

Inferring local processes from macro-scale phenological pattern: a comparison of two methods

Albert B. Phillimore^{1,2*}, Konstantinos Proios¹, Naiara O'Mahony^{1,3}, Rodolphe Bernard¹, Alexa M. Lord¹, Sian Atkinson⁴ and Richard J. Smithers⁵

¹Imperial College London, Ascot, UK; ²University of Edinburgh, Edinburgh, UK; ³University of Durham, Durham, UK;

⁴The Woodland Trust, Grantham, UK; and ⁵Ricardo-AEA, Harwell, UK

Summary

1. Understanding the processes responsible for macro-scale spatial and temporal phenological patterns is a critical step in developing predictive phenological models. While phenological responses may involve the integration of multiple environmental cues, the spring phenology of many plant and animal species appears to be especially sensitive to temperature.

2. As a result of the success of citizen science schemes in mobilizing amateur naturalists, for some parts of the world, there now exist extensive data sets of phenological timings, spanning many species, locations and years. In macroecology, two types of models – time windows and growing degree-days – are widely used to predict phenology on the basis of temperature.

3. Here, we compare the performance of the two methods in predicting spatiotemporal variation in the timing of *Quercus robur* first leafing. The methods agree on the time at which leafing becomes sensitive to temperature and provide weak support for a delay in initiation of thermal sensitivity with increasing latitude due to a day-length requirement. Both methods explain c. 50% of the variation in first dates and identify plasticity, rather than local adaptation, as the major cause of spatial covariation between temperature and phenology. For a 1°C rise in spring temperatures we predict that a plastic response of first leafing will give rise to an advance of about seven days.

4. Synthesis: Time-window and growing degree-day methods provide remarkably congruent insights into the processes underpinning geographic variation in *Quercus robur* first leafing dates. We find that a spatially invariant plastic response to temperature dominates spatiotemporal phenological variation, which means that it may be reasonable to substitute space for time to project how this species will respond to climate change. This study demonstrates the contribution that top-down macroecological approaches can make to our understanding of the processes that give rise to intraspecific phenological variation.

Key-words: day-length, growing degree-days, local adaptation, macroecology, phenology, plant–climate interactions, plasticity, time window

Introduction

Phenology is the study of the timing of life-history events that occur in a seasonal and repeated pattern (Forrest & Miller-Rushing 2010). Observations of key life-history events, collected either deliberately or incidentally, extend as far back as the 800s in Japan (Aono & Kazui 2008) and 1700s in Europe (Marshall 1789; Sparks & Carey 1995). Over the past two decades, interest in phenology has burgeoned, as the value of these data in providing evidence for the effect of changes in climate variables on species and their biotic inter-

actions has been recognized (Walther *et al.* 2002; Parmesan & Yohe 2003; Forrest & Miller-Rushing 2010).

Interannual variation in ambient temperature is a strong negative correlate of the timing of many spring events (Crick & Sparks 1999; Menzel & Fabian 1999; Fitter & Fitter 2002; Willis *et al.* 2008). While it is widely accepted that an individual's phenological response may require the integration of more than a single environmental variable (Visser *et al.* 2010), temperature is often able to explain a large proportion of spatial and temporal variation in the timing of events of fungi, plants and animals (Sparks & Carey 1995). Most studies that seek to describe temperature's effect on phenology employ a variant of one of two main types of statistical approach, time-window correlations and growing-degree day

*Correspondence author. E-mail: albert.phillimore@ed.ac.uk

(GDD) models, although alternatives to these methods have been proposed (e.g. Gienapp, Hemerik & Visser 2005, Van de Pol & Cockburn 2011).

Time-window approaches typically involve regression of phenological observations collected across time or space on the mean temperature during a specified period, often on or several months. At their simplest, these models have four parameters, the start and end date of the window, an intercept and slope. Usually, the intercept and slope are estimated using linear regression, and models for several alternative time windows are compared using R^2 or AIC. More sophisticated variations on the time-window approach include nonlinear regression (Harrington, Woiwod & Sparks 1999) and relaxation of the assumption that temperatures on all days within the window affect phenology equally (Roberts 2012). Time-window models are easily applied and interpreted, even in the absence of any mechanistic understanding of the relationship between temperature and phenology, and this is probably one explanation for their popularity particularly in vertebrate studies (e.g. Crick & Sparks 1999) and in macroecological analyses that consider multiple species (e.g. Roy & Sparks 2000). Furthermore, treating the relationship between temperature and phenology as a regression slope means it is straightforward to adopt quantitative genetic reaction norm based methods. Reaction norms describe how a single genotype's phenotype (phenology) changes with respect to a change in environment (temperature). Temperature versus phenology reaction norms have variously been used to estimate the additive genetic component of variation in the plastic response among individuals in a single population (Nussey *et al.* 2005; Husby *et al.* 2010), test whether population-level reaction norms over space and time differ (Phillimore *et al.* 2010, 2012) and test whether populations differ in their plasticity (Husby *et al.* 2010; Phillimore *et al.* 2010).

Growing degree-day (GDD) approaches, in comparison, assume a linear effect of temperature on development rate via enzyme activity (Bonhomme 2000). The approach can be traced back to Réaumur, who in the eighteenth century proposed the idea of heat units – obtained by summing mean daily temperatures between an arbitrary date of onset and the date of an observed phenological event. Here, we describe a simple three-parameter form of a thermal-time GDD model that uses daily average temperature data and considers only the forcing phase (i.e. we ignore any chilling requirement see Chuine 2000). The parameters are start time, base temperature and cumulative GDD requirement. From the start time, daily temperatures above the base temperature are summed, and the model predicts the phenological event to occur once the cumulative GDD requirement has been met. A number of approaches exist for identifying the best parameters under the GDD model, but most involve the iterative application of combinations of parameters to the temperature data and estimation of one of several related metrics that quantify the difference between the predicted and observed phenology (Bonhomme 2000). GDD phenological models have been used most extensively in applied areas of ecology, such as forestry (Kramer 1995; Chuine 2000; Rötzer, Grote

& Pretzsch 2004), and pest management (Wilson & Barnett 1983).

In several parts of the world, Europe and North America in particular, various organizations have harnessed amateur naturalists' interest in phenology and coordinated the recording of observations across many species, years and locations (Dickinson, Zuckenberg & Bonter 2010). For abundant and charismatic species, the resulting data sets can comprise thousands or, in a few cases, hundreds of thousands of individual observations. Macroecological studies of the arising phenological data in relation to climate have provided invaluable insights into various aspects of pattern (e.g. Crick & Sparks 1999; Parmesan & Yohe 2003; Thackeray *et al.* 2010). However, if our goal is to project the effects of climate change on phenology, we need to identify the processes that underpin the relationship between variation in the environment and phenology. Previously, bottom-up approaches, whereby GDD parameters capturing process are estimated for a single population and then used to predict spatiotemporal phenological pattern, have been applied with some success (Chuine, Cambon & Comtois 2000). However, local adaptation may cause populations to differ in the cues that they respond to and in their responses to these cues: either scenario would necessitate that parameters be estimated for each population in turn, a very labour-intensive process. The objective of our study is to take the reverse approach; rather than projecting pattern on the basis of an understanding of process, we apply both the time-window and GDD approaches top-down and seek to infer process from pattern.

In this study, we use spatiotemporal data on the first leafing of *Quercus robur*, one of Northern Europe's dominant tree species. The ecological importance of *Q. robur* is portrayed in cameo by the tree's role as the base of a tri-trophic chain (oak → caterpillar → passerine), which has been much studied in the context of the impacts of climate on phenology and fitness (Visser & Both 2005). An earlier correlative study revealed that *Q. robur* leafing in the UK had advanced and that temperatures in the February to April time window were strong negative correlates of first leaf date (Sparks & Carey 1995). Given the long generation time of oaks, we can surmise that most of the interannual variation in their leafing phenology is likely to arise via phenological plasticity. Indeed, a study of *Q. robur* clones planted in phenological gardens across Europe reported that a large proportion of the inter-site variation in leafing dates could be ascribed to a plastic response to temperature (Kramer 1995). Altitudinal transplant experiments of *Q. petraea*, a close relative of *Q. robur*, revealed high levels of temperature-mediated plasticity, such that a 1 °C rise in mean spring temperature caused leaf unfolding to advance by approximately 6 days (Vitasse *et al.* 2010). On a Europe-wide scale, some genetic divergence in the timing of *Q. robur* budburst has been reported, with more northerly provenances flushing earlier than their southern counterparts when they were grown together in a common garden environment (Jensen & Hansen 2008). This pattern is consistent with countergradient local adaptation (Conover & Schulz 1995).

We combine a data set of *Q. robur* first leafing that comprises 8336 observations from across the UK over the period 1998–2009 with daily temperature data interpolated across a 5 km × 5 km grid (Perry & Hollis 2005; Perry, Hollis & Elms 2009). To these data, we apply novel implementations of the time-window and GDD models, designed to identify the thermal cues for phenology, and test whether the relationship between temperature and phenology observed across the UK can be explained via plasticity alone or whether local adaptation plays a role. We deal with non-independence associated with multiple observations in the same year or a locality by adopting mixed-modelling approaches. Day-length requirements have been shown to influence when a tree first becomes sensitive to temperature (Körner & Basler 2010), and we test whether spatial variation in day-length contributes to spatial variation in phenology. We also compare the performance of time-window and GDD approaches. Specifically, we are interested in the statistical performance of the approaches and in whether they provide congruent or conflicting insights into the processes that underpin spatiotemporal variation in leafing dates.

Materials and methods

SPATIOTEMPORAL DATA

Our phenological dataset spans the period 1998–2009. Over this period the central England yearly average temperatures (Parker, Legg & Folland 1992) were consistently well above the long-term mean, but the trend over this 12-year period was non-significant. Between 1998 and 2009 citizen scientists submitted 8336 spatially referenced observations of *Q. robur* first leafing to the UK Phenology Network (www.naturescalendar.org.uk). ‘First leaf’ was defined as when the first leaf on a tree was fully open and recognisably the shape, if not the full size, of the adult leaf. Recorders were advised that: ‘If you are finding it difficult to decide when to record, wait until the event is occurring in three plants of the same species in close proximity to each other. Record the trendsetters rather than the extraordinary’. Individual records are liable to be subject to a degree of bias by the effort of an individual recorder and the local abundance of the species. We anticipate that such bias would only pose problems for our analyses if a correlation between temperature and the degree of bias were to exist over space or time (Phillimore *et al.* 2012). We have no reason to suspect that such a bias does exist.

We used data on daily average air temperatures that had been interpolated between recording stations at a scale of 5 km × 5 km and covers most of the UK (Perry & Hollis 2005; Perry, Hollis & Elms 2009; www.metoffice.gov.uk/climatechange/science/monitoring/ukcp09/). Phenology and temperature data were placed in a spatial relational database using POSTGRESQL version 8.3.5 (PostgreSQL Global Development Group), and we used POSTGIS version 1.3.4 to store and query the data via the RPOSTGRESQL R library. Phenological records were matched with daily temperature data for the relevant 5 km × 5 km grid cell and year. The procedure of assigning interpolated mean temperature point estimates to trees means that the introduction of measurement error is unavoidable, particularly given variation in altitude and aspect within 5-km grid cells. Day-length (time from sunrise to sunset) in minutes was calculated for each day by applying sunrise and sunset equations (Meeus 1991) to the

centroid of every 5 km × 5 km grid cell. Phenological observations matched with 5 km temperature and day-length data were assigned to 150 km × 150 km grid cells, which were treated as populations in analyses. We decided upon grid cells of this size to (i) provide sufficient replication of grid cells and numbers of observations per grid cell to permit estimation of trends across grid cells and years (ii) minimise the contribution of measurement error in interpolated temperatures to relationships estimated across grid cells. The data set comprises observations from the majority of the UK, although sample sizes vary greatly over space (150-km grid cells: range = 1–1721 records per grid cell, median = 121.5) and are largest for the south-east and smallest in the far North. Sample sizes also vary among years, from a low of 50 records in 1998 to a maximum of 1174 in 2002 (median = 774.5).

TIME-WINDOW MODELS

The slope of phenology on average temperature over time versus over space can be used to identify the relative contributions of phenotypic plasticity and local adaptation (Phillimore *et al.* 2010, 2012). We assume that the cue that determines phenology is the mean temperature during a particular time window or strongly correlated with this temperature. We also assume at the population level that the mean plastic response of phenology to temperature is linear and that this response does not vary among populations. We assess the validity of these assumptions in the results section. Under this model, over long time periods, the interannual or temporal slope of phenology on temperature will be due to the contributions of plasticity and microevolution. In comparison, over a time period that is short relative to the generation time of the organism, we assume that temporal slope can be treated as an estimate of population-level phenological plasticity, due to microevolution making a negligible contribution. Differential survival of different age classes that vary in phenology can also contribute to an environment–phenology relationship (van de Pol, Osmond & Cockburn 2012), but we do not expect this to have a strong effect on the patterns reported for *Q. robur* leafing, guidance to recorders for the UK Phenology Network requests that they ‘Use mature trees (30 years old or more), as young trees show different responses’.

Plasticity may also contribute to the slope of phenology on temperature between populations (the spatial slope). However, for a resident species, sufficient generations may have elapsed for local adaptation (towards optima that covary with temperature) to contribute to the spatial slope. A spatial slope of the same sign as the temporal slope but steeper would be consistent with cogradient local adaptation. Conversely, a spatial slope of the same sign but shallower than the temporal slope would be consistent with countergradient local adaptation and may arise where local adaptation of phenology runs in the counterdirection to the effect of plasticity (Phillimore *et al.* 2010).

We estimated the slopes of phenological lag (the number of days that elapsed between the end of the time window and the phenological event) on average temperature separately across time and space. We tested whether these slopes differed significantly, using the bivariate mixed-modelling approach described in Phillimore *et al.* 2010. We treated phenological lag and temperature as a bivariate response with year and grid cell (150 km × 150 km) as random effects. Models were fitted in MCMCglmm (Hadfield 2010), and we used the default priors (for (co)variance components they are drawn from the inverse Wishart distribution with $V = 1$ and $\nu = 0.0002$, and for the mean they are drawn from a normal distribution with mean = 0

and variance = 10^8). We obtained posterior sample sizes of 2000 by running models for 23 000 iterations, sampling every 10th iteration and discarding the first 3000 iterations as burn-in. For each random effect, we obtained a posterior distribution for the variance–covariance matrix of phenological lag and average temperature. Dividing the covariance between phenological lag and temperature by the variance in temperature, we obtained a posterior distribution for the slope of phenological lag on temperature for that random effect. We estimated the spatial slope and temporal slope on the basis of grid cell and year as random effects, respectively. We tested for local adaptation on the basis of the posterior distribution of the slope difference. The null hypothesis, that the spatial slope of phenological lag against average temperatures can be attributed to plasticity alone, can be rejected if the 95% highest posterior density (HPD) of the spatial slope minus the temporal slope does not include zero.

In the above framework for estimating temperature-mediated plasticity and local adaptation, we assume that the time window over which *Q. robur* first leafing is most sensitive to ambient temperature is known, but this is not the case. For this reason, we considered a large number of time windows and compared models based on the degree to which ambient temperature explained variation in phenology.

First, we considered latitudinally invariant time windows (LITWs), where the start and end points of the window were the same for all locations. We examined different time windows by varying both the start date (from days 1 to 66 in 5-day intervals) and duration (from 5 to 120 days in 5-day intervals) of time window. We only considered LITWs where the end of the time window was before the 97.5 percentile of the spatiotemporal distribution of first leaf dates. We compared the predictive power of temperature in LITWs using a measure of pseudo- R^2 that was developed specifically for this bivariate response mixed model to treat one of the responses (average temperature, T) as a predictor of the other (phenological lag, P) across k random effects (eq. 1, Phillimore *et al.* 2010).

$$R^2 = \frac{\sum \sigma_{PR_k}^2}{\sum \sigma_{PR_k}^2 + \sum \sigma_{TR_k}^2} \quad \text{eqn 1}$$

The second type of time-window model that we considered was one where the start date of the window was initiated by a specific day-length having been reached. We explored parameter space of day-length-initiated time windows (DLITWs) by identifying the most southerly 5 km × 5 km grid cell (49.96°N, −5.21°W) and finding the day-length that corresponded to days 1–66 in 5-day intervals. The values returned were 490, 496, 504, 515, 527, 541, 556, 572, 588, 606, 624, 642, 660 and 679 min. For each observation, we identified the date in the corresponding 5 km × 5 km grid cell on which these day-lengths were reached/exceeded, and these were the start dates for DLITWs. We considered the same range of values for time-window duration as described for LITW models. In all cases, the start date was delayed further north, although the difference in the start date between the far north and south declined from 31 ordinal days on day 1 to just 3 days on ordinal day 66. We considered all DLITWs with end dates ≤ 150 days.

A consequence of considering DLITWs is that spatial variance in lag time, the response variable in the pseudo- R^2 , differs among models (Phillimore *et al.* 2012). Across most of the parameter space that we considered, the variance in DLITW model lag times across space will be reduced, depressing the pseudo- R^2 values. Therefore, we compared among DLITW models and between DLITW and LITW models by recalculating the pseudo- R^2 (eqn 1) using only the temporal random effect, which we refer to as a temporal R^2 .

GROWING DEGREE-DAY MODELS

Growing degree-day (GDD) models make a prediction of the day on which a phenological event should take place. We explored a large area of parameter space, corresponding to all combinations of start dates (both latitudinally invariant and day-length-initiated, as described above for time-window model), base temperature (0–6 °C in 0.2 °C intervals) and cumulative GDD requirement (100–600 GDDs in 20 GDD intervals). In addition, we explored models where the base temperature or cumulative GDD requirement covaried with latitude, as these represent axes along which *Q. robur* could be locally adapted (Table 1). However, finding a gradient in one or more GDD parameters would not constitute conclusive evidence for local adaptation, as we would not be able to rule out the possibility that an unmeasured variable influences one or more parameter. We scaled latitude, such that the most southerly (49.96°N) and northerly (58.57°N) 5 km × 5 km grid cells corresponded to 0 and 1, respectively. The product of scaled latitude and base temperature slope (−2 to 2 in intervals of 0.2) was added to the base temperatures, allowing base temperatures to differ by −2 to +2 °C across the latitudinal range. Similarly, the product of scaled latitude and the cumulative GDD slope (−100 to 100 in intervals of 10) was added to the cumulative GDD values, allowing the cumulative GDD requirement to differ by −100 to +100 across the observed latitudinal range.

After the start date, for each observation, we summed all temperatures above the base temperature in the corresponding 5 km × 5 km grid cell and calculated the day on which the cumulative GDD requirement for that cell was met. Temperatures below the base were treated as contributing zero GDDs.

We compared the performance of models with different parameters by calculating the unexplained variance (σ_U^2) of the difference between observed phenology and predicted phenology around an expectation of zero. Perfect prediction would therefore correspond to zero variance. To make biological interpretation of these values more intuitive, we calculated a pseudo- $R^2_{1:1}$, which quantified the degree to which observed first leafing was predicted by a 1:1 relationship with GDD predictions

$$R^2_{1:1} = 1 - \frac{\sigma_U^2}{\sigma_T^2} \quad \text{eqn 2}$$

where σ_T^2 is the variance in observed leafing dates around the grand mean. Models with negative $R^2_{1:1}$ were possible.

We ran models corresponding to all of the combinations of parameters described above, using a mixed-modelling approach with an intercept of zero and the following structure

$$y_{ij} = 0 + \beta_j + \beta_i + \beta_{ij} + e_{ij} \quad \text{eqn 3}$$

where y is the difference between observed and predicted phenology in year i and for 150 km × 150 km grid cell j . β_j , β_i and β_{ij} are random normal deviates corresponding to the random effects of grid cell, year and the grid cell × year interaction. e_{ij} is a residual term. For each model, we calculated σ_U^2 as the sum of the variance components estimated across the posterior distribution of all random effects. We calculated σ_T^2 as the sum of the variance components for grid cell, year, grid cell × year interaction and residual terms, estimated using a mixed-effects model with observed phenology as the response variable and with the grand mean estimated. We fitted these models using ASReml3 (VSN International, Hemel Hempstead, UK, Gilmore *et al.* 2009).

We projected expected phenological advances under climate warming using the best set of parameters under the GDD model where all parameters were invariant with latitude. We simulated temperature increase by raising all daily temperature values by a fixed amount

(0.1–5°C in intervals of 0.1 intervals). To each simulated temperature increase, we then applied the best set of GDD parameters and calculated the predicted first leaf dates. Projected advance was calculated as the projected phenology under warming minus the projected phenology using baseline temperatures. We then applied a mixed-modelling approach with grid cell, year and their interaction as random effects and calculated the grand mean phenological advance for each simulated temperature increase.

Table 1 provides a summary of models, parameters, their biological interpretation and the methods used to explore parameter space.

Table 1. A summary of methods, parameters and their interpretation

Method	Parameters [procedure for searching parameter space]*	Biological inference	Statistical Model	Model Comparison
Time window	Intercept [MCMC] Start date [It] End date [It] Delay of start with day-length [It] Temporal slope [MCMC] Spatial slope [MCMC]	The critical time window during which phenology is sensitive to temperature. Whether the start of the critical window is set by day-length or is latitudinally invariant. Interpreted as an estimate of the plastic response. Interpreted as the result of plasticity + local adaptation. The temporal minus spatial slope (Δb) is interpreted as an estimate of local adaptation.	Bayesian generalized linear mixed model: Response = phenology (or lag time), average temperature; fixed effects = grand means; random effects = population, year and residual.	Pseudo- R^2 for LITW models Temporal R^2 for comparing among DLITW models or between these models and LITW models.
GDD	Start date [It] Base temperature [It] Cumulative GDD requirement [It] Delay of start with day-length [It] Relationship between latitude and cumulative GDD requirement [It] Relationship between latitude and base temperature [It]	The three GDD parameters describe the mechanistic basis of a plastic response. On the basis of the GDD parameters for a single location, the time period over which trees are sensitive to temperature can be calculated post hoc. Comparison with latitudinally invariant start date tests provides a simple test of whether there is a day-length requirement. A latitudinal trend in the cumulative GDD requirement is one form that local adaptation could take. A latitudinal trend in the base temperature is a second form that local adaptation could take.	Linear mixed model: Response = observed phenology minus GDD prediction†; fixed effects = grand mean = 0; random effects = population, year, population–year interaction and residual.	Pseudo- $R^2_{1,1}$.

*It, iterative; ML, maximum likelihood; MCMC, Metropolis-coupled Monte Carlo; DLITW, day-length-initiated time windows; GDD, growing degree-day; LITW, latitudinally invariant time windows.

†Note that this is equivalent to a model with observed phenology as the response variable and predicted phenology as an offset.

Table 2. Summary of model outputs for highest values for GDD models with different combinations of free parameters

Model	Start date*	Base temperature intercept (°C)	Base temperature slope with latitude	Cumulative GDD requirement intercept	Cumulative GDD requirement slope with latitude	$R^2_{1,1}$	Spatial variance	Temporal variance	Spatial × Temporal variance	Residual variance
NULL	–	–	–	–	–	0	44.77	35.67	2.32	79.79
LI	56	1.6	–	360	–	0.470	4.43	2.37	1.01	78.31
	56	0.2	–	420	60	0.479	3.40	2.72	1.23	77.40
	56	0.2	0.8	420	–	0.477	3.52	2.67	1.29	77.55
DLI	56	1.4	–	360	–	0.476	3.58	2.63	1.06	77.85
	56	0.6	–	400	20	0.479	3.16	3.04	1.15	77.35
	56	0.6	0.4	400	–	0.478	3.12	2.99	1.18	77.56

*For DLI start date models, this value is the start date for the most southerly grid cell, and the start date will be delayed as latitude increases, see materials and methods.

Results

The earliest and latest median first leaf dates for any grid cell × year combination with multiple observations were 90 days (March 30th/April 1st) in the south of Britain in 1998 and 142.5 days (May 22nd/23rd) in the north of Britain in 2001, respectively. A mixed-effects model (referred to as the null model in Table 2) that included year and 150 km × 150 km grid cell as cross-classified random effects plus their interac-

tion revealed substantial variation in first leafing date among locations and years, while the interaction explained little of the variance. However, the largest variance component was within a grid cell in a single year. When latitude and longitude were added to the above model as fixed effects, a strong latitudinal trend in first leafing dates was revealed ($b = 1.09$ days/ $^{\circ}$ N), consistent with first leafing being 9.4 days later in the most northerly 5 km \times 5 km grid cell than the most southerly, and the spatial variance component declined from 44.77 to 3.21. No significant longitudinal trend was identified.

TIME-WINDOW ANALYSIS

Mean temperature during several overlapping LITWs yielded pseudo- R^2 values in excess of 0.5 (Fig. S1), with the highest values ($=0.53$ – 0.54) obtained for the periods 56–125, 56–130, 61–130 and 61–135 days. We only report parameter estimates in detail for the former, as it also yielded the highest temporal R^2 ($=0.95$, see Fig. 1b). Over this time window, the spatial ($b = -7.12$, 95% HPD $= -8.30$ to -5.92) and temporal ($b = -7.00$, -8.16 to -5.58) slopes were significantly negative and very similar (Fig. 1c). The residual slope within a grid

cell in a single year was also significantly negative, but shallower ($b = -4.01$, -4.37 to -3.64). Of the other 38 models within 0.05 units of the highest pseudo- R^2 model, the difference between the spatial and temporal slope was only significant for two of these (Fig. S1b). Similarly, of the 14 other models within 0.05 units of the highest temporal R^2 model, the slope difference was only significant for one of these (Fig. 1b).

Across the DLITW models, average temperature during a similar time period was identified as being most predictive (Fig. 1d). The time window that yielded the highest temporal R^2 ($=0.95$) spanned the same period as identified using LITWs in the south (56–125 days) and was 7 days later in the far north. Using this window, the slopes of phenology on temperature over space ($b = -6.54$, -7.69 to -5.17) and time ($b = -7.34$, -8.61 to -6.12) were again similar, and the difference in slopes was non-significant (Fig. 1f). The residual slope was also shallower but significantly negative ($b = -3.97$, -4.36 to -3.61). Of the 12 other models within 0.05 units of the highest temporal R^2 model, the slope difference was only significant for one of these (Fig. 1e).

When we use temporal R^2 to compare the best performing LITW and DLITW models, we find no evidence to favour

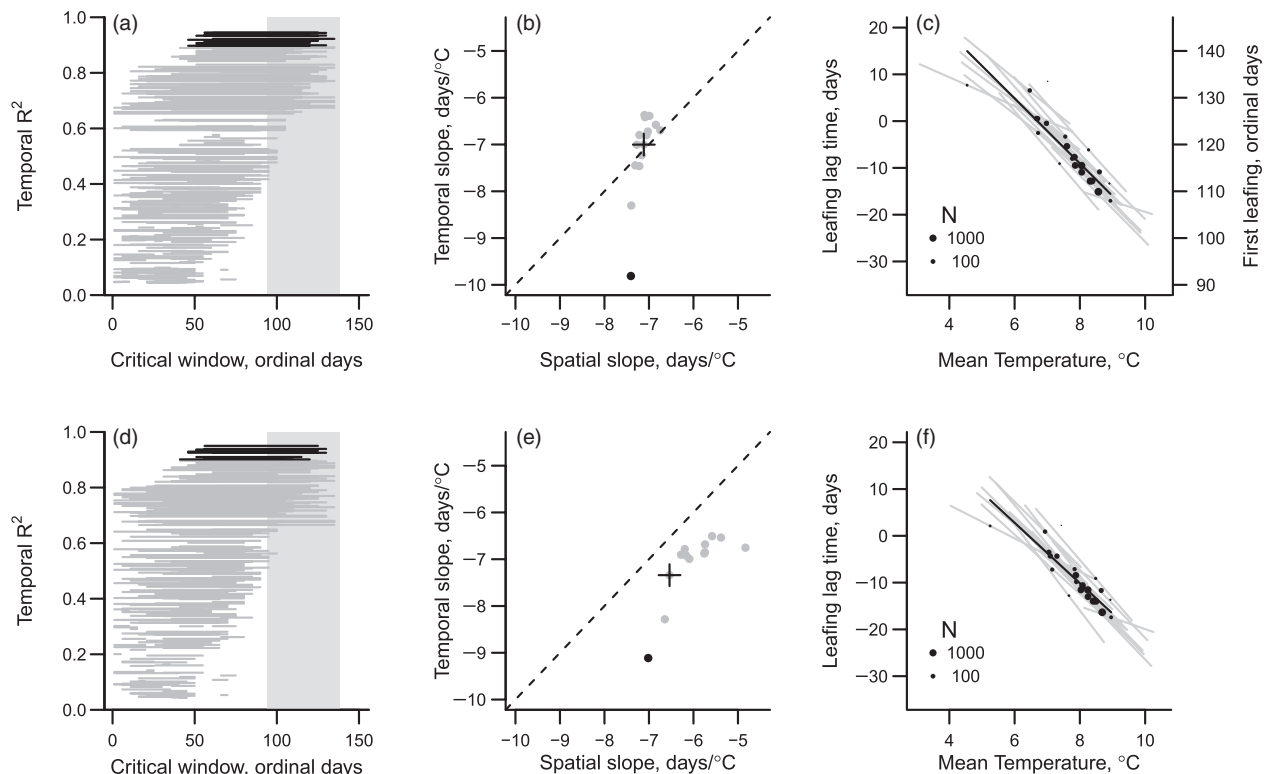


Fig. 1. A comparison of the performance of latitudinally invariant (a–c) and day-length-initiated (d–f) time-window models. (a) and (d) show how the explanatory power of temperature, quantified as temporal R^2 , varies over different time periods. In the case of (d), the time period corresponds to the most southerly grid cell, and the actual time window will be delayed in more northern grid cells, pending reaching a day-length threshold. The light grey box corresponds to the 95% confidence interval of observed first leafing dates. High R^2 models (those within 0.05 R^2 units of the highest R^2) are highlighted in black. (b) and (e) show slope estimates for the high R^2 models. Grey points = spatial and temporal slopes do not differ significantly, black points = spatial and temporal slopes differ significantly. The best model is highlighted with a cross. (c) and (f) show the relationship between average temperatures over the time window that gave rise to the highest R^2 and spatiotemporal variation in the lag time between the end of the time window and the phenological event. Grey lines represent the interannual slopes of lag time on temperature within a grid cell (temporal slopes). Black points represent grid cell mean temperature and lag time values, with the diameter of the point proportional to the log of the sample size. The black line represents the slope estimated across grid cells (spatial slope).

one type of window over the other, perhaps unsurprising given the largely overlapping time periods over which temperatures are averaged.

Taken together, our results are largely consistent with the null hypothesis that plasticity is responsible for spatial covariation between temperature and first leafing dates. All of the LITWs and DLITWs overlapped substantially with the distribution of observed phenological events (Fig. 1a & d), meaning that for some years and locations, the time window over which we consider average temperature to be a cue extends beyond the timing of first leafing. Returning to the assumptions we made in the using the space versus time slope comparison to test for local adaptation, visual inspection of Figure 1C & D reveals that slope estimates for different populations are remarkably consistent and that a linear slope across space provides a good fit to the data.

GDD ANALYSIS

There was little to choose between the highest $R^2_{1:1}$ GDD models where various parameters were held latitudinally invariant or allowed to covary with latitude (Table 2). A GDD model where all parameters were held latitudinally invariant was able to explain almost 47% of the variance in first leafing (Fig. 2a). Model explanatory power was improved ($R^2_{1:1} = 0.477$) when we incorporated and used a day-length requirement to determine the start date of the GDD model (Fig. 2b). Allowing for a covariance between the cumulative GDD requirement or base temperature and latitude also led to very marginal improvements in model explanatory power, the best GDD models being those where the cumulative GDD requirement increased with latitude ($R^2_{1:1} = 0.479$).

Of the 279 parameter combinations that yielded $R^2_{1:1}$ values within 0.01 units of the globally best model, 83% of these started on day 56 and the remaining 17% on day 51. Across these models, there was a pronounced negative correlation between the base temperature and the cumulative GDD requirement. In all cases of high $R^2_{1:1}$, the variance in leafing times between grid cells and between years was largely explained, while the residual variance (within a single grid

cell and year) remained similar to that estimated for the null model (Table 2).

Taking the GDD model with all parameters held invariant with latitude, we project an average of a 6.91-day advance in phenology for a 1 °C rise in air temperature. The relationship between temperature increase and mean phenological advance was slightly concave but close to linear across the range considered, with a projected advance of 26.63 days at the upper extreme of 5 °C.

COMPARISON OF PREDICTIONS

The days over which temperature was identified as influencing phenology were mainly overlapping between time-window and GDD methods, both for latitudinally invariant and day-length-initiated models (Fig. 3 a & c). Using both methods, we identified identical start dates. Under GDD models, the end date of the sensitive period occurred before the end of the time window in warm years, but extended a little beyond it in cold years. There was also a pronounced tendency for temperature sensitivity to span a longer time period further north, although this tendency was less pronounced under the best day-length-initiated model (Fig 3c).

Predicted phenology in grid cell \times year combinations under time-window and GDD models was highly consistent, regardless of whether start date was latitudinally invariant or day-length initiated (Fig. 3b & d). The only systematic bias was a tendency for the very earliest predicted dates under time-window approaches to be earlier than the very earliest dates predicted from a GDD model.

Discussion

A plastic response of first leafing to temperature explained approximately half of the spatiotemporal variation in first leafing, irrespective of whether we adopted a time-window or GDD approach. Note that we assume the temporal trend is plastic rather than proving this, given that our data do not permit assessment of trends for individual trees. Under time-window and GDD models, we project that a 1 °C rise in

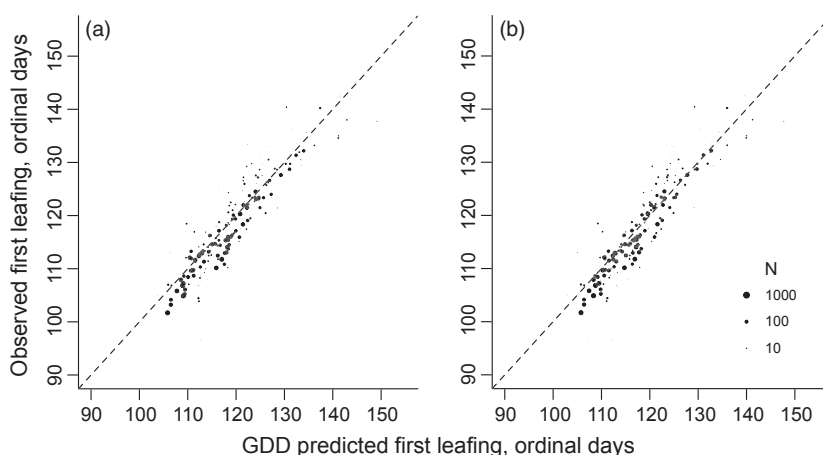
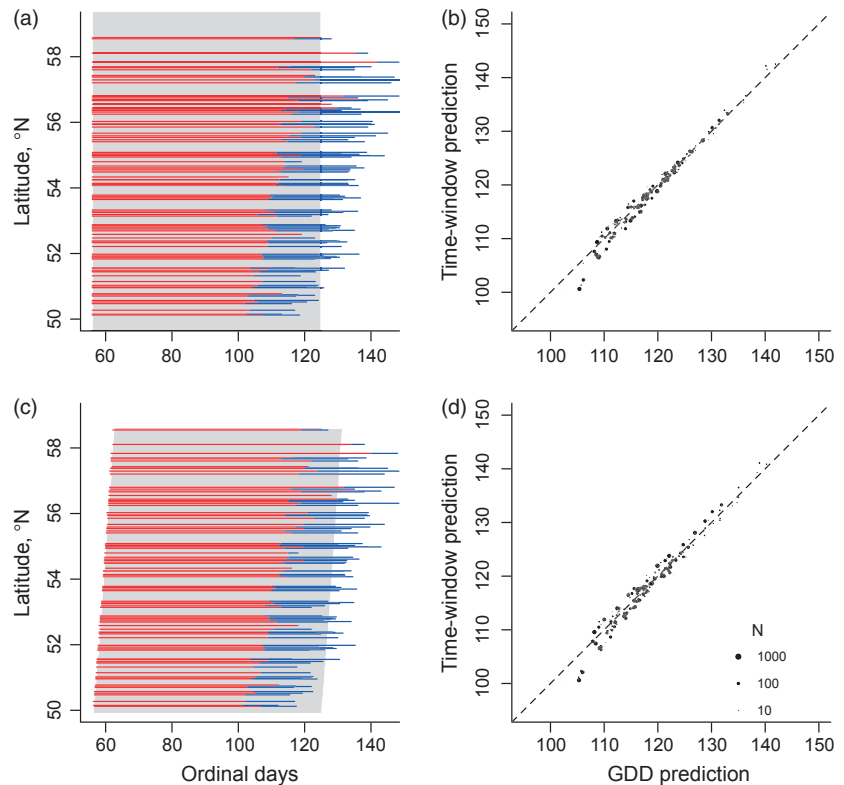


Fig. 2. The relationship between predicted and observed phenology under the highest $R^2_{1:1}$ (a) latitudinally invariant and (b) day-length-initiated GDD models, where all other parameters are held latitudinally invariant. Data points are mean values for grid cell \times year combinations. The dashed line corresponds to unity.

Fig. 3. Plots (a) and (c) compare the time periods during which phenology is sensitive to temperature under time-window (grey polygon) and GDD approaches (horizontal lines) under latitudinally invariant (a) and day-length-initiated (c) models. GDD model parameters are reported in Table 2 (keeping base temperature and the cumulative GDD requirement latitudinally invariant). Lines correspond to mean values in 50 km \times 50 km grid cells, with red lines indicating the days that contribute to the GDD model in warm years and blue lines indicating the additional days that contribute in cold years. Plots (b) and (d) show the relationship between the predicted values in grid cell (150 km) \times year combinations under the best time-window and GDD models. The dashed line corresponds to unity, and the diameter of points is proportional to the log sample size.



daily temperatures would cause first leafing to advance by 7.3 and 6.9 days earlier, respectively. These findings are in qualitative agreement with the temperature-mediated phenological plasticity across *Q. robur* clones (~ 2 days, $^{\circ}\text{C}^{-1}$) planted in different phenological gardens (Kramer 1995) and *Q. petraea* (c. 6 days, $^{\circ}\text{C}^{-1}$) subject to elevation transplants (Vitasse *et al.* 2010). Our elevated slope estimates as compared with Kramer's are likely to be due to us considering average temperature during a short time window, whereas Kramer temperatures spanned the entire period November 1st – leaf unfolding. Substantial temperature-mediated plasticity of spring phenology appears to be commonplace among temperate species (Sparks & Carey 1995; Charmantier *et al.* 2008; Phillimore *et al.* 2012). Obtaining estimates of temperature-mediated phenological plasticity for other species that are likely to be affected by oak phenology (e.g., competitors, understory plants, invertebrates and birds) should be a priority for future work. These estimates can then be used to project the phenology of each species relative to the oak yardstick under future climate scenarios (Visser & Both 2005).

With both time-window and GDD models, we found *Q. robur* to be most sensitive to ambient temperatures from around day 56, extending for anything between 34 and 96 days. This period overlaps substantially with the time periods identified by studies of *Q. robur* in Norfolk (UK) (Sparks and Carey 1995) and *Q. petraea* subject to altitudinal transplants in the Pyrenees (Vitasse *et al.* 2010). In terms of the variation in phenology explained, there was little to choose between models where the start date was latitudinally invariant versus day-length initiated, with the latter performing margin-

ally better. If temperature sensitivity is initiated by a day-length requirement, a start date around 56 days corresponds to a period where the geographic variation in day-length is less pronounced than earlier in the year; consequently, we may have little statistical power to detect what may be a small effect size. Interestingly, a study of *Q. robur* clones planted across European phenological gardens did not identify a photoperiod threshold for leafing (Kramer 1995).

We found little evidence that first leafing is locally adapted with respect to temperature. While the slopes of first leafing on average temperature were significantly different over space versus time for only a minority of the high R^2 time windows, there was a general trend for them to be shallower over space than time, consistent with countergradient variation (Phillimore *et al.* 2012). GDD results largely agreed with those from time-window analyses inasmuch as that models where parameters were held to be invariant with latitude (consistent with variation being due to plasticity alone) explained very nearly as much of the phenological variation as models where one of the GDD parameters covaried with latitude (consistent with a role for local adaptation). The very weak evidence we found for the heat sum requirement increasing with latitude actually is contrary to the expectation if countergradient variation were present. Moreover, this finding is difficult to reconcile with the results of a common garden trial of oak provenances from across Europe, which identified local adaptation of flushing time on a continent-wide scale and a weak negative correlation between provenance latitude and flushing date, with the two UK provenances adhering to this general trend (Jensen & Hansen 2008). One possible

explanation for the weak and slightly conflicting evidence for local adaptation that we find may be that the value of *Q. robur* as a commercial crop has meant that provenances have been transplanted across much of Europe (Jensen & Hansen 2008), and this may have masked the effects of local adaptation. We note that another tree species that is native to the UK, *Pinus sylvestris*, shows local adaptation on a relatively fine geographic scale (Salmela *et al.* 2011). Two implications of the lack of evidence we find for phenological local adaptation of *Q. robur* are that it may be acceptable to substitute space for time to project the effects of temperature on phenology under future climate change scenarios (Phillimore *et al.* 2012) and that first leafing may not need to evolve to cope with at least a moderate amount of warming.

Irrespective of the modelling approach, a substantial proportion (in the region of half) of the variation in first dates was left unexplained. This residual variation is likely to comprise phenological observation error, plasticity with respect to unaccounted for environmental variables and genetic differences. Explanatory power will also be diminished by the discrepancy between the 5-km grid cell interpolated temperature data and the temperatures experienced by individual trees. However, by considering only first dates, we are likely to have substantially overestimated the proportion of the variance in oak first leafing that we can explain. This is because we ignore the substantial phenological among-tree variation that is known to exist within a single location and year (Crawley & Akhteruzzaman 1998), the basis of which is likely to be partially environmental and partially genetic.

STATISTICAL CONSIDERATIONS

One of our aims in this study was to compare the insights and statistical performance of time-window and GDD models, all of which we found to be remarkably similar. Below, we outline the relative strengths and weaknesses of each approach, as implemented in this study.

The Achilles heel of the time-window approach is the biological implausibility of average temperature in the time window being a cue, given that the time window extends beyond the timing of many of the events (Gienapp, Hemerik & Visser 2005; Phillimore *et al.* 2012). A comparison of the time periods that the time-window and GDD approaches identify as being important for *Q. robur* leafing (Figs. 3a & c) may partially resolve the problem. It appears that under warmer conditions, the time required for a GDD model is much shorter than the extent of the time window, whereas under colder conditions, the time required for the GDD model can extend beyond the time window. Considering this finding alongside the similarity of predictions between time-window and GDD approaches (Figs 3b & d), we suggest that even if the actual thermal cues underpinning phenology are in fact closer to a GDD model, the time-window approach may still provide a useful approximation and can deliver useful insights. The time-window approach offers the additional benefit that only two parameters (start date and end date) need to be searched iteratively, while the remainders are optimized in the model.

The GDD approach has the advantage that the model is biologically motivated and plausible. However, while the GDD model requires fewer parameters, the iterative exhaustive searching of parameter space is a highly inefficient optimization approach. We also only considered latitude as one axis along which local adaptation may cause GDD parameters to vary geographically, but other axes, such as the average timing of the last frost, are perhaps better candidates and could be incorporated into refinements in the future.

In summary, we have shown that time-window and GDD approaches provide congruent insights into the cues and processes that underpin geographic variation in the spring phenology of *Q. robur* and both approaches have comparable explanatory power. Macro-scale data sets, such as the ones considered here, have considerable potential to contribute to our understanding of the processes underpinning phenological variation.

Acknowledgements

We thank citizen scientist contributors to the UK Phenology Network and the Woodland Trust for providing access to the data. We are grateful to D. Inouye, L. McInnes, A. Pigot and two anonymous reviewers for comments that greatly improved the manuscript and A. Bento, M. Crawley, R. Ennos, J. Hadfield, A. Humphreys, D. Orme and R. Lande for insightful discussions. A.B.P. was funded by an Imperial College London Junior Research Fellowship and NERC advanced fellowship (Ne/I020598/1). The UK Climate Projections data have been made available by the Department for Environment, Food and Rural Affairs (Defra) and Department for Energy and Climate Change (DECC) under license from the Met Office, Newcastle University, University of East Anglia and Proudman Oceanographic Laboratory. These organizations accept no responsibility for any inaccuracies or omissions in the data, nor for any loss or damage directly or indirectly caused to any person or body by reason of, or arising out of, any use of these data.

References

- 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.
- Bonhomme, R. (2000) Bases and limits to using 'degree.day' units. *European Journal of Agronomy*, **13**, 1–10.
- Charmanier, A., McCleery, R.H., Cole, L.R., Perrins, C., Kruuk, L.E.B. & Sheldon, B.C. (2008) Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*, **320**, 800–803.
- Chuine, I. (2000) A unified model for budburst of trees. *Journal of Theoretical Biology*, **207**, 337–347.
- Chuine, I., Cambron, G. & Comtois, P. (2000) Scaling phenology from the local to the regional level: advances from species-specific phenological models. *Global Change Biology*, **6**, 943–952.
- Conover, D.O. & Schulz, E.T. (1995) Phenotypic similarity and the evolutionary significance of countergradient variation. *Trends in Ecology & Evolution*, **10**, 248–252.
- Crawley, M.J. & Akhteruzzaman, M. (1998) Individual variation in the phenology of oak trees and its consequences for herbivorous insects. *Functional Ecology*, **2**, 409–415.
- Crick, H.Q.P. & Sparks, T.H. (1999) Climate change related to egg-laying trends. *Nature*, **399**, 423–424.
- 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.
- Fitter, A.H. & Fitter, R.S.R. (2002) Rapid changes in flowering time in British plants. *Science*, **296**, 1689–1691.
- Forrest, J. & Miller-Rushing, A.J. (2010) Toward a synthetic understanding of the role of phenology in ecology and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 3101–3112.

- Gienapp, P., Hemerik, L. & Visser, M.E. (2005) A new statistical tool to predict phenology under climate change scenarios. *Global Change Biology*, **11**, 600–606.
- Gilmore, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2009) ASReml user guide release 3.0.
- Hadfield, J.D. (2010) MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software*, **33**, 1–22.
- Harrington, R., Woiwod, I. & Sparks, T. (1999) Climate change and trophic interactions. *Trends in Ecology and Evolution*, **14**, 146–150.
- Husby, A., Nussey, D.H., Visser, M.E., Wilson, A.J., Sheldon, B.C. & Kruuk, L.E.B. (2010) Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit (*Parus major*) populations. *Evolution*, **64**, 2221–2237.
- Jensen, J.S. & Hansen, J.K. (2008) Geographical variation in phenology of *Quercus petraea* (Matt.) Liebl and *Quercus robur* L. oak grown in a greenhouse. *Scandinavian Journal of Forest Research*, **23**, 179–188.
- Körner, C. & Basler, D. (2010) Phenology Under Global Warming. *Science*, **327**, 1461–1462.
- Kramer, K. (1995) Phenotypic plasticity of the phenology of seven European tree species in relation to climatic warming. *Plant, Cell and Environment*, **18**, 93–104.
- Marsham, R.A. (1789) Indications of Spring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **79**, 154–156.
- Meeus, J.H. (1991) *Astronomical Algorithms*. Willmann-Bell, Richmond, Virginia.
- Menzel, A. & Fabian, P. (1999) Growing season extended in Europe. *Nature*, **397**, 659.
- Nussey, D.H., Postma, E., Gienapp, P. & Visser, M.E. (2005) Selection on heritable phenotypic plasticity in a wild bird population. *Science*, **310**, 304–306.
- Parnesan, C. & Yohe, G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.
- Parker, D.E., Legg, T.P. & Folland, C.K. (1992) A new daily Central England temperature series, 1772–1991. *International Journal of Climatology*, **12**, 317–342.
- Perry, M. & Hollis, D. (2005) The generation of monthly gridded datasets for a range of climatic variables over the UK. *International Journal of Climatology*, **25**, 1041–1054.
- Perry, M., Hollis, D. & Elms, M. (2009) *Climate Memorandum No 24: The Generation of Daily Gridded Datasets of Temperature and Rainfall for the UK*. National Climate Information Centre, Met Office, National Climate Information Centre. Exeter, UK.
- Phillimore, A.B., Hadfield, J.D., Jones, O.R. & Smithers, R.J. (2010) Differences in spawning date between populations of common frog reveal local adaptation. *Proceedings of the National Academy of Sciences*, **107**, 8292–8297.
- Phillimore, A.B., Stålhandske, S., Smithers, R.J. & Bernard, R. (2012) Dissecting the contributions of plasticity and local adaptation to the phenology of a butterfly and its host plants. *American Naturalist*, **180**, 655–670.
- van de Pol, M. & Cockburn, A. (2011) Identifying the critical climatic window that affects trait expression. *American Naturalist*, **177**, 698–707.
- van de Pol, M., Osmond, H.L. & Cockburn, A. (2012) Fluctuations in population composition dampen the impact of phenotypic plasticity on trait dynamics in superb fairy-wrens. *Journal of Animal Ecology*, **81**, 411–422.
- Roberts, A.M.I. (2012) Comparison of regression methods for phenology. *International Journal of Biometeorology*, **56**, 707–717.
- Rötzer, T., Grote, R. & Pretzsch, H. (2004) The timing of budburst and its effect on tree growth. *International Journal of Biometeorology*, **48**, 109–118.
- Salmela, M.J., Cavers, S., Cottrell, J.E., Iason, G.R. & Ennos, R.A. (2011) Seasonal patterns of photochemical capacity and spring phenology reveal genetic differentiation among native Scots pine (*Pinus sylvestris* L.) populations in Scotland. *Forest Ecology and Management*, **262**, 1020–1029.
- Sparks, T.H. & Carey, P.D. (1995) The responses of species to climate over two centuries: an analysis of the Marsham phenological record, 1736–1947. *Journal of Animal Ecology*, **64**, 321–329.
- Thackeray, S.J., Sparks, T.H., Frederiksen, M., Burthe, S., Bacon, P.J., Bell, J.R. *et al.* (2010) Trophic level asynchrony in rates of phenological change for marine, freshwater and terrestrial environments. *Global Change Biology*, **16**, 3304–3313.
- van de Pol, M. & Cockburn, A. (2011) Identifying the critical climatic window that affects trait expression. *The American Naturalist*, **177**, 698–707.
- van de Pol, M., Osmond, H.L. & Cockburn, A. (2012) Fluctuations in population composition dampen the impact of phenotypic plasticity on trait dynamics in superb fairy-wrens. *Journal of Animal Ecology*, **81**, 411–422.
- Visser, M.E. & Both, C. (2005) Shifts in phenology due to global climate change: the need for a yardstick. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 2561–2569.
- Visser, M.E., Caro, S.P., van Oers, K., Schaper, S.V. & Helm, B. (2010) Phenology, seasonal timing and circannual rhythms: towards a unified framework. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 3113–3127.
- Visser, M.E. & Holleman, L.J.M. (2001) Warmer springs disrupt the synchrony of oak and winter moth phenology. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **268**, 289–294.
- Vitasse, Y., Bresson, C.C., Kremer, A., Michalet, R. & Delzon, S. (2010) Quantifying phenological plasticity to temperature in two temperate tree species. *Functional Ecology*, **24**, 1211–1218.
- Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.-M., Hoegh-Guldberg, O. & Bairlein, F. (2002) Ecological responses to recent climate change. *Nature*, **416**, 389–395.
- Willis, C.G., Ruhfel, B., Primack, R.B., Miller-Rushing, A.J. & Davis, C.C. (2008) Phylogenetic patterns of species loss in Thoreau's woods are driven by climate change. *Proceedings of the National Academy of Sciences*, **105**, 17029–17033.
- Wilson, L.T. & Barnett, W.W. (1983) Degree-days: an aid in crop and pest management. *California Agriculture*, **37**, 4–7.

Received 14 September 2012; accepted 15 January 2013
Handling Editor: David Gibson

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. a) The extent to which average temperature during a LITW explains spatiotemporal variation in first leafing. b) Slope estimates for the high R2 models.