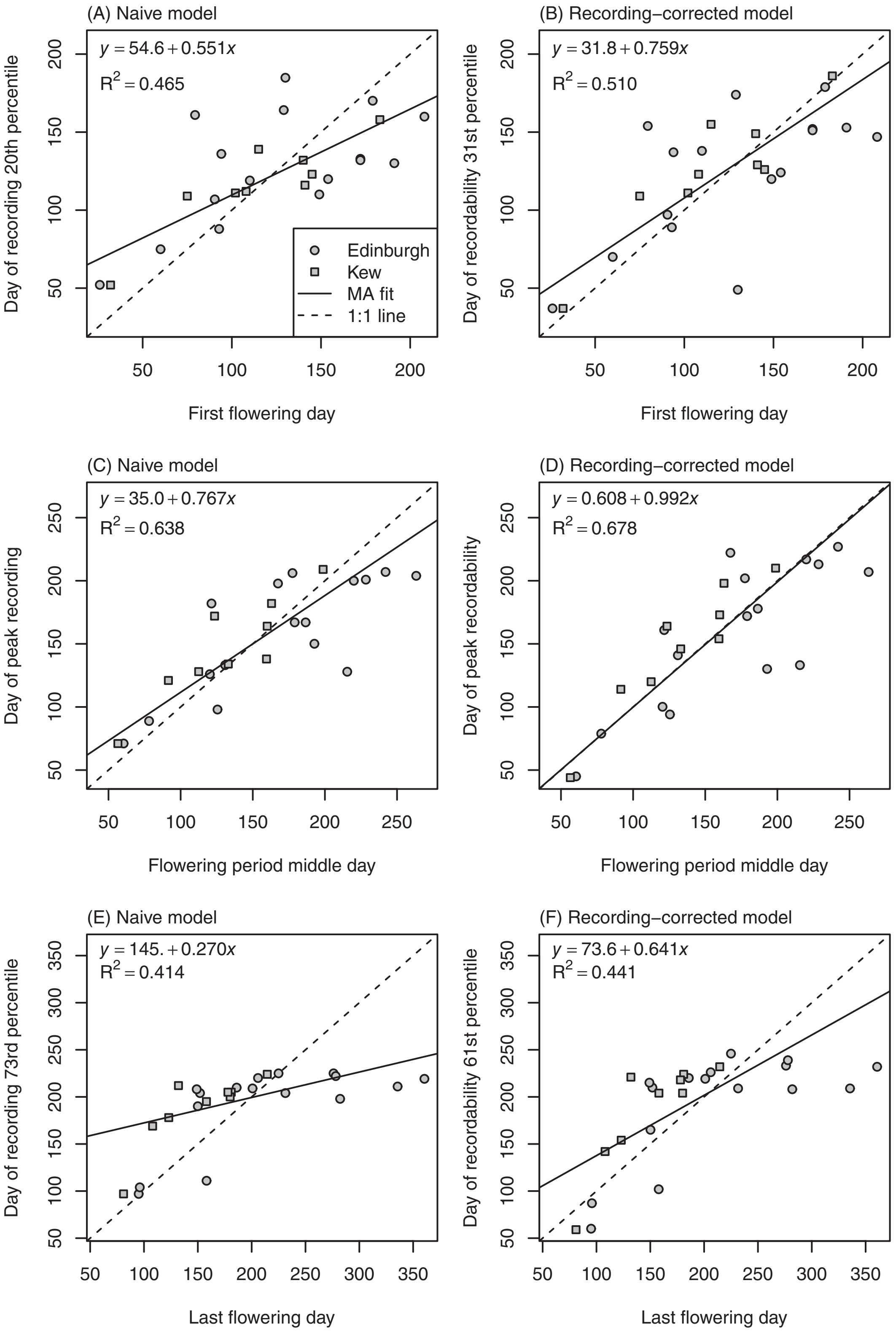


Figure 5. Relationships between flowering period start, middle, and end dates monitored in two botanic gardens and (A, C, E) recording estimated by the naive model and (B, D, F) recordability estimated by the corrected model. A, B, E, F, the percentiles used were chosen to maximize correspondence between the model and the observations. Major-axis (MA) regression lines, equations and R^2 values are given. Outliers identified in Figure 4 are omitted.

cation of the model is necessary. Insect-pollinated species with distinctive flowers and relatively inconspicuous fruits are likely to be the most suitable.

The estimated recordability peaks explained around 70% of the observed interspecific variation in observed peak flowering. The residual variation will partly have resulted from the comparison of national-level estimates from the model with single-site estimates from the botanic gardens. Furthermore, the gardens may be situated outside the geographical or altitudinal range of the species or in warm urban microclimates. Also, there was a slight mismatch between the years of the monitoring and modelling data. However, it also suggests a level of imprecision in the modelling, which means that the approach advanced here is probably more useful for broad studies across many species, rather than for detailed characterization of individual species' phenology. Biological recording data are available for many more species and locations than even in the best existing direct phenological monitoring recording schemes, such as Nature's Calendar (Collinson & Sparks, 2008; Amano *et al*., 2010) or the International Phenology Gardens (Chmielewski & Rötzer, 2001). Therefore, the main advantage of using biological recording data for phenological study may be the potential to analyse phenology trends in space, as well as over time, for many more species than is currently possible (Roy & Asher, 2003).

Similar approaches should also be useful for phenological research on other well-recorded taxonomic groups. However, a lack of equivalent benchmark taxa may prevent direct application of the model to other taxa. For example, in temperate areas virtually all terrestrial insects are inactive at some time of the year – usually winter. Therefore, it may be impossible to identify benchmark taxa equally recordable throughout the year (Bishop *et al*., 2013). However, the model could be modified for these groups in at least two ways. First, the model could make use of some other source of reliable information on recording effort, such as numbers of unique recording lists over time. A second possibility relies on some of the species being subject to standardized phenological monitoring, such as monitoring of within-year abundance. It should be possible to use this information on the true phenology to infer the pattern of recording activity responsible for generating the biological recording dates of those species, and use this to estimate the recordability of other species. This would be a particularly powerful approach as it would directly link the recordability concept advanced here with the underlying abundance of the species.

Our model sits within the broader research on accounting for recording effects in opportunistic biological recording data. Other approaches have been developed to address problems, including estimation of species richness (Hill, 2012), population trends (Hill, 2012; Isaac *et al*., 2014), and range shifts (Hassall & Thompson, 2010; Strien, Swaay & Termaat, 2013). Other authors have also considered biases from volunteer recording or citizen science schemes designed to monitor phenology directly (Sparks, Huber & Tryjanowski, 2008; Van Strien *et al*., 2008; Gonsamo & D'Odorico, 2014). However, to our knowledge, this study is the first attempt to model recording effects in the derivation of phenology metrics from biological recording data. We find that raw recording dates give biased phenology estimates, but that knowledge on recorder activity can be used to correct this bias. Specifically, flowering peaks may be well suited to estimation from biological recording data; however, the timing of first and last phenological events may be more elusive across large numbers of species. Therefore, future studies should exploit the potential of biological recording data to gain new understanding of the factors influencing peak phenological change. In particular, the wide reach of biological recording should allow consideration of a broader range of taxa and spatial scales than have thus far been permitted by direct long-term phenological monitoring.

ACKNOWLEDGEMENTS

We are indebted to the many volunteers who have contributed to the Botanical Society of Britain and Ireland recording scheme and the phenological monitoring at the Royal Botanic Garden Edinburgh. We thank Myra Miller, Marion Piper, David Rowe and Margaret Sorrell for collecting phenology data at RBG Kew. The Biological Record Centre receives support from the Joint Nature Conservation Committee and the Natural Environment Research Council (via National Capability funding to the Centre for Ecology & Hydrology, project NEC04932). We thank Chris Preston and an anonymous reviewer for helpful comments on an earlier draft.

REFERENCES

Amano T, Smithers RJ, Sparks TH, Sutherland WJ. 2010. A 250-year index of first flowering dates and its response to temperature changes. *Proceedings of the Royal Society B: Biological Sciences* **277:** 2451–2457.

Bishop TR, Botham MS, Fox R, Leather SR, Chapman DS, Oliver TH. 2013. The utility of distribution data in predicting phenology. *Methods in Ecology and Evolution* **4:** 1024–1032.

Blockeel TL, Bosanquet SDS, Hill MO, Preston CD. 2014. *Atlas of British & Irish Bryophytes*. Newbury, UK: Pisces Publications.

Briggs WL, Henson VE. 1995. *The DFT: an owners' manual for the discrete Fourier transform*. Philadelphia, PA: Siam.

Chambers LE, Altwegg R, Barbraud C, Barnard P, Beaumont LJ, Crawford RJ, Durant JM, Hughes L, Keatley MR, Low M, Morellato PC, Poloczanska ES, Ruoppolo V, Vanstreels RET, Woehler EJ, Wolfaardt AC. 2013. Phenological changes in the Southern Hemisphere. *PLoS ONE* **8:** e75514.

Chapman DS. 2013. Greater phenological sensitivity to temperature on higher Scottish mountains: new insights from remote sensing. *Global Change Biology* **19:** 3463–3471.

Chapman DS, Haynes T, Beal S, Essl F, Bullock JM. 2014. Phenology predicts the native and invasive range limits of common ragweed. *Global Change Biology* **20:** 192–202.

Chmielewski F-M, Rötzer T. 2001. Response of tree phenology to climate change across Europe. *Agricultural and Forest Meteorology* **108:** 101–112.

Clark RM, Thompson R. 2011. Sampling bias in the determination of first and last occurrences. *Plant Ecology & Diversity* **4:** 201–211.

Cleland EE, Allen JM, Crimmins TM, Dunne JA, Pau S, Travers SE, Zavaleta ES, Wolkovich EM. 2012. Phenological tracking enables positive species responses to climate change. *Ecology* **93:** 1765–1771.

Collinson N, Sparks T. 2008. Phenology – nature's calendar: an overview of results from the UK phenology network. *Arboricultural Journal* **30:** 271–278.

Dickson J, Macpherson P, Watson K, Tait T. 2000. *The changing flora of Glasgow: urban and rural plants through the centuries*. Edinburgh: Edinburgh University Press.

Gonsamo A, D'Odorico P. 2014. Citizen science: best practices to remove observer bias in trend analysis. *International Journal of Biometeorology* **58:** 2159–2163.

Hassall C, Thompson DJ. 2010. Accounting for recorder effort in the detection of range shifts from historical data. *Methods in Ecology and Evolution* **1:** 343–350.

Hassall C, Thompson DJ, French GC, Harvey IF. 2007. Historical changes in the phenology of British Odonata are related to climate. *Global Change Biology* **13:** 933–941.

Heidelberger P, Welch PD. 1983. Simulation run length control in the presence of an initial transient. *Operations Research* **31:** 1109–1144.

Hill MO. 2012. Local frequency as a key to interpreting species occurrence data when recording effort is not known. *Methods in Ecology and Evolution* **3:** 195–205.

Hodgson JA, Thomas CD, Oliver TH, Anderson BJ, Brereton T, Crone E. 2011. Predicting insect phenology across space and time. *Global Change Biology* **17:** 1289–1300.

Isaac NJ, Strien AJ, August TA, Zeeuw MP, Roy DB. 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution* **5:** 1052–1060.

Kauserud H, Heegaard E, Büntgen U, Halvorsen R, Egli S, Senn-Irlet B, Krisai-Greilhuber I, Dämon W, Sparks T, Nordén J, Høiland K, Kirk P, Semenov M, Boddy L, Stenseth NC. 2012. Warming-induced shift in European mushroom fruiting phenology. *Proceedings of the National Academy of Sciences of the United States of America* **109:** 14488–14493.

Martin AD, Quinn KM, Park JH. 2011. MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software* **42:** 1–21.

Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, Alm-KüBler K, Bissolli P, Braslavská O, Briede A, Chmielewski FM, Crepinsek Z, Curnel Y, Dahl Å, Defila C, Donnelly A, Filella Y, Jatczak K, Måge F, Mestre A, Nordli Ø, Peñuelas J, Pirinen P, Remišová V, Scheifinger H, Striz M, Susnik A, Van Vliet AJH, Wielgolaski F-E, Zach S, Zust ANA. 2006. European phenological response to climate change matches the warming pattern. *Global Change Biology* **12:** 1969–1976.

Miller-Rushing AJ, Høye TT, Inouye DW, Post E. 2010. The effects of phenological mismatches on demography. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365:** 3177–3186.

Moody A, Johnson DM. 2001. Land-surface phenologies from AVHRR using the discrete fourier transform. *Remote Sensing of Environment* **75:** 305–323.

Moussus JP, Julliard R, Jiguet F. 2010. Featuring 10 phenological estimators using simulated data. *Methods in Ecology and Evolution* **1:** 140–150.

Pearman DA, Preston CD, Rothero GP, Walker KJ. 2008. *The Flora of Rum: an Atlantic island reserve*. Truro, UK: Privately Published.

Phillimore AB, Hadfield JD, Jones OR, Smithers RJ. 2010. Differences in spawning date between populations of common frog reveal local adaptation. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 8292–8297.

Preston CD, Pearman DA, Dines TD. 2002. *New Atlas of the British and Irish flora: an atlas of the vascular plants of Britain, Ireland, the Isle of Man and the Channel Islands*. Oxford, UK: Oxford University Press.

Raftery AE, Lewis SM. 1992. One long run with diagnostics: implementation strategies for Markov Chain Monte Carlo. *Statistical Science* **7:** 493–497.

Robbirt KM, Davy AJ, Hutchings MJ, Roberts DL. 2011. Validation of biological collections as a source of phenological data for use in climate change studies: a case study with the orchid *Ophrys sphegodes*. *Journal of Ecology* **99:** 235–241.

Rose F, O'Reilly C, Smith DP, Collings M. 2006. *The Wild Flower Key: how to identify wild flowers, trees and shrubs in Britain and Ireland*. London: Penguin Books.

Roy D, Asher J. 2003. Spatial trends in the sighting dates of British butterflies. *International Journal of Biometeorology* **47:** 188–192.

Settele J, Scholes R, Betts R, Bunn S, Leadley P, Nepstad D, Overpeck J, Taboada M. 2014. Terrestrial and inland water systems. In: Field C, Barros V, Dokken D, Mach K, Mastrandrea M, Bilir T, Chatterjee M, Ebi K, Estrada Y, Genova R, Girma B, Kissel E, Levy A, MacCracken S, Mastrandrea P, White L, eds. *Climate* *change 2014: impacts, adaptation, and vulnerability. Part A: global and sectoral aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York: Cambridge University Press, 271–360.

Sherry RA, Zhou X, Gu S, Arnone JA, Schimel DS, Verburg PS, Wallace LL, Luo Y. 2007. Divergence of reproductive phenology under climate warming. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 198–202.

Smith PM, Dixon ROD, Cochrane MP. 2002. *Plant Life of Edinburgh and the Lothians*. Edinburgh: Edinburgh University Press.

Sowter F. 1949. *Arum maculatum* L. *The Journal of Ecology* **37:** 207–219.

Sparks T, Huber K, Tryjanowski P. 2008. Something for the weekend? Examining the bias in avian phenological recording. *International Journal of Biometeorology* **52:** 505–510.

Stace C. 2010. *New Flora of the British Isles*. Cambridge, UK: Cambridge University Press.

Stenseth NC, Mysterud A. 2002. Climate, changing phenology, and other life history traits: nonlinearity and match–mismatch to the environment. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 13379–13381.

Strien AJ, Swaay CA, Termaat T. 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology* **50:** 1450–1458.

Taskinen S, Warton DI. 2011. Robust estimation and inference for bivariate line-fitting in allometry. *Biometrical Journal* **53:** 652–672.

Thackeray SJ, Sparks TH, Frederiksen M, Burthe S, Bacon PJ, Bell JR, Botham MS, Brereton TM, Bright PW, Carvalho L, Clutton-Brock TIM, Dawson A, Edwards M, Elliott JM, Harrington R, Johns D, Jones ID, Jones JT, Leech DI, Roy DB, Scott WA, Smith M, Smithers RJ, Winfield IJ, Wanless S. 2010. Trophic level asynchrony in rates of phenological change for marine, freshwater and terrestrial environments. *Global Change Biology* **16:** 3304–3313.

Van Strien AJ, Plantenga WF, Soldaat LL, Van Swaay CA, WallisDeVries MF. 2008. Bias in phenology assessments based on first appearance data of butterflies. *Oecologia* **156:** 227–235.