Predicting insect phenology across space and time

J. A. HODGSON*, C. D. THOMAS*, T. H. OLIVER†, B. J. ANDERSON*, T. M. BRERETON‡ and E. E. CRONE§

Department of Biology Area 18, Wentworth Way, University of York, York YO105DD, UK, †Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, OX108BB, UK, ‡Butterfly Conservation, Manor Yard, East Lulworth, Dorset BH205QP, UK, §Department of Ecosystem and Conservation Sciences, College of Forestry and Conservation, University of Montana, Missoula, MT 59812, USA

Abstract

Many species appear to be undergoing shifts in phenology, arising from climate change. To predict the direction and magnitude of future changes requires an understanding of how phenology depends on climatic variation. Species show large-scale spatial variation in phenology (affected by differentiation among populations) as well as variation in phenology from year-to-year at the same site (affected predominantly by local plasticity). Teasing apart spatial and temporal variation in phenology should allow improved predictions of phenology under climate change. This study is the first to quantify large-scale spatial and temporal variation in the entire emergence pattern of species, and to test the relationships found by predicting future data. We use data from up to 33 years of permanent transect records of butterflies in the United Kingdom to fit and test models for 15 butterfly species. We use generalized additive models to model spatial and temporal variation in the distribution of adult butterflies over the season, allowing us to capture changes in the timing of emergence peaks, relative sizes of peaks and/or number of peaks in a single analysis. We develop these models using data for 1973–2000, and then use them to predict phenologies from 2001 to 2006. For six of our study species, a model with only spatial variation in phenology is the best predictor of the future, implying that these species have limited plasticity. For the remaining nine species, the best predictions come from a model with both spatial and temporal variation in phenology; for four of these, growing degree-days have similar effects over space and time, implying high levels of plasticity. The results show that statistical phenology models can be used to predict phenology shifts in a second time period, suggesting that it should be feasible to project phenologies under climate change scenarios, at least over modest time scales.

Keywords: adaptation, butterfly monitoring scheme, climate change, emergence time, growing degree-days, latitude, Lepidoptera, phenotypic plasticity

Received 21 December 2009 and accepted 25 July 2010

Introduction

Many species appear to be undergoing shifts in phenology as a result of climate change (Sparks & Carey, 1995; Crick *et al.*, 1997; Menzel & Fabian, 1999; Roy & Sparks, 2000; Stefanescu *et al.*, 2003), and phenological changes constitute a high proportion of all of the evidence that species are responding to climate change (Walther *et al.*, 2002; IPCC, 2007). Thanks to comprehensive biological recording, butterflies have become a model group in detecting the impacts of climate change, including impacts on phenology (Sparks & Yates, 1997; Roy & Sparks, 2000; Forister & Shapiro, 2003; Stefanescu *et al.*, 2003; Dell *et al.*, 2005; Gordo & Sanz, 2006; Menzel *et al.*, 2006). However, understanding and predicting these changes remains more of a challenge (Visser, 2008). The climate change literature is replete with statements about the

potential detrimental effects of phenology shifts because they could lead to significant changes in population growth rates and in community interactions (van Asch *et al.*, 2007; Doi *et al.*, 2008). For example, temporal mismatches between plants and their pollinators could reduce seed set in plants (including crops) and reduce food availability for the pollinators (Memmott *et al.*, 2007). To predict whether such issues will become greater in the future demands an ability to predict phenology shifts, based on climatic variation.

In insects, a variety of life history parameters (e.g., induction of diapause, size at pupation, maternal investments, etc.) may evolve in order to maximize fitness under the prevailing climate and the historic range of variability of that climate (Davidson, 1944; Lees, 1950; Manly, 1974; Visser, 2008). Insects respond to very specific cues which, over their evolutionary history, have helped them to stay in synchrony with important seasonal events, such as the bud burst of trees (van Asch *et al.*, 2007). Temperature and day length (photoperiod) may

Correspondence: J. A. Hodgson, tel. + 441904 328623, e-mail: jh69@york.ac.uk

both be used as cues (Bale *et al.*, 2002). Because different species have evolved to use different cues, and because climate change affects temperature but not day length, this could lead to maladaptive responses (Doi *et al.*, 2008).

A number of studies have used year-to-year variation in phenology (usually earliest emergence or mean emergence date) regressed on yearly climate to predict future changes (e.g., Roy & Sparks, 2000; Gordo & Sanz, 2006). As well as temporal changes, many species also show spatial variation in phenology (e.g., Zhou *et al.*, 1995; Langvatn *et al.*, 1996; White *et al.*, 1997), but a given climate variable may have different effects over space and over time (Pollard, 1991; Rock *et al.*, 1993; Thompson & Clark, 2006; Doi *et al.*, 2008; Doi & Takahashi, 2008; Forkner *et al.*, 2008). For example, the leafing date of Japanese trees varies less with the gradient of temperature over space than it does with changes in temperature over time at a single site (Doi & Takahashi, 2008). If phenology varies to different extents in space and in time this implies that species do not simply respond physiologically to temperature cues. Instead they are likely to be, in part, locally adapted and/or responsive to spatially fixed cues, such as daylength (Bradshaw & Holzapfel, 2001; Visser, 2008).

Teasing apart the spatial variation in phenology, which is affected by local differentiation of populations, and the temporal variation, which is affected by phenotypic plasticity within a population, should allow much better predictions of phenology under climate change than looking at one (space/time) dimension in isolation. However, few studies have considered both spatial and temporal variation in phenology (Rock *et al.*, 1993; Doi *et al.*, 2008; Doi & Takahashi, 2008), and none of these has explicitly attempted to predict the future.

In this study, we use data from up to 33 years of permanent transect records of butterflies in the United Kingdom to fit and test models of the phenology of 15 butterfly species, modelling both spatial and temporal variation in relation to the climate. We use generalized additive models (GAMs) (Wood, 2006) to model the distribution of adult butterflies over the season, independently of the overall abundance at a particular site and year, allowing us to capture changes in the timing of emergence peaks, relative sizes of peaks and/or number of peaks in a single analysis. We develop these models using data for the time period 1973–2000, and then use each model to predict phenologies for the period 2001–2006. The testing data enables us to assess our ability to predict detailed phenological patterns into the future. Furthermore, the relative predictive power of alternative models that have different predictor variables allows us to make inferences about the factors that regulate phenology in different species.

Methods

Butterfly data

We used butterfly abundance data from UK Butterfly Monitoring Scheme (collected and maintained by Butterfly Conservation and the Centre for Ecology and Hydrology; Pollard & Yates, 1993; Asher *et al.*, 2001). Butterflies of all species are recorded along a permanent transect at weekly intervals between April and September at a network of sites across the country. The number of sites taking part in the scheme has increased from one in 1973 to 750 in 2006. In order to fit models to roughly half the data, and use the other half to test model predictions, we split the data set into 1973–2000 and 2001–2006 periods.

We focused on the 15 species that are known to have multiple generations per year in at least part of their British range, since these have the most complex phenological patterns and might be expected to be the most difficult to model and predict future changes. We ran all the analyses separately for each species, using only sites where the species was found at a reasonable frequency (usually using the condition that the median count over all weeks and years should be greater than one, but this condition had to be relaxed for four species that were particularly rare and/or temporally variable: holly blue, wall brown, small blue and wood white; common English and Latin names are given in Table 2).

Climate data

We used historical monthly mean temperature data at the 10 km grid resolution, from averages of the 5 km grid resolution data available from the Met Office ('UKCP09' data, Met Office, 2009). We decided to use growing degree-days above 5° Celsius (henceforth 'GDD5') as a measure of the warmth available for development and reproduction that is relevant to all butterfly species (even though there will also be various effects of winter minima, summer maxima, rainfall, cloudiness, etc.) (Romo & Eddleman, 1995; Bryant *et al.*, 2002). From the monthly mean temperatures we calculated the GDD5 for each UKBMS site (henceforth actual GDD5) in each year; defining 'year' as the months between October of the previous year (just after the previous transect season finished) and September (when the transect season in question finished). Across the United Kingdom, GDD5 has been increasing in the time period of this study, but there is also considerable year-to-year variation (Fig. 1). The average increase in GDD5 between the period 1973–2000 and the period 2001–2006 is c.200, which is nearly half of the interannual variation (variation in medians among years), and 10% of the spatial variation in GDD5 that exists across all British 10 km grid squares within a single year (Fig. 1).

Because we wanted to separate the effects of spatial variation in the climate from temporal variation in the climate, we also calculated (a) mean GDD5 for each site over the period 1973–2000 (henceforth 'site GDD5') and (b) the difference between each site's actual GDD5 in each year and the site mean (henceforth 'GDD5 differential'). Hence each site and year has three interrelated GDD5 measures: (actual GDD5)$_t$ = (site GDD5) + (GDD5 differential)$_t$ where t indexes years.

Statistical analyses

We used GAMs with a negative binomial error structure to model the emergence pattern as a flexible, smooth function of the time of year (measured as weeks from the first week in April). By using 2-D and 3-D smooths, we could represent the fact that emergence pattern would change gradually with the climate over space and/or time. We used the GAM implementation in the MGCV package in R (Wood, 2003, 2008; R Development Core Team, 2009) with the extension to negative binomial family in the MASS package (Venables & Ripley, 2002). Because the data consisted of butterfly counts (of a certain species at a given site, week and year), but we wanted to investigate the emergence pattern independently of the
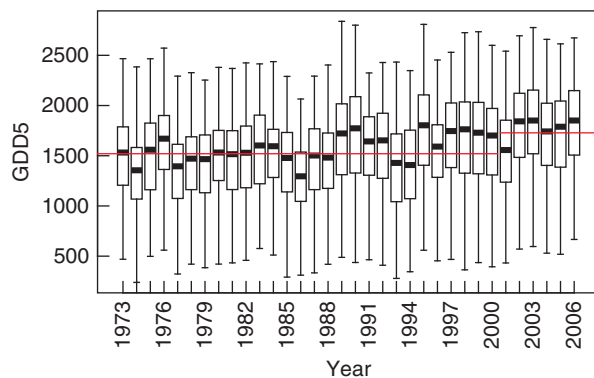
Fig. 1 The distribution of GDD5 across all 10 km squares in the United Kingdom (boxplots of the median, interquartile range and extremes) in each of the 33 years of the study. Red lines show the overall mean for the period of fitting data (1973–2000) and the period of testing data (2001–2006).

overall abundance of butterflies, we always included the average abundance for the site and year as a linear covariate in the GAM. Specifically, this linear covariate was calculated as: $\text{mean}[\ln(\text{count}_i + 1/27)]$ where i indexes the week, and there are usually 26 weeks (sometimes fewer because of missed surveys). The mean is calculated in log space because the negative-binomial GAM has a log link function, and a constant $1/27$ is added to the count because zeros cannot be logged. Using this average abundance as a covariate means that the variation remaining to be explained is the distribution of individuals among the weeks of a given year – which is the main quantity of interest – and that the observation error structure relevant to such count data is retained.

We fitted seven GAMs, representing different hypotheses about the drivers of phenology, to each species' data for the early time period (1973–2000). We evaluated the GAMs on their ability to predict the later time period (2001–2006). Details of the seven models used are given in Table 1. The spatial component of climate is represented either by site GDD5 or by site northing: the latter should be a better predictor of phenology if the species responds more to day length than temperature. The temporal component is represented by GDD5 differential. If the basic responses to temperature are the same everywhere (i.e., different populations are not adapted to the climate differently) then the actual GDD5 experienced by each site in each year might be the best predictor of phenology.

Model comparison

The exact likelihood of each observation (x) given our model can be calculated as the probability density at x of the negative binomial distribution with mean given by the GAM prediction (according to the explanatory variables specific to that observation), and shape parameter as fitted by the GAM (the same

Table 1 Summary of 7 GAM models fitted to each species' data

Code number	Model variables*	Max. no. of parameters	Interpretation
00	Abundance	2	No climate effects and no seasonality
0	00 + smooth (week)	12	No climate effects
1	00 + smooth (week × site GDD5)	25	Spatial variation only (suggests response to fixed site-specific cues that are correlated with site temperature; possibly local photoperiod adaptation)
2	00 + smooth (week × northing)	25	Spatial variation only (suggests response to fixed cues that are correlated with latitude; probably photoperiod)
3	00 + smooth (week × actual GDD5)	25	Space-for-time substitution (suggests plastic response to temperature)
4	00 + smooth (week × site GDD5 × GDD5 diff)	41	Different spatial and temporal effects (suggests site-specific responses, as in model 1, combined with flexible temperature responses)
5	00 + smooth (week × northing × GDD5 diff)	41	Different spatial and temporal effects (suggests likely photoperiod response, as in model 2, combined with flexible temperature response)

*Each site and year has three interrelated GDD5 measures: $(\text{actual GDD5})_t = (\text{site GDD5}) + (\text{GDD5 differential})_t$. Site GDD5 is the mean GDD5 for the site for the period 1973–2000.

shape parameter for all observations under a given model). Taking logs of these probabilities and summing over all observations (either observations in the fitting period, or observations in the testing period) gave us the log likelihood, *L*. The proportion of deviance 'explained' was taken to be how far *L* lies between Lnull (likelihood under model 00, see Table 1) and Lsat. Lsat is the maximum likelihood possible for those data: that which would occur under a model where predictions are exactly the same as observations (i.e., a saturated model) and where a Poisson distribution (no overdispersion) is assumed. So proportion deviance explained = 1−(*L*-Lsat)/(Lnull-Lsat).

Our results focus on comparing the log likelihood of the testing (2001–2006) data under each GAM. Using a separate test data set effectively penalizes overfitted models, and the model with the highest log likelihood can be said to be the best model (i.e., no GAM parameters were estimated for any model using the testing data). We can also use −2*L* to calculate Akaike weights (Burnham & Anderson, 2002) to give us a 'degree of belief' in a certain model compared to others. However, the Akaike weighting is strongly dependent on the assumption that observations are independent – an assumption that was not really tenable for our data, where there were multiple observations from exactly the same site. To show which of our conclusions are robust to this assumption, we report raw model weights, and weights where each observation only counts as 1/12 of a degree of freedom (*L* becomes *L*/12). The factor 12 was chosen based on the autocorrelation of residuals in the testing data. Residuals from consecutive weeks at the same site and the same year had rank correlation coefficients >0.5, but correlation dropped quickly with lag >1 week, so roughly every other week's data could be considered a pseudoreplicate. Residuals from the same week at the same site for different years had rank correlation coefficients >0.5, irrespective of the time lag in years – suggesting that of 6 years of data, five may be pseudoreplicates. Dividing by two for nonindependent weeks and by six for nonindependent years resulted in the factor 12. Although this 'correction factor' is crude, it should be noted that there is no perfect way of assessing independence, and that the main results in this paper are robust to quite large changes in the weighting.

Results

For all species except the small tortoiseshell, the GAMs predict 15–30% of the deviance in the testing data (Table 2, Fig. 2). The amount of deviance explained is only slightly lower in the testing than in the training data (Table 2), showing relatively good explanatory performance of the models. There is strong support for at least some spatio-temporal variability in phenology for all the species (i.e., models incorporating variation in space and/or time perform consistently and significantly better than 'null' week-only models; final two columns in Table 2).

For six species, phenologies are spatially variable but seem little affected by year-to-year changes in temperature; a model with only spatial variation in phenology is the best predictor of the 'future' 2001–2006 dynamics (model 1 or 2 have the best fit: Table 2, Fig. 3). Of these

six species, latitude alone is the best predictor for five species (common blue, speckled wood, brown argus, small blue, wood white), whereas for the remaining one (holly blue) the site's average GDD5 is the best predictor. This implies that, for these six species, phenology is relatively fixed at a given site (or that it responds to climate variables that are not correlated with GDD5).

The best-fitting models for the remaining nine species suggest that they exhibit some temporal variation in their phenology, depending on the thermal conditions from year to year and site to site. For four of these species (small white, large white, Adonis blue, wall brown), the model based on actual GDD5 (the GDD5 experienced in year X at site Y, model 3) is the best at predicting the future – their phenologies appear entirely flexible with respect to GDD5 (Table 2, Fig. 3). For the remaining five species (green-veined white, small tortoiseshell, small heath, small copper, comma), fixed spatial variables (northing or average site GDD5) combine with responses to annual variation in GDD5 to determine observed emergence patterns (model 4 or 5 best; Table 2, Fig. 3). For these five species, the model based on actual GDD5 (model 3) still outperforms purely spatial models (models 1–2, Table 2, Fig. 3). This demonstrates that, even though there may not be perfect space-for-time equivalence, there is a degree of consistency in the response to temperature in different places (Table 2, Fig. 3).

For nine species, northing is a consistently better spatial covariate than site average GDD5, i.e. the Akaike weight of models with northing vs. those with GDD5 is close to 1 (Table 2). This suggests that day length could play a role in determining their phenologies, without necessarily implying site-specific local adaptation. For six species, the opposite is true (low weight of models with northing, compared to site average GDD5; Table 2); this suggests that phenological changes are triggered by cues that are correlated with the average GDD5 of each site, and this could potentially include population-specific photoperiod responses. Hypotheses about population and species differences in photoperiod responses remain to be tested experimentally.

There is great variation between species in exactly how phenology varies over space and time. The number of emergence peaks appears to change in space and/or time for seven out of the 15 species (Figs 4 and 5, Fig. S1). Six of these species show a greater number of peaks in the southern (hotter) parts of their ranges than in the north [common blue (Fig. 4a), wood white, small blue, small tortoiseshell, small copper, small heath (Fig. S1)]. Two of these species show an increased tendency to produce an extra peak in the south in relatively hot years [small copper (Fig. 5a and b), small heath (Fig. S1)]. The speckled wood (Fig. 4b) appears to be an exception

Table 2 Summary of results for each species

Species common English name	Latin name	No. obervations in 1973–2000 period	No. obervations in 2001–2006 period	Best % explained in fitting data (1973–2000)	Best % explained in testing data (2001–2006)	Best model at predicting testing data	Second best model at predicting testing data	Akaike weight of model 3 relative to 1 and 2*	Akaike weight of model 3 with df correction*	Akaike weight of the models with northing vs. those with site GDD5*	Akaike weight of models with northing with df correction*	Akaike weight of model 0 relative to all others*	Akaike weight of model 0 with df correction*
Holly blue	*Celastrina argiolus*	38 157	33 980	18	15	1	0	0.00000	0.00003	0.00000	0.00260	0.08465	0.08465
Common blue	*Polyommatus icarus*	21 900	17 347	23	23	2	5	0.00000	0.00000	1.00000	1.00000	0.00000	0.00000
Speckled wood	*Pararge aegeria*	42 782	33 303	18	16	2	3	0.00000	0.00299	1.00000	0.99998	0.00000	0.00000
Brown argus	*Aricia agestis*	2488	2151	24	22	2	0	0.00000	0.00011	1.00000	0.88283	0.12504	0.12504
Small blue	*Cupido minimus*	877	933	37	30	2	5	0.00001	0.18354	0.99999	0.76765	0.06272	0.06272
Wood white	*Leptidea sinapis*	911	803	37	23	2	3	0.08686	0.31741	0.99784	0.53666	0.07144	0.07144
Small white	*Pieris rapae*	29 420	20 736	20	19	3	4	1.00000	1.00000	0.00000	0.25855	0.00000	0.00000
Large white	*Pieris brassicae*	20 281	15 059	20	20	3	4	1.00000	1.00000	0.04052	0.43445	0.00000	0.00000
Adonis blue	*Polyommatus bellargus*	1030	591	36	27	3	1	0.99997	0.60770	0.00000	0.18239	0.05358	0.05358
Wall brown	*Lasiommata megera*	35 070	15 463	25	20	3	5	1.00000	0.99991	1.00000	0.99950	0.00000	0.00000
Green-veined white	*Pieris napi*	35 484	20 578	20	19	4	5	1.00000	1.00000	0.00000	0.00000	0.00000	0.00000
Small tortoiseshell	*Aglais urticae*	14 272	8252	10	4	4	3	1.00000	0.99984	0.00000	0.00000	0.00000	0.00000
Small heath	*Coenonympha pamphilus*	20 111	12 892	22	21	5	4	0.99988	0.64557	1.00000	0.93065	0.00000	0.00000
Small copper	*Lycaena phlaeas*	7184	3087	22	18	5	4	1.00000	0.90559	1.00000	0.91856	0.00001	0.00001
Comma	*Polygonia calbum*	2021	2538	21	18	5	3	1.00000	0.77540	1.00000	0.75212	0.00335	0.00335

*Model weights are by definition between 0 and 1, but are only shown to five decimal places.
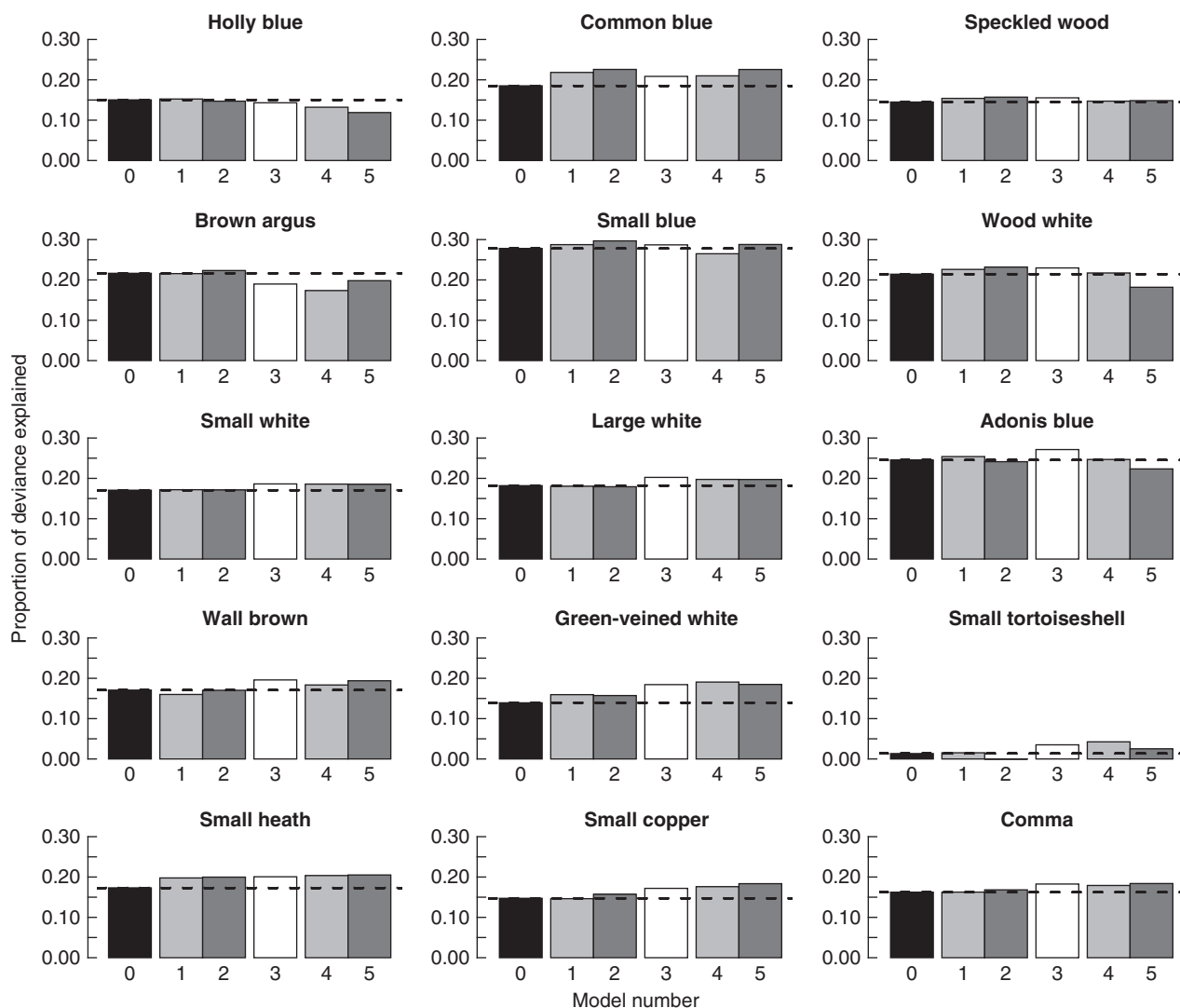
Fig. 2 Comparison of deviance in an independent test dataset explained by the GAMs 0–5. Bars are shaded to aid the comparison of similar models: black when phenology is a smooth function of week of the year, with no spatial or temporal variation, light grey when site GDD5 is the spatial covariate, dark grey when northing is the spatial covariate, and white when actual GDD5 is the covariate (combining both spatial and temporal variation). Key to models: 0, week only; 1, site GDD5; 2, northing; 3, actual GDD5; 4, site GDD5 × GDD5 diff; 5, northing × GDD5 diff. Horizontal dotted line is always at the level of GAM 0, for comparison. GAM, generalized additive model; GDD5, growing degree-days above 5 °C.

in that there are fewer emergence peaks in the far south (2) than further north (3). However, because of the complex life cycle of this species, fewer peaks might not imply fewer generations per year (see Discussion).

For other species, the number of peaks does not seem to change, but the relative sizes of the different peaks do change. The holly blue (Fig. 4c), green veined white, and large white (Fig. S1), show higher late summer peaks in the south than in the north. The brown argus (Fig. 4d) shows a higher first peak in the south. In the comma butterfly, the relative size of the middle peak is larger in the south than in the north (Fig. 5c and d). Finally, for some species the timing of the peaks changes, with earlier

emergences in the south and in hotter years (e.g., wall brown, green-veined white, large white, Fig. S1).

Discussion

This is the first study to quantify large-scale spatial and temporal variation in the entire emergence pattern of species, and to test the relationships found by predicting future data. Even though a large proportion of the deviance remained unexplained, predictions from our models were surprisingly robust (similar amounts of deviance were explained in the testing as in the training data). Because the data we fitted consisted of transect
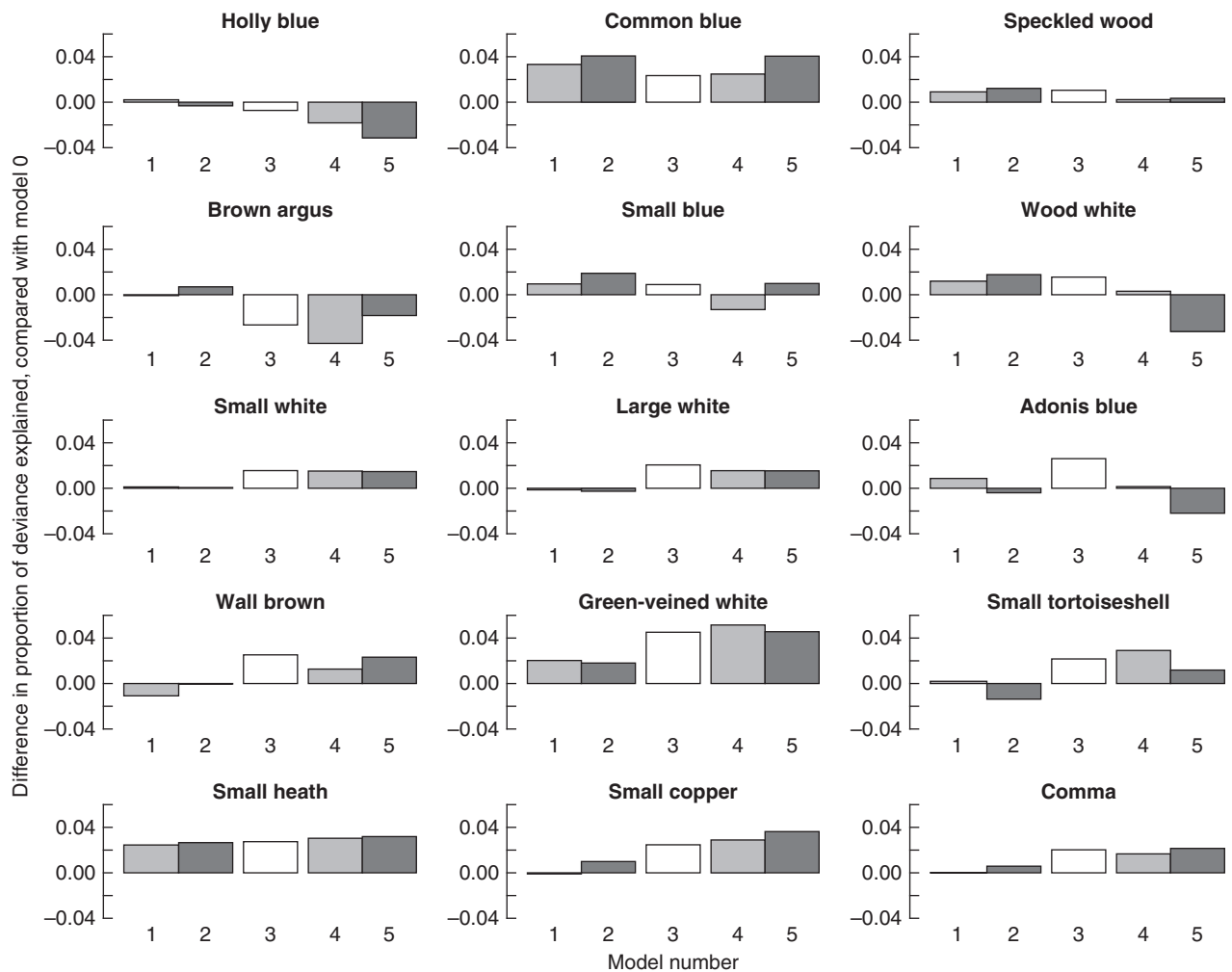
Fig. 3 Comparison of deviance in an independent test dataset explained by GAMs 1–5 compared with GAM 0 (where phenology is a smooth function of week of the year, with no spatial or temporal variation). Bars are shaded to aid the comparison of similar models: light grey when site GDD5 is the spatial covariate, dark grey when northing is the spatial covariate, and white when actual GDD5 is the covariate (combining both spatial and temporal variation). Key to models: 0, week only; 1, site GDD5; 2, northing; 3, actual GDD5; 4, site GDD5 × GDD5 diff; 5, northing × GDD5 diff. For the absolute amounts of deviance explained, see Fig. 2. GAM, generalized additive model; GDD5, growing degree-days above 5 °C.

counts in particular weeks and sites, which are inherently variable for many reasons, we would never expect predictions to have a very high precision, but this should not prevent them from predicting large-scale trends. The results show that statistical phenology models can be used to predict phenology shifts in a second time period, suggesting that it should be feasible to project phenologies under climate change scenarios, at least over modest time scales.

Our method isolates 'phenology' in the sense of the distribution of adults over the season, independently of the overall abundance in a particular site or year. This is a desirable first step to understanding the underlying biology: it is more meaningful than other commonly used metrics such as the date of first observation, mean

date or date range, which can be biased by changes in overall abundance or changes in voltinism. The entire pattern of emergence is relevant to the interactions between species, such as between herbivores and plants or between plants and pollinators. Observing the distribution of adults over the season in relation to habitat and climate factors also permits fuller consideration of how climate change alters the interactions among species, and how it might change selection pressures.

The variety of phenological patterns observed

One of the desirable features of our GAM modelling approach is its flexible ability to represent the variety of phenological patterns seen in butterflies, especially
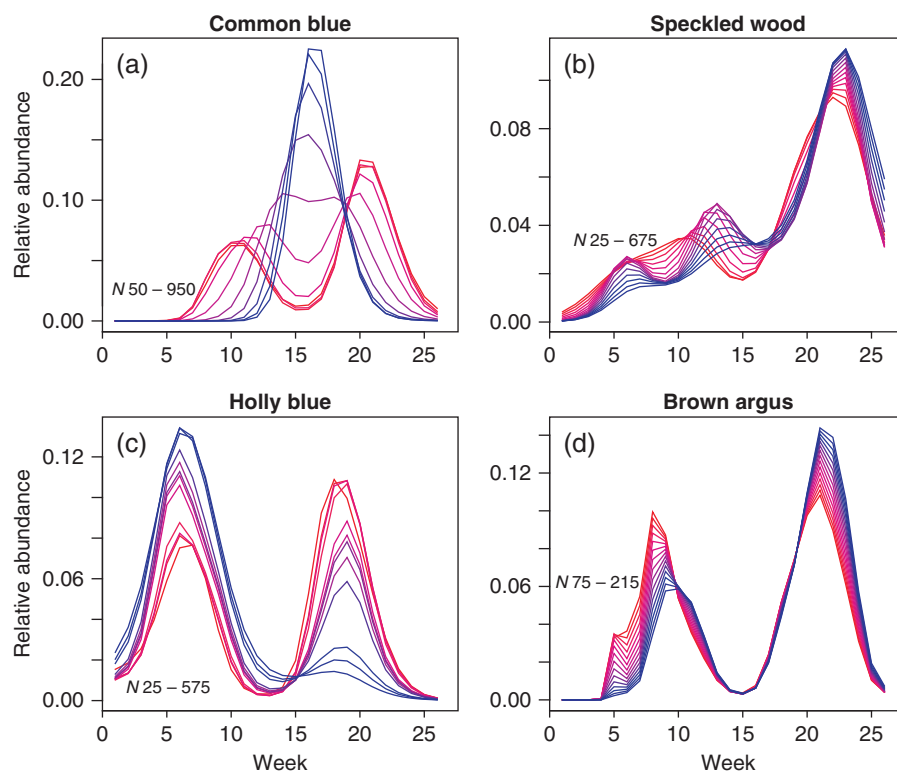
Fig. 4 Spatio-temporal variation in phenology of selected species where the best-predicting generalized additive model (GAM) included only spatial variation. In each panel, the unlogged, GAM-fitted distribution of butterflies across the 26 weeks of transect surveys is shown by lines, standardized so that the sum of all weeks equals 1. Different lines represent different values of Northing: equally spaced steps between the southernmost (red) and the northernmost (blue) parts of the species' data (the range, in British National Grid 10 km squares is also printed in the panel). For example, the common blue shows one peak in the north and two in the south. Equivalent figures for all other species are given in supplementary material, Fig. S1.

multivoltine butterflies. The patterns of emergence peaks in the GAMs (Figs 4 and 5 and S1) are largely what we would expect given what is known about the voltinism of these species. For example, the common blue was already known to have univoltine populations in the north and bivoltine ones in the south and this is picked up very clearly in the GAM (Fig. 4a). The northern peak is timed midway between the two southern peaks, suggesting that it would not be simple for a population to switch between the two modes over time, and thus is not surprising that the northing-only model predicted best for this species.

The speckled wood also shows a complex pattern, broadly consistent with what is known about its phenology. This species has a complex life cycle, overwintering in both larval and pupal stages (Nylin *et al.*, 1995; Blakeley, 1997). The pupae emerge somewhat before the larvae in spring, producing two small peaks in the first half of the flight season (Fig. 4b). Going from north to south, the two small peaks overlap more and more until they appear as one extended peak. So we do

not think this decrease in number of emergence peaks represents a decrease in the numbers of generations the speckled wood achieves in a year – in fact it may represent an increase. There could be two generations per year in the south but only three generations per 2 years in the north (the third peak being the offspring of the 1st peak and the 2nd peak being the offspring of the 3rd peak in the previous year). It is a little surprising that the northing-only model gave the best predictions for this species, because it has been observed elsewhere that development times are very dependent on temperature, and that the flight period can be extended (start earlier and finish later) in relatively warm years at a single site (Shreeve, 1986). This highlights the potential for more detailed species-specific models to explain more of the observed variation by including additional climatic and other variables (see below).

For other species, the GAMs pick up changes in the relative sizes of the different peaks. The holly blue (Fig. 4c), green veined white, and large white (Fig. S1), show higher late summer peaks in the south than in the
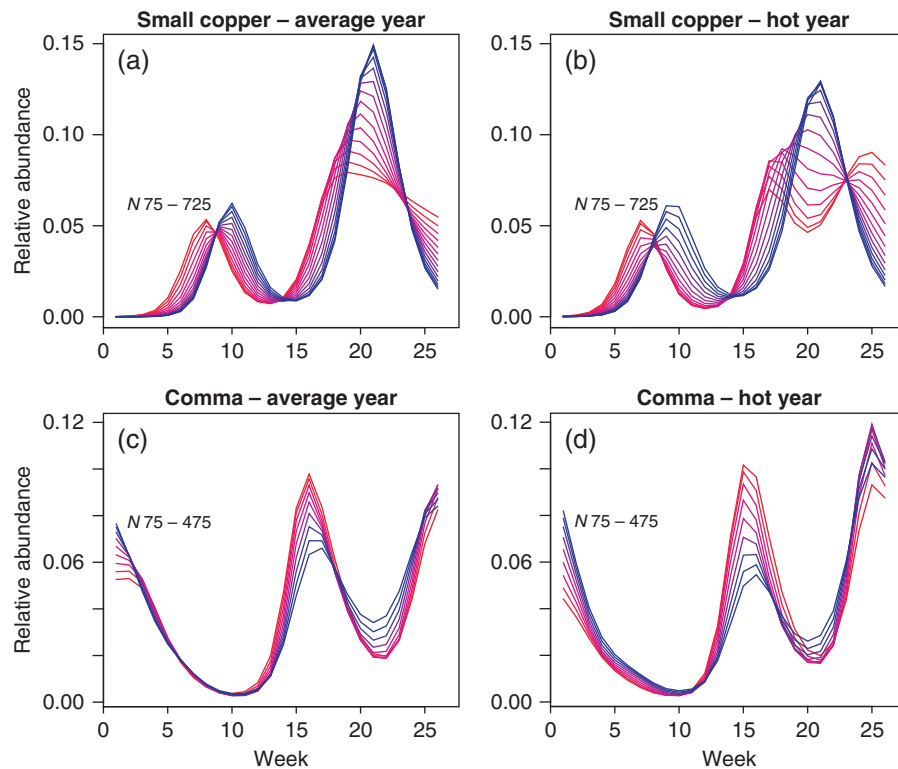
Fig. 5 Spatio-temporal variation in phenology of selected species where the best-predicting generalized additive model (GAM) included temporal as well as spatial variation. In each panel, the unlogged, GAM-fitted distribution of butterflies across the 26 weeks of transect surveys is shown by lines, standardized so that the sum of all weeks equals 1. Different lines represent different values of Northing: equally spaced steps between the southernmost (red) and the northernmost (blue) parts of the species' data (the range, in British National Grid10 km squares is also printed in the panel). To illustrate temporal variation, predictions for an average year [growing degree-days above 5 °C (GDD5) differential = 0] and a hot year (GDD5 differential = 200), are shown in adjacent panels. For example, the small copper exhibits two distinct peaks in mid-late summer in the hot south, but only in hot years. Equivalent figures for all other species are given in supplementary material, Fig. S1.

north. For the holly blue and the green-veined white, the second peak in the north is smaller than the first, and may represent only a partial second generation. For the large white the second peak is always substantially bigger than the first, so it seems more reasonable to assume that there are two complete generations per year, but that breeding success and/or survival during summer are higher in the warm south. The brown argus (Fig. 4d) shows a higher first peak in the south, seemingly because the first adult peak is extremely small in northern locations rather than because summer breeding success is increased in the north (although the latter is possible because of reduced parasitism in the north; Menendez *et al.*, 2008). In the comma butterfly, the relative size of the middle peak is larger in the south than in the north (Fig. 5c and d). Two types of comma adults are produced in the mid-summer peak: bright orange, active individuals that breed again in the same year, with their offspring emerging in late summer; and darker, less active individuals that will overwinter directly as adults and that will not breed until the

following year (Nylin, 1989, 1992). Therefore, the shift most likely reflects an increase in the frequency of two generations per year in the south.

Insight into the drivers of phenology

Our statistical models can potentially capture the net effects of phenological adaptations (the average and slope of the reaction norms) that differ between populations, and make successful predictions without necessarily knowing what the climate triggers are. It is likely, as has been shown for many insects, that phenological events are triggered by a combination of temperature and day-length cues. For example, the developmental pathways of the comma butterfly, for which model 5 (indicating an interaction effect of latitude and GDD5 differential) was the best predictor, are known to be affected by both photoperiod and temperature (Nylin, 1989, 1992). Any ongoing evolution of the reaction norms will cause mismatches between our models and future data. To predict this evolution is a very difficult task, and

would require more detailed knowledge of the traits that underlie the realized phenology (Visser, 2008). However, our results suggest testable predictions about which species respond to which triggers.

The relative performance of our five alternative models suggest that different species currently have different amounts of regional adaptation of phenology, and different levels of plasticity in phenology. Some species show quite flexible space-for-time substitution (i.e., the phenology can simply be predicted by knowing the temperature accumulated in a given site/year), but others show patterns of emergence that are less flexible over time. Six out of the 15 species' phenologies were best fit solely by fixed spatial effects (latitude or average GDD5 of a site), four species showed flexible space-for-time substitution, while five species showed complex patterns that were best fitted by an interaction between spatial and temporal variation in phenology (i.e., change in phenology over time was dependent on the spatial location of the site).

Species with fixed spatial effects have, prima facie, no plasticity and therefore may only respond to climate change gradually, through evolutionary changes (e.g., evolution of critical photoperiods that induce diapause, Bradshaw & Holzapfel, 2001). When species exhibit space-for-time equivalence this suggests that reaction norms are the same in all populations, and that no evolution is required to respond to moderate climate change. Therefore we may expect these species to 'keep up' with climate change, and we might also have greater confidence in the predictions of our models for future years for these species. When species exhibit space–time interactions, this suggests that reaction norms are different in different geographic areas. Therefore we may expect these species to show an intermediate rate of response to climate change, or to respond in some geographic areas but not others. It is not clear whether year-to-year plasticity necessarily increases fitness, compared to a fixed phenology strategy – presumably this would also depend on the phenological responses of the resources, mutualists and natural enemies of the species under consideration. It would be very interesting to design some experimental work on some selected species, to test whether the extent of plasticity, and the type of physiological cues, are consistent with the predictions here.

Limitations

We have focused on how the GDD5 summed over a whole year affects the entire sequence of emergence, and in doing so we have inevitably glossed over many of the climatic and biological details that underlie these phenomena. Better fits could undoubtedly be obtained

in future studies by including additional climate variables, particularly those that show little correlation with GDD5. The small tortoiseshell in particular could be an interesting subject for further study because so much variation was left unexplained by temperature or northing. The speckled wood deserves further consideration because its phenology responds to early and late season temperatures (Shreeve, 1986) that were not fully encapsulated within our GDD5 approach, and the same may also be true for other species. It may be argued that better predictions of the future would be obtained from a mechanistic model of the species' development, life-history and population dynamics. In several experimental studies it has been shown that degree-days have quite a direct relationship to the development rate of each larval instar of insects (Manel & Debouzie, 1997; Gu & Novak, 2006; Thompson & Clark, 2006; Raworth, 2007; Trnka et al., 2007). These relationships are often used to build sophisticated models of the phenology of certain well-known species, such as crop pests (e.g., Collier & Finch, 1985; Trnka et al., 2007). However, such models usually need to be calibrated differently for different geographic areas, reflecting the variability of development conditions in the field relative to laboratory conditions, microclimate differences among sites and individuals, and the genetic differentiation of populations (Collier & Finch, 1985; Bryant et al., 2002; Trnka et al., 2007). Because physiological data will be required for large numbers of populations of every species, it is unlikely to be feasible to parameterize such models for the majority of species for which we would like to predict phenological responses.

While our model projections performed well in testing data, the predictions were only for a relatively short testing period (2001–2006), immediately following the model training period (1973–2000), and the climate during the testing period remained largely within the bounds of conditions experienced during the training period (Fig. 1). As such, the testing data were independent, but we remain cautious about the capacity of such models (or other models) to project phenologies far into the future. Reaction norms may evolve, especially once the new climate falls outside the range of the conditions previously experienced by those populations, and cause future phenologies to deviate systematically from model projections. This would be true even for species that already have high levels of phenotypic plasticity in phenology. For multivoltine species, such as these butterflies, extra generations may be observed under future, hotter climates, which could accelerate the pace of demographic and evolutionary responses.

Although we remain cautious about long-term projections, the approach we have taken appears suitable to

model and predict short-term phenology patterns in space and time. The statistical models presented here can be parameterized from the sort of phenology data that are regularly collected by volunteer recorders, and this approach can straightforwardly be applied to large numbers of species. The models have a predictive power which is reasonably high and surprisingly consistent across species. We see this as an important step towards anticipating which species might be threatened as a result of phenological mismatches as the climate changes.

References

Asher J, Warren M, Fox R, Harding P, Jeffcoate G, Jeffcoate S (2001) *The Millennium Atlas of Butterflies in Britain and Ireland*. Oxford University Press, Oxford.

Bale J, Masters G, Hodkinson I *et al.* (2002) Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. *Global Change Biology*, **8**, 1–16.

Blakeley D (1997) *Overwintering biology of Pararge aegeria*. MPhil thesis, University of Leeds.

Bradshaw WE, Holzapfel CM (2001) Genetic shift in photoperiodic response correlated with global warming. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 14509–14511.

Bryant SR, Thomas CD, Bale JS (2002) The influence of thermal ecology on the distribution of three nymphalid butterflies. *Journal of Applied Ecology*, **39**, 43–55.

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York.

Collier RH, Finch S (1985) Accumulated temperatures for predicting the time of emergence in the spring of the cabbage root fly, *Delia radicum* (L) (Diptera, Anthomyiidae). *Bulletin of Entomological Research*, **75**, 395–404.

Crick HQP, Dudley C, Glue DE, Thomson DL (1997) UK birds are laying eggs earlier. *Nature*, **388**, 526–526.

Davidson J (1944) The relationship between temperature and rate of development of insects at constant temperatures. *Journal of Animal Ecology*, **13**, 26–38.

Dell D, Sparks TH, Dennis RLH (2005) Climate change and the effect of increasing spring temperatures on emergence dates of the butterfly *Apatura iris* (Lepidoptera : Nymphalidae). *European Journal of Entomology*, **102**, 161–167.

Doi H, Gordo O, Katano I (2008) Heterogeneous intra-annual climatic changes drive different phenological responses at two trophic levels. *Climate Research*, **36**, 181–190.

Doi H, Takahashi M (2008) Latitudinal patterns in the phenological responses of leaf colouring and leaf fall to climate change in Japan. *Global Ecology and Biogeography*, **17**, 556–561.

Forister ML, Shapiro AM (2003) Climatic trends and advancing spring flight of butterflies in lowland California. *Global Change Biology*, **9**, 1130–1135.

Forkner RE, Marquis RJ, Lill JT, Le Corff J (2008) Timing is everything? Phenological synchrony and population variability in leaf-chewing herbivores of *Quercus*. *Ecological Entomology*, **33**, 276–285.

Gordo O, Sanz JJ (2006) Temporal trends in phenology of the honey bee *Apis mellifera* (L.) and the small white *Pieris rapae* (L.) in the Iberian Peninsula (1952–2004). *Ecological Entomology*, **31**, 261–268.

Gu WD, Novak RJ (2006) Statistical estimation of degree days of mosquito development under fluctuating temperatures in the field. *Journal of Vector Ecology*, **31**, 107–112.

IPCC (2007) *Climate Change 2007 – Impacts, Adaptation and Vulnerability: Contribution of Working Group II to the Fourth Assessment Report of the IPCC*. Cambridge University Press, Cambridge.

Langvatn R, Albon SD, Burkey T, CluttonBrock TH (1996) Climate, plant phenology and variation in age of first reproduction in a temperate herbivore. *Journal of Animal Ecology*, **65**, 653–670.

Lees AD (1950) Diapause and photoperiodism in the fruit tree red spider mite (*Metatetranychus ulmi* Koch). *Nature*, **166**, 874–875.

Manel S, Debouzie D (1997) Modeling insect development time of two or more larval stages in the field under variable temperatures. *Environmental Entomology*, **26**, 163–169.

Manly BFJ (1974) Estimation of stage-specific survival rates and other parameters for insect populations developing through several stages. *Oecologia*, **15**, 277–285.

Memmott J, Craze PG, Waser NM, Price MV (2007) Global warming and the disruption of plant–pollinator interactions. *Ecology Letters*, **10**, 710–717.

Menendez R, Gonzalez-Megias A, Lewis OT, Shaw MR, Thomas CD (2008) Escape from natural enemies during climate-driven range expansion: a case study. *Ecological Entomology*, **33**, 413–421.

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

Menzel A, Sparks TH, Estrella N, Roy DB (2006) Altered geographic and temporal variability in phenology in response to climate change. *Global Ecology and Biogeography*, **15**, 498–504.

Met Office (2009) *UKCP09: Gridded observation data sets*. Available at http://www.metoffice.gov.uk/climatechange/science/monitoring/ukcp09/index.html (accessed 8 October 2009).

Nylin S (1989) Effects of changing photoperiods in the life-cycle regulation of the Comma butterfly, *Polygonia-c-album* (Nymphalidae). *Ecological Entomology*, **14**, 209–218.

Nylin S (1992) Seasonal plasticity in life-history traits - growth and development in *Polygonia-c-album* (Lepidoptera, Nymphalidae). *Biological Journal of the Linnean Society*, **47**, 301–323.

Nylin S, Wickman PO, Wiklund C (1995) Life-cycle regulation and life-history plasticity in the speckled wood butterfly – are reaction norms predictable? *Biological Journal of the Linnean Society*, **55**, 143–157.

Pollard E (1991) Changes in the flight period of the hedge brown butterfly *Pyronia-tithonus* during range expansion. *Journal of Animal Ecology*, **60**, 737–748.

Pollard E, Yates TJ (1993) *Monitoring Butterflies for Ecology and Conservation*. Chapman & Hall, London.

Raworth DA (2007) Initiation of oviposition after winter diapause in the spider mite *Tetranychus urticae* (Acari : Tetranychidae): prediction and historical patterns. *Population Ecology*, **49**, 201–210.

R Development Core Team (2009) *R 2.9.1: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rock GC, Stinner RE, Bacheler JE, Hull LA, Hogmire HW (1993) Predicting geographical and within-season variation in male flights of 4 fruit pests. *Environmental Entomology*, **22**, 716–725.

Romo J, Eddleman L (1995) Use of degree-days in multiple-temperature experiments. *Journal of Range Management*, **48**, 410–416.

Roy DB, Sparks TH (2000) Phenology of British butterflies and climate change. *Global Change Biology*, **6**, 407–416.

Shreeve TG (1986) The effect of weather on the life cycle of the speckled wood butterfly *Pararge aegeria*. *Ecological Entomology*, **11**, 325–332.

Sparks TH, Carey PD (1995) The responses of species to climate over 2 centuries – an analysis of the Marsham phenological record, 1736–1947. *Journal of Ecology*, **83**, 321–329.

Sparks TH, Yates TJ (1997) The effect of spring temperature on the appearance dates of British butterflies 1883–1993. *Ecography*, **20**, 368–374.

Stefanescu C, Penuelas J, Filella I (2003) Effects of climatic change on the phenology of butterflies in the northwest Mediterranean Basin. *Global Change Biology*, **9**, 1494–1506.

Thompson R, Clark RM (2006) Spatio-temporal modelling and assessment of within-species phenological variability using thermal time methods. *International Journal of Biometeorology*, **50**, 312–322.

Trnka M, Muska F, Semeradova D, Dubrovsky M, Kocmankova E, Zalud Z (2007) European Corn Borer life stage model: regional estimates of pest development and spatial distribution under present and future climate. *Ecological Modelling*, **207**, 61–84.

van Asch M, Tienderen PH, Holleman LJM, Visser ME (2007) Predicting adaptation of phenology in response to climate change, an insect herbivore example. *Global Change Biology*, **13**, 1596–1604.

Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*, 4th edn. Springer, New York.

Visser ME (2008) Keeping up with a warming world; assessing the rate of adaptation to climate change. *Proceedings of the Royal Society B-Biological Sciences*, **275**, 649–659.

Walther GR, Post E, Convey P *et al.* (2002) Ecological responses to recent climate change. *Nature*, **416**, 389–395.

White MA, Thornton PE, Running SW (1997) A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Global Biogeochemical Cycles*, **11**, 217–234.

Wood SN (2003) Thin-plate regression splines. *Journal of the Royal Statistical Society (B)*, **65**, 95–114.

Wood SN (2006) *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC, New York.

Wood SN (2008) Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society (B)*, **70**, 495–518.

Zhou XL, Harrington R, Woiwod IP, Perry JN, Bale JS, Clark SJ (1995) Effects of temperature on aphid phenology. *Global Change Biology*, **1**, 303–313.

Supporting Information

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

Figure S1. phenology of all species as predicted by their best GAM. In each panel, the unlogged, GAM-fitted distribution of butterflies across the 26 weeks of transect surveys is shown by lines, standardized so that the sum of all weeks equals 1. Different lines represent different values of Northing: equally-spaced steps between the southernmost (red) and the northernmost (blue) parts of the species' data (the range, in British National Grid 10 km squares is also printed in the panel). For species where the best model has both spatial and temporal variation, two panels are shown: one for an average year in the earlier time period (GDD5 differential = 0) and one for an average year in the later time period (GDD5 differential = 200).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.