BUILDING PHENOLOGICAL MODELS FROM PRESENCE/ABSENCE DATA FOR A BUTTERFLY FAUNA

James H. Thorne,[1,4] Joshua O'Brien,[1] Matthew L. Forister,[2] and Arthur M. Shapiro[3]

[1]*Department of Environmental Science and Policy, University of California, 1 Shields Avenue, Davis, California 95616 USA*
[2]*Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York 11794-5245 USA*
[3]*Center for Population Biology, University of California, Davis, California 95616 USA*

Abstract. Species phenology is increasingly being used to explore the effects of climate change and other environmental stressors. Long-term monitoring data sets are essential for understanding both patterns manifest by individual species and more complex patterns evident at the community level. This study used records of 78 butterfly species observed on 626 days across 27 years at a site in northern California, USA, to build quadratic logistic regression models of the observation probability of each species for each day of the year. Daily species probabilities were summed to develop a potential aggregate species richness (PASR) model, indicating expected daily species richness. Daily positive and negative contributions to PASR were calculated, which can be used to target optimum sampling time frames. Residuals to PASR indicate a rate of decline of 0.12 species per year over the course of the study. When PASR was calculated for wet and dry years, wet years were found to delay group phenology by up to 17 days and reduce the maximum annual expected species from 32.36 to 30. Three tests to determine how well the PASR model reflected the butterfly fauna dynamics were all positive: We correlated probabilities developed with species presence/absence data to observed abundance by species, tested species' predicted phenological patterns against known biological characteristics, and compared the PASR curve to a spline-fitted curve calculated from the original species richness observations. Modeling individual species' flight windows was possible from presence/absence data, an approach that could be used on other similar records for butterfly communities with seasonal phenologies, and for common species with far fewer dates than used here. It also provided a method to assess sample frequency guidelines for other butterfly monitoring programs.

Key words: butterflies; California coast range; climate change; long-term monitoring; phenological modeling; phenology; potential aggregate species richness; presence/absence data.

INTRODUCTION

Many ecological processes are not replicable in the laboratory or in manipulative experiments, and the only way to understand them is through systematic, long-term monitoring (Diamond 1983). Biological monitoring data can be used to discern the signal of environmental change, including global warming, by decoding signals in the biological dynamics of species sampled over long periods of time. Phenology, the timing of life history events (Gurevitch et al. 2002), is of particular interest because it reflects annual cycles. Life cycles can be affected by changing environmental conditions, and phenological patterns have been correlated with environmental phenomena for a wide range of species and environments. These include warmer temperatures associated with first leaf and flower dates for lilac, apple, and grape (Wolfe et al. 2005, Cayan et al. 2001, Chiune et al. 2004), shifts in the timing of breeding for birds in Europe (Sanz 2003), and changes in the

abundance of zooplankton in the English Channel (Southward et al. 1995; other shifts reviewed in Walther et al. [2002]). Inouye et al. (2000) found differences in phenological shifts between high and low altitudes, indicating that shifts may not be uniform across the landscape.

Because butterflies are charismatic and easily observed, seasonal records have been kept in a number of countries, in some cases for many decades (Roy and Sparks 2000, Roy et al. 2001, Whitfield 2001, Konvicka et al. 2003, Stefanescu et al. 2003). These records have provided insight into the phenological properties of single species and of entire faunas. The best-known butterfly-monitoring project is the British Butterfly Monitoring Scheme (BMS), established in 1976, which includes more than 100 sites throughout the British Isles that are monitored weekly for 58 species (Pollard et al. 1995). Other studies of butterfly or moth fauna phenology include Emmel and Emmel (1962), Emmel (1963), Morisita (1967), Opler and Langston (1968), Forister and Shapiro (2003), and others reviewed in Shapiro et al. (2003).

We asked whether community-level trends in a butterfly fauna could be discerned by aggregating

models of individual butterfly species' phenology to reflect the composition of species at a site. Particularly, we wanted to identify individual species' contributions to the positive and negative rates of seasonal species turnover at a site. We also investigated whether aggregate predictive models were effective in detecting the differences in the timing of faunal phenology and changes in species richness between wet and dry years. The technique we applied to build the potential aggregate species richness (PASR) model was a quadratic logistic regression, using a butterfly monitoring data set spanning 27 years and containing only presence and absence records, to model individual species' seasonal flight windows.

The method yields improved estimates of first flight date and abundance from only presence and absence data. By combining phenological models from all of the species in the fauna, we created a fine-grained picture of species turnover in the site's fauna throughout a typical year, an interpretation made possible by the probabilistic modeling approach. Given that many historical species records only provide presence or absence information, this approach may have broad application in helping to build views of species' phenological patterns, particularly in areas with seasonal climates. This paper complements and extends the previous work by Shapiro et al. (2003), which examined between-year and between-site variation in species' phenology.

METHODS

Study location and monitoring methods

Monitoring for this project was done by a single individual (A. M. Shapiro; see Plate 1). It began in 1976 at Gates Canyon, California, USA, a small east-draining canyon in the coast range 70 km northeast of San Francisco (Appendix). Monitoring was conducted at two-week intervals by walking a 4-km transect from the base of the canyon up the road that climbs through it. All adult butterfly species observed (presences) were recorded. The transect samples several butterfly habitats: agriculture, mixed-oak woodlands, riparian, and chaparral (Shapiro et al. 2003). Species abundance records have been kept since 1999.

Development of the potential aggregate species richness model

Probability of observation for each species on each day of the year was assessed using a quadratic logistic regression. All species' probabilities were summed by day to develop a PASR model, which characterizes the potential species richness for each day of the year. A total of 10 720 species presences and 38 108 absences were entered to a database (Access 2004) as 1s and 0s, representing all species seen and not seen on 626 site visits from 1976 through 2002. Species' observation dates were represented as days numbered 1–365, beginning on 1 January, and all years were collapsed

into one. Records from 29 February were collapsed into 28 February (day 59).

We used the SAS (SAS 2000) system's Proc NLIN to calculate a quadratic logistic regression for all species, with species' presences modeled as a function of day and day squared. The model for the quadratic logistic regression is

$$P = \frac{\exp[a + b_1(\text{date}) + b_2(\text{date} \times \text{date})]}{1 + a + b_1(\text{date}) + b_2(\text{date} \times \text{date})}$$

where P is the probability of a presence, e is the base of the natural logarithm and a, b_1, and b_2 are the model parameters. Error is treated as binomially distributed. The curve of best fit for a species gives an estimate of the daily probability of a species' observation. For most species, the observation probability increases to some maximum, then decreases. Including a quadratic term (date × date) in the model allows the curve to take a humped form appropriate for fitting such unimodally distributed data.

Each species' daily observation probability was recorded in a matrix (Supplement). We summed the probabilities of all the species on each day of the year to arrive at the predicted aggregate species richness (PASR) model. We portray PASR as a year-long curve with the directly measured species richness values from each site visit.

Model validation

To test whether the PASR approach yields biologically credible results, we checked modeled flight windows in three ways. First, we tested the hypothesis that modeled species probability corresponds to actual species abundance by measuring the correlation between each species' maximum probability and its actual maximum abundance, observed since 1999. Second, we examined the timing of contributions to the overall species richness curve made by butterflies with various ecological characteristics. Finally, we compared daily estimates of species richness from the PASR model to those obtained using a smoothing spline (Venables and Ripley 2002). We conducted these analyses because the direct estimates of logistic regression accuracy and precision are difficult to interpret (Robert and Casella 2004).

To evaluate the relationship between each species' maximum predicted observation probability and observed abundance, we used abundance data (number of individuals per species per visit) that A. M. Shapiro recorded at the site during the last five years (141 visits, 2117 species/visit combinations for which numbers of individuals were recorded). For each species, we identified the maximum number of individuals ever observed, and used that as an indicator of its relative abundance. We extracted the maximum probability of observation of each species from the PASR model and calculated a nonparametric correlation (Kendall's tau) between these two variables.

To test whether known life history characteristics of different butterfly species were significantly correlated

JAMES H. THORNE ET AL.

TABLE 1. Life history characteristics of the 78 butterfly species recorded at Gates Canyon, California, USA.

Species	Brood number	Overwinter life stage	Weediness	Residence
Adelpha bredowii	multiple	larva	no	resident
Anthocharis sara sara	two	pupa	no	resident
Apodemia mormo	one	larva	no	nonresident
Atalopedes campestris	multiple	unknown	yes	resident
Atlides halesus	multiple	pupa	no	resident
Battus philenor	multiple	pupa	no	resident
Brephidium exilis	multiple	unknown	yes	nonresident
Callophrys augustinus iroides	two	pupa	no	resident
Callophrys dumetorum	two	pupa	no	resident
Callophrys johnsoni	two	pupa	no	nonresident
Celastrina echo	two	pupa	no	resident
Cercyonis pegala boopis	one	larva	no	resident
Cercyonis silvestris	one	larva	no	nonresident
Chlosyne leanira	one	larva	no	unknown
Chlosyne palla	one	larva	no	resident
Coenonympha tullia california	two	larva	no	resident
Colias eurytheme	multiple	larva	yes	resident
Danaus plexippus	multiple	immigrant	yes	migrant
Dione vanillae	multiple	unknown	yes	nonresident
Epargyreus clarus	one	larva	no	nonresident
Erynnis pacuvius	unknown	larva	no	unknown
Erynnis persius	two	larva	no	unknown
Erynnis propertius	two	larva	no	resident
Erynnis tristis	multiple	larva	no	resident
Euchloe ausonides	two	pupa	no	resident
Euchloe hyantis hyantis	one	pupa	no	nonresident
Euphydryas chalcedona	one	larva	no	resident
Everes comyntas	multiple	larva	yes	nonresident
Glaucopsyche lygdamus	one	pupa	no	resident
Heliopetes ericetorum	two	unknown	no	migrant
Hesperia columbia	two	unknown	no	resident
Hylephila phyleus	multiple	unknown	yes	nonresident
Icaricia icarioides	one	pupa	no	resident
Leptotes marina	multiple	unknown	yes	nonresident
Lerodea eufala	multiple	unknown	yes	nonresident
Limenitis lorquini	multiple	larva	no	resident
Lycaena arota arota	one	egg	no	resident
Lycaena gorgon	one	egg	no	resident
Lycaena helloides	multiple	egg	yes	unknown
Lycaena xanthoides	one	egg	no	unknown
Nymphalis antiopa	one	adult	no	resident
Nymphalis californica	one	adult	no	migrant
Nymphalis milberti	one	adult	no	unknown
Ochlodes agricola	one	unknown	no	resident
Ochlodes sylvanoides	two	larva	no	resident
Papilio eurymedon	one	pupa	no	resident
Papilio multicaudatus	two	pupa	no	resident
Papilio rutulus	two	pupa	no	resident
Papilio zelicaon	multiple	pupa	yes	resident
Pholisora catullus	multiple	larva	yes	nonresident
Phyciodes campestris	multiple	larva	no	nonresident
Phyciodes mylitta	multiple	larva	yes	resident
Pieris napi	two	pupa	no	resident
Pieris rapae	multiple	pupa	yes	resident
Plebejus acmon	multiple	larva	yes	resident
Poanes melane	multiple	unknown	no	resident
Polites sabuleti sabuleti	multiple	pupa	yes	nonresident
Polygonia satyrus	two	adult	no	resident
Polygonia zephyrus	two	adult	no	resident
Pontia protodice	multiple	pupa	yes	nonresident
Pontia sisymbrii	one	pupa	no	nonresident
Precis coenia	multiple	adult	yes	resident
Pyrgus communis	multiple	larva	yes	resident
Pyrgus scriptura	multiple	pupa	no	nonresident
Satyrium auretorum	one	egg	no	resident
Satyrium californica	one	egg	no	resident
Satyrium saepium	one	egg	no	resident
Satyrium sylvinus	one	egg	no	resident
Satyrium tetra	one	egg	no	resident
Speyeria callippe subspecies	one	larva	no	nonresident

TABLE 1. Continued.

Species	Brood number	Overwinter life stage	Weediness	Residence
Speyeria coronis	one	larva	no	resident
Strymon melinus	multiple	pupa	yes	resident
Thorybes pylades	one	larva	no	resident
Vanessa annabella	multiple	all	yes	resident
Vanessa atalanta	multiple	adult	no	resident
Vanessa cardui	multiple	immigrant	yes	migrant
Vanessa virginiensis	multiple	adult	no	resident
Zerene eurydice	two	adult	no	resident

Notes: These species were used to test the modeled flight windows by assessing whether potential aggregate species richness could be significantly decomposed into species representing various groups. Note that some of the characteristics recorded for a species at the site may be different from the characteristics the same species shows at other places in the United States.

with the timing of their predicted phenology, we extracted several descriptors from the probability curve of each species: the date of maximum probability; the value of maximum probability; and an estimate of the seasonal length of the flight period, defined as the length of time during which a species' observation probability was at least 20% of its yearly maximum. We also identified date of first and last observation, calling these "date of first flight" and "date of last flight."

These phenological properties were compared with a number of life history characteristics, defined by our observational experience and the literature (Table 1). Attributes included were: the number of broods per year (three categories), the overwintering life stage for each species (five categories), residency on site (three categories), and whether the species is weedy or non-weedy. "Weedy" species are defined as those associated with habitat disturbance, are highly dispersive, and have local population dynamics characterized by a high turnover rate. Resident species are those that breed on site every year and spend all year at the site; nonresidents are not there all year round, and may breed on the site but not in all years and have no persistent populations (this category includes strays); migrants are those species that breed on-site, but not in every generation. We did not include over-summer life stages because only two species have summer adult reproductive diapause, and these then reproduce late the same season. All species with summer diapause in any immature stage convert to winter diapause without a break and do not resume development until the following spring.

We used a one-way ANOVA (SAS Institute 2002) to determine whether these life history characteristics can explain variation in flight period characteristics of the butterfly species at Gates Canyon. Rarely observed species were not well characterized by this approach, so we excluded 19 species observed fewer than 20 times during the study (<3% of the visits) from the ANOVA models. Quadratic logistic results for these species are the most uncertain of the species contributing to PASR. We also removed species from each life history characteristic group whose classification was uncertain. For the analysis of overwintering category, we excluded nonresidents as listed in Table 1.

The previous tests examined how well individual species probability models reflected biological traits. To see how well the PASR model fit observed species richness per visit, we compared the PASR curve to a curve generated using a smoothing spline fit to the observed species richness data. Smoothing splines are an efficient nonparametric curve-fitting technique, useful when a function may not have a polynomial distribution (Venables and Ripley 2002). The spline approach generated a number for potential species richness for each day, while the PASR model represents the sum of each species' probability for each day.

Analysis of species turnover, effects of precipitation, and species richness trends

We conducted three analyses to examine community phenology using the PASR model: First, we decomposed PASR to derive the rate of species gain and loss for each day of the year, metrics not directly obtainable from the raw data. Second, we used mean annual precipitation to split the study into wet and dry years, and reran the analysis to examine the effect of wet years on phenology. Third, yearly residuals between the number of species observed per visit and the PASR model were used to examine temporal species richness trends.

Constant species richness values can mask rapid turnover in faunal composition. Therefore, we calculated the rate at which species were gained and lost from the observed fauna. Our measure was the daily change in probability of observation (positive or negative) for all species. Species whose probability of observation increased on a given day contributed to that day's positive turnover; species whose probability of observation decreased, contributed to the negative turnover for the same day. We used the PASR probability-by-day matrix to extract these measures. We also calculated the net turnover of species per day. These calculations permitted a detailed description of the faunal seasonal dynamics at the site.

To examine the effect of precipitation on butterfly phenology, we split the monitoring data set into two subsets, representing drier and wetter years ($N = 13, 14$, respectively). Because the site is in a mediterranean

climate, rainfall was measured by water year (defined here as 30 August–29 August), with each water year assigned to the calendar year in which it ended. The range of precipitation for our defined dry years was 21.9–47.6 cm/yr, with a mean ($\pm$SD) of 36.9 $\pm$ 7.9 cm/yr; for wet years the range was 55.0–118.5 cm/yr with a mean of 84.5 $\pm$ 20.4 cm/yr.

Precipitation data from a nearby Vacaville, California, weather station (Western Regional Climate Center 2004) were used for the classification. The dry years from the study were: 1976, 1977, 1979, 1981, 1985, 1987, 1988, 1989, 1990, 1991, 1992, 1994, and 2001. The wet years for the study were: 1978, 1980, 1982, 1983, 1984, 1986, 1993, 1995, 1996, 1997, 1998, 1999, 2000, and 2002. We ran independent quadratic logistic regressions for each species in each subset. We generated two PASR curves and compared the daily richness values to determine the effects of wet years on species phenology and richness.

To test for a temporal trend in species richness, we examined the residuals derived from the difference between actual and modeled species number for each site visit ($n = 626$). We used multiple regression of the residuals on date, date squared, and year in order to determine the relationship of the residuals to year. Date and date squared were included in the model to account for their potential confounding effects on the relationship between residuals and year.

Results

Seventy-eight species were observed at least once at Gates Canyon during the study period. Seventeen species (22% of the fauna) were observed fewer than 20 times.

Construction of the potential aggregate species richness model (PASR)

We ran the SAS quadratic logistic regression code, regressing the observations for each species against day and day squared (Fig. 1, Supplement). Of the 78 species, 13 had yearly maximum observation probabilities under 5% (0.9% for the lowest species, *Leptotes marina*), 24 had maximum observation probabilities between 5% and 50%, and 42 had maximum probabilities over 50% (Supplement). The range of dates of maximum abundance for most species spanned from day 34 to 294. Four species' maximum predicted occurrences fall in the first three months, 44 species' maxima are in the second quarter, 23 species in the third quarter, and two species, *Brephidium exilis*, and *Phyciodes campestris*, have maxima in October. In addition, the model identified the very early date of 1 January for four species, *Polygonia satyrus, Vanessa atalanta, Polygonia zephyrus,* and *Nymphalis antiopa,* and the very late date of 31 December for a single species, *Vanessa annabella.* These late abundance species are all adult hibernators. Flight period, defined as the number of days when the probability of a species' observation is 20% or greater

of its total probability, ranged from 26 days to 365 days, with a mean ($\pm$SD) of 241 $\pm$ 95 days.

We summed species' probabilities across all days to produce the PASR model, a single curve representing the number of expected species for each day of the year (Fig. 2, Supplement). The maximum modeled number of species was 30.526 on Day 144 (May 24). The modeled curve reflects the actual number of species observed per visit (represented as dots), but is lower at its maximum than observed values due to the low probability of some species that were actually observed.

Model validation

Species' maximum predicted probabilities of observation were highly correlated with their maximum observed abundances ($R = 0.8317$, $P < 0.0001$, $n = 69$). This confirmed that the maximum probability of observation could be used as a surrogate for abundance in later tests.

We summed the daily species probabilities by life history characteristics (Table 1) to determine daily contributions made by butterflies with different characteristics to the PASR model (Fig. 3a–d). We present the statistics for each set of one-way ANOVAs (Table 2), conducted to assess influence of life history characteristic on potential species richness by date. The 17 rarely seen species (seen <3% of the time) excluded from the analysis were: *Apodemia mormo, Callophrys johnsoni, Cercyonis pegala boopis, Cercyonis silvestris, Chlosyne leanira, Dione vanillae, Epargyreus clarus, Erynnis pacuvius, Euchloe hyantis hyantis, Heliopetes ericetorum, Hesperia columbia, Leptotes marina, Lycaena arota arota, Lycaena xanthoides, Nymphalis milberti, Phyciodes campestris, Pyrgus scriptura, Speyeria callippe* subspecies, and *Speyeria coronis.*

Among overwintering life stage categories, adult hibernators had significantly earlier maximum probability of observation than any other group. However, that maximum was significantly lower than that reached by larval overwinterers. Egg stage overwintering species had significantly shorter flight windows than those of larval or adult stage overwinterers. The negative date of first flight projected by the model for adult overwintering species indicates a significantly earlier date than species with other overwintering life stages. Finally, species that overwinter as pupae have earlier first flight observations than eggs.

Multiple-brooded species had significantly later dates of maximum probability of observation, longer flight periods, and later dates of last flight than univoltine or bivoltine species. Nonresident species dates' of maximum probability of observation were significantly different from resident species, and they had significantly later dates of first flight. There was no difference in length of flight window between nonresidents and residents.

Weedy species did not differ in average maximum abundance, but did differ in all categories relating to
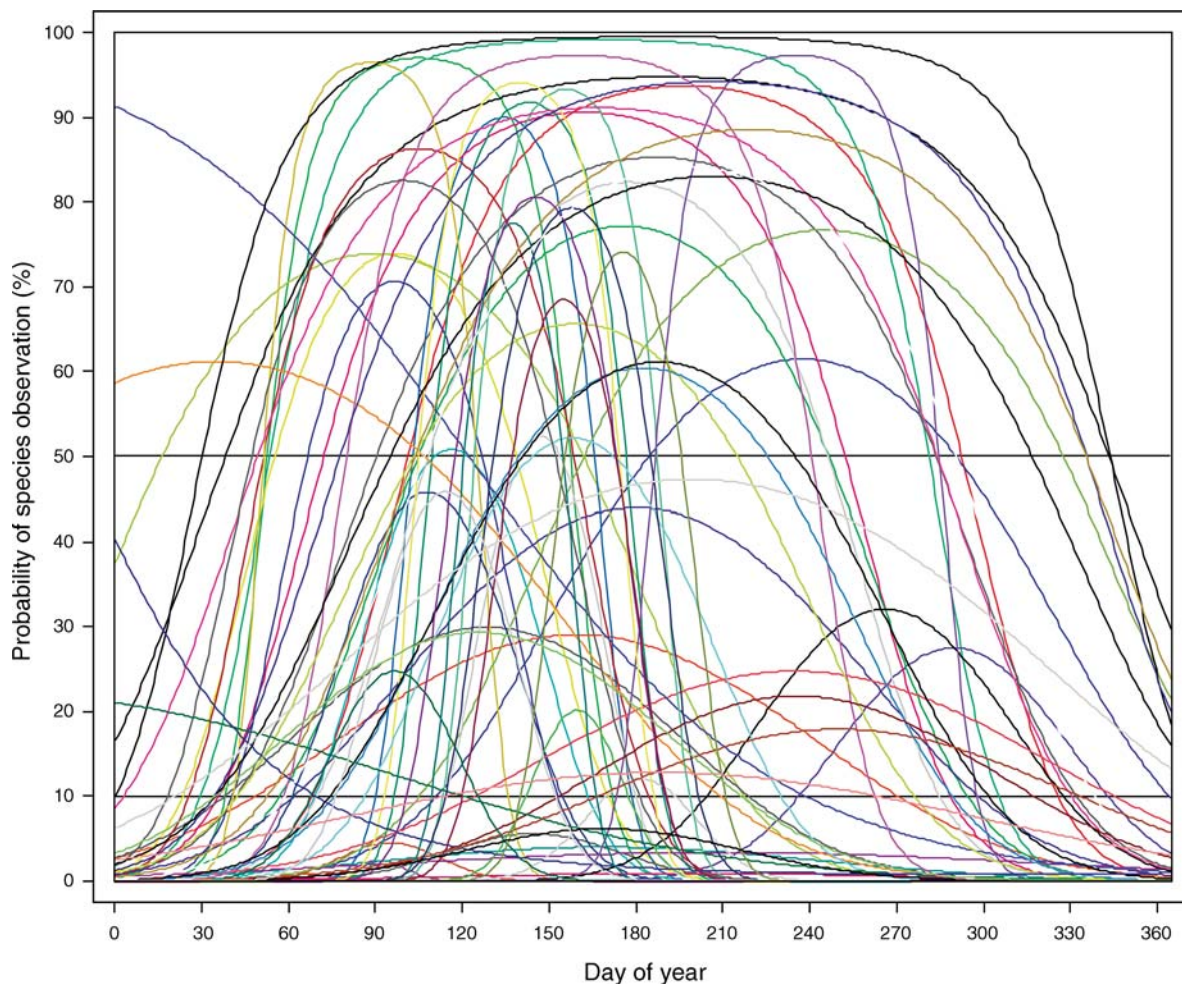
FIG. 1. These curves represent the probabilities of observation for butterfly species found at Gates Canyon, California, USA, on any given day. Butterflies at the site for which the quadratic logistic regression used was not able to generate a curve are not included. A large number of species have low maximum probability of observations. The variation in curve width represents the length of modeled flight windows of different species. Reference lines are provided at 10% and 50%.

flight phenology. Weedy species' maximum observation probabilities were significantly later than non-weedy species, length of flight window was significantly longer, and first and last flight dates (observations) were significantly later.

The comparison of the PASR curve with a spline-fitted curve (Fig. 2) showed that both methods identified a similar maximum level of potential species richness at the same time of year. The spline is better able to fit a second hump in the data, representing the contribution to species richness by multivoltine butterflies. However, the curve generated using the spline technique is unable to identify the proportional contribution of each species to the overall potential species richness, whereas PASR summarizes the dynamics of all the species used to compose it.

The quadratic logistic species modeling and PASR approach accurately represented many of the dynamics of individual species' phenology faunal dynamics and

abundance for the Gates Canyon butterfly fauna. Since these biological checks were all positive, we felt the PASR model could be used for other, exploratory, analyses.

Analysis of species richness trends, effects of precipitation, and species turnover

We calculated turnover, the rate of species accumulation and loss at the site (Fig. 4). From the beginning of the year, new species entered the fauna at a rate that accelerated up to 0.32 species/day on 29 April (day 119). Species loss from the site began on 5 April (day 95) and accelerated around 10 April (day 100) (a few weeks before the maximum species gain). Species loss reached −0.30 species/day loss on 25 June (day 176), with remaining species being more stable until a second, smaller wave of loss peaked on 13 October (day 286) at −0.17 species/day. Less than 5% of the fauna have yet to emerge by 6 July (day 187), while 42% of the species
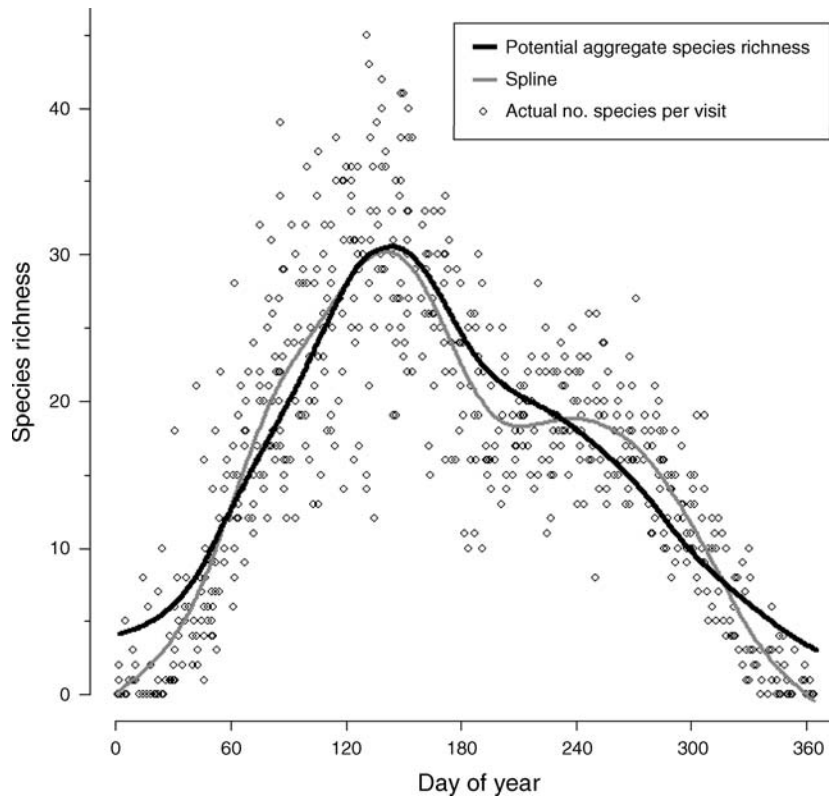
FIG. 2. The black-line curve represents the summed probabilities of 78 species of butterfly found at Gates Canyon for every day of the year. The circles refer to the actual number of species seen on at each visit over 27 years. The second curve (gray line) was fitted by a spline, an efficient curve-fitting technique, but one that loses information about the individual contribution of each species to daily predicted species richness.

have already completed their life cycle for the year. Total turnover of species was marked by four distinct peaks, the first and last of which represent almost exclusively gain or loss; the middle two, 2 May and 27 June (days 122 and 178) represent the combination of the two.

The PASR models for wet and dry years were markedly different from each other (Fig. 5). Peak richness occurred on 23 May (day 133) for dry years. In wet years, peak PASR was 17 days later, on 30 May (day 150). Peak PASR in wet years was 30.00; in dry years it was 32.36. The wet-years curve lags the dry-years curve at the onset of the season, and was on average 3.37 species lower before day 150. The timing in decline of species richness was nearly the same for dry and wet years.

Using a least squares fit of residuals against year, date, date squared, and their second-order interactions, our analysis shows an average decline of daily species richness of 0.12 species per year. This rate of decline is equivalent to a loss, over the course of the study, of 3.24 species on any given day. There was also a significant positive interaction between year and date (t ratio = 2.56, $P = 0.01$), indicating a shift in yearly phenology to

later in the year. Note that the majority of the wet years in the study were in its second half.

DISCUSSION

Both the individual species models and PASR revealed important details about butterfly dynamics at Gates Canyon. Individual species models permitted an estimate of when the best times might be for observing them, and confirmed that many species are rare in this system. The whole-community model showed phenological sensitivity to rainfall, permitted quantitative estimates of species richness by day, and permitted detection of the loss of 3.24 species from the species pool over the course of the study. The length of our data set permitted some rare species to be modeled, but this approach could be used for more common species with much less data, perhaps four or five years worth. The approach is well suited to describe butterfly phenological patterns in seasonal systems.

Model results confirmed early conclusions about biological observations (Preston 1948); specifically, that many species are rare, with an associated very low probability of observation. About 22% of the species at Gates Canyon were seen <20 times in 626 visits (<3.2% of the time). Rare species have important ramifications
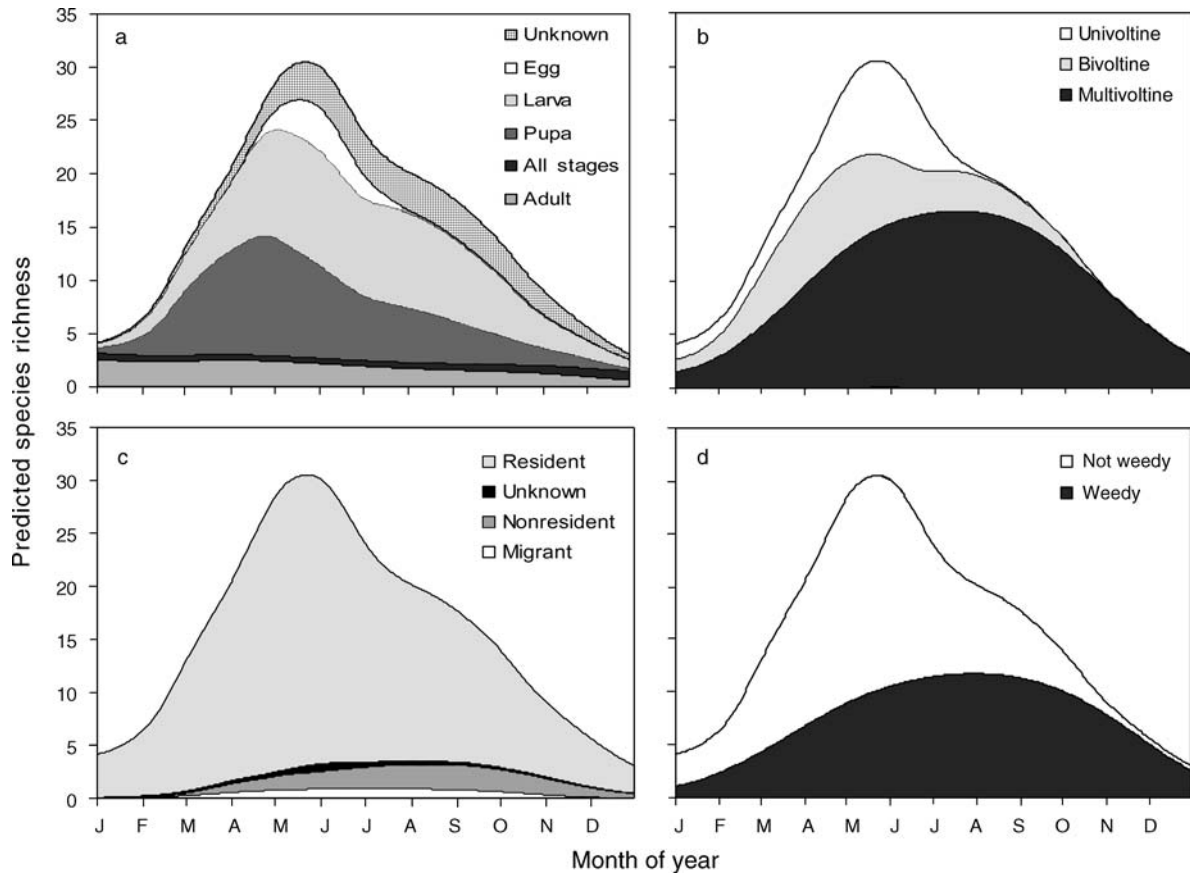
Fig. 3. These curves represent the relative contribution to the potential aggregate species richness model made by the life history characteristics of each species of butterfly. Species for which certain characteristics were unknown were excluded from the graphics.

for projects trying to monitor species decline, as it could be misleading to use the same criteria to measure decline for both common and rare species. Criteria on how to interpret rare-species trends need to be developed. For

this study, commonness or rarity affected modeled flight windows. Rare species were sometimes fit by curves with relatively sharp peaks, implying short flight windows, when the opposite was sometimes true. Examples of this

TABLE 2. Results from ANOVA tests of life history characteristics on various modeled components of butterfly seasonality.

Variable	Mean date of maximum abundance	Length of flight window (d)	Date of first observation
Overwintering stage	($F_{3,46} = 6.93$****)	($F_{3,46} = 4.05$**)	($F_{3,46} = 20.07$****)
Adult ($n = 8$)	70.7[b]	241.0[a]	−49.9[a]
Pupa ($n = 19$)	143.7[a]	176.2[a,b]	55.9[b]
Larva ($n = 16$)	171[a]	208.1[a]	67.1[b,c]
Egg ($n = 7$)	166[a]	97.7[b]	117.4[c]
Voltinism	($F_{2,56} = 17.12$****)	($F_{2,56} = 54.63$****)	NS
Multiple-brooded ($n = 28$)	203.4[a]	267.5[a]	69.8
Bivoltine ($n = 15$)	113.0[b]	157.0[b]	34.7
Univoltine ($n = 16$)	121.9[b]	92.4[c]	75.9
Residence	($F_{2,53} = 7.86$**)	NS	($F_{2,53} = 7.25$**)
Nonresident ($n = 7$)	219	190	124.1
Resident ($n = 48$)	148	185	55.6
Weediness	($F_{1,57} = 40.53$****)	($F_{1,57} = 29.43$****)	($F_{1,57} = 4.05$*)
Non-weedy ($n = 39$)	126.1	153.1	49.8
Weedy ($n = 20$)	221.1	267.6	87.4

Notes: We extracted mean date of maximum abundance (numeric date), length of flight window (in days), and date of first observation (numeric date). Superscript letters identify which categories were significantly different from each other within each analysis (different letters) and which were not significantly different (the same letter). Values in parentheses are ANOVA results for comparisons within each life history characteristic. Levels of significance are: NS, nonsignificant; * $P < 0.05$; ** $P < 0.01$; **** $P < 0.0001$.
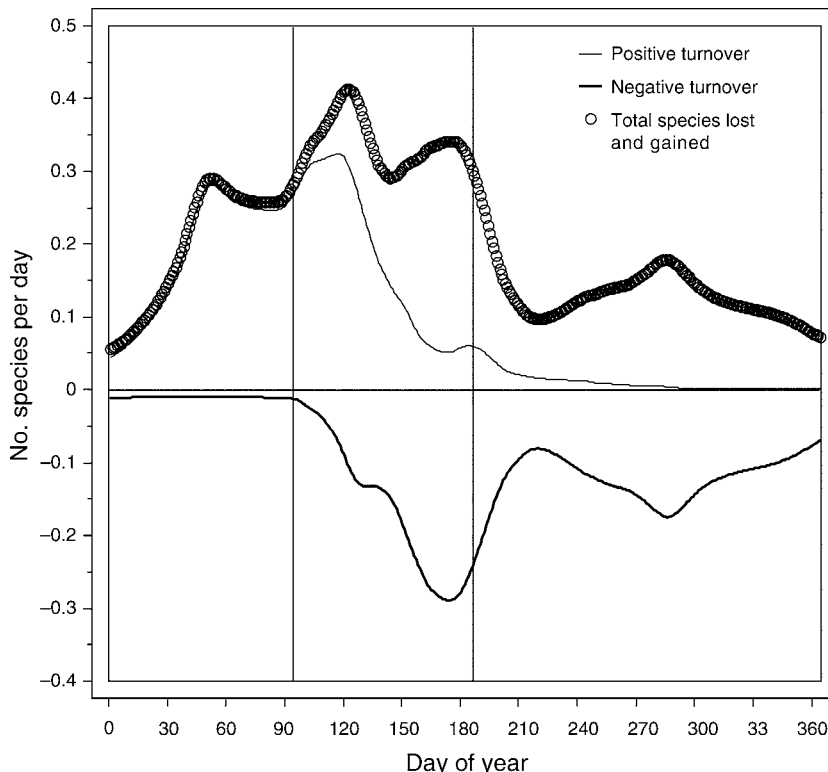
Fig. 4. The addition, loss, and overall rate of transition of species at Gates Canyon across a single model year. The two vertical lines represent day 95, when negative species turnover begins, and day 187, at which point <5% of all modeled species for the year have yet to emerge, and 42% of the modeled species have already finished their yearly life cycle. The greatest day of species loss is day 175.

potential misinterpretation include the rare species *Lycaena arota arota,* which is single brooded, and *Hesperia columbia,* which is double brooded. Modeled flight windows were narrow for the two species, but *H.*
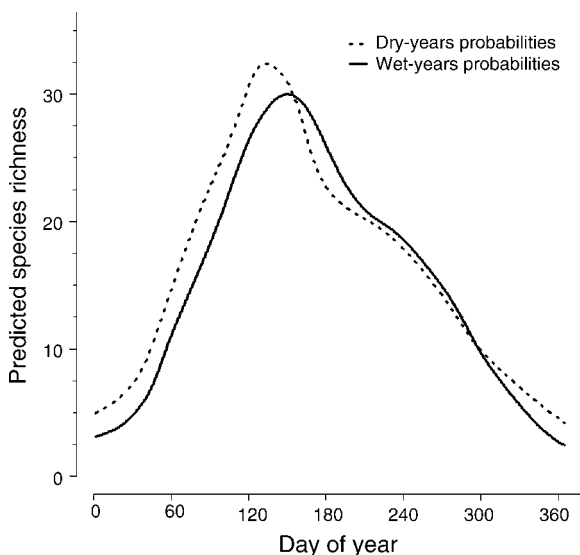

Fig. 5. Potential aggregate species richness for wet and dry years. Wet years delay the phenological cycle by as much as 17 days and reduce the maximum predicted species by 2.

columbia flies for much longer time than the other. Both were excluded from the ANOVA models due to rarity.

Wet years were shown to delay the maximum PASR date by, on average, 17 days, when there were also two less potential species than in dry years. The majority of the study's wet years was in the last decade, and potentially help explain the measured decline in species. This explanation potentially confounds other hypothesized effects that include habitat fragmentation and increasing temperatures. The primary effect of wet years was to delay the onset of species' life cycles, which effectively narrows the length of time available to some early-season species before the vegetation dries up. This likely affects univoltine species that have a short season to begin with. Forister and Shapiro (2003) similarly found that the average date of first spring flight for butterflies in nearby Central Valley, California, locations was significantly delayed in wetter years.

Delay of species life cycles is potentially an important problem; much of the California foothills butterfly fauna flies only in the spring, when there is good soil moisture (hence young, tender plant growth) and sufficient sunshine. Since both wet and dry years dry out at the same time (Fig. 5), late springs effectively compress the time available for early-season species. An example (not however, found at Gates Canyon) is

PLATE 1. (Left) The Woodland kipper (*Ochlodes sylvanoides*) is unique among resident species at Gates Canyon in apparently having two successive broods beginning in midsummer and extending into late autumn. Photo credit: Erik Runquist. (Right) Art Shapiro on his ~624th visit to Gates Canyon. Photo credit: J. M. O'Brien.

Euphydryas editha, which needs to reach its third instar (diapausing stage) before its ephemeral spring-annual host senesces, which pushes it towards very early emergence (Hellmann et al. 2004). In these foothills, most (non-weedy) species seen in spring are uni- or bivoltine, while the weedy species are multivoltine and increase in probability of observation as the season progresses. Their date of maximum probability of observation is later in the season. Often, weedy species come out quite early in highly disturbed areas in California, but at Gates Canyon, they colonize the area as the season progresses. Nonresident species have a later maximum abundance because they are moving in from other sites.

Shapiro et al. (2003) noted a set of early-spring species, a set of late-spring species, and a set of spring-bivoltine species that overlaps the other groups in this fauna. These phenomena are evident in the PASR model. The sharp drop in species richness at midyear corresponds to the disappearance of California Buckeye (*Aesculus californica*) as a nectar source. Shapiro et al. (2003) suggested that the flight windows of several species have been selected to correspond to the availability of this resource. They also noted that very few species emerge in the last one-third of the season, not only at Gates Canyon, but in all parts of temperate

North America examined so far. That observation is validated here, but a general explanation of the phenomenon, which is not limited to mediterranean climates, remains elusive. Note that at Gates Canyon, adult nectar sources (mostly coyote bush, *Baccharis pilularis*, a native chaparral shrub) are abundant in autumn and are very heavily exploited by the autumn fauna, consisting almost entirely of weedy species. Other resources available late in the flight season are riparian trees, which are used as larval hosts by a set of non-weedy multivoltines (usually two to three brooded). These species (*Limenitis lorquini, Adelpha bredowii, Papilio rutulus, P. multicaudatus*) contribute to the "late-season" portion of the PASR curve.

The PASR model permitted a detailed description of the average daily species turnover. Species loss at the site did not begin until day 95 (5 April), indicating that a single day's monitoring effort prior to this date could potentially capture early-season species richness. By day 156 (5 May) there remained only 10% of the total potential species to be added, and 20% of the season's species had disappeared. By day 187 (6 July) only 5% of the species pool remained to be added, while 42% were done for the year, and <1% of the species remained to be added on day 253, while 63% were finished for the season. This approach provides a constructive way to

think about sampling dates. Combined with the knowledge of whether a year is wet or dry, sampling windows that maximize the chance of observing that year's species richness can be developed.

The PASR curve fit the actual species richness data as well as spline fitting in the early season, and identified a similar maximum predicted abundance. The spline curve better described dynamics in the mid-portion of the year when weedy lower elevation species enter the site, and both approaches were similar in the last one-third of the year. However, the spline approach, fitted to the number of species observed each visit, did not yield information about the relative daily contributions to site richness of each species that the PASR approach provided.

At the beginning and end of the year, PASR indicated between three and four potential species, higher than the spline curve and most observed values (between zero and two). This is due to the way logistic regression weights deviate from the predicted curve; improved fit in the middle of the curve is selected, resulting in slight misfit at the ends. Likewise, expected richness near the height of the season underfit the actual average number of species observed. In fitting a single curve to data combined from many years, the curve is bound to be broader and flatter than any of the single years, resulting in such underfitting. The occasional appearance of species that contribute little to the overall probability of observation, but contribute an entire individual on days they are actually seen, contributes to curve underfitting. These are the rare species that could be treated differently when calculating loss in species richness.

A few species that overwinter as adults had negative probabilities of observation at the beginning of the year. Hibernators generally live from August to April, and many of them can emerge at almost any time over that time, although there are exceptions: *Nymphalis antiopa* doesn't come out between November and 25 January. Most of these species, *Polygonia zephyrus* excepted, fly in summer, too (we do not understand where this species is during the warm season; only hibernators were observed at Gates Canyon). This group, therefore, has a long flight window, which is not well represented using a yearly cycle that starts on January first. The nonsensical negative value for date of first observation is a reflection of the model's inability to portray this dynamic.

Species flight window modeling depends on the pooling of data from many years, and treats all data points as if they were drawn from a single distribution. In fact, phenology is not constant from year to year, and one effect of the method is, thus, to make the flight window appear broader than it may actually be for any single year. By including a quadratic term in the PASR model approach, the probability of presence through the year has a rising and a falling limb that represents butterfly dynamics for many species in a seasonally variable climate. Our approach was useful for estimating daily potential species richness, which served as the basis

for positive and negative species turnover models, which cannot be directly measured from the observation data. Calculated observation probabilities for univoltines reflected observed patterns in the field. Those for bivoltine and multivoltine species were not as representative, as the modeled flight window smoothed the two or more peak abundances into a single curve.

The PASR approach permitted a way to think about the timing of field monitoring for a suite of species highly sensitive to local weather conditions. The date of maximum probability from the PASR model can indicate dates to optimize the chance to record faunal diversity, which is not directly measurable from the original presence/absence data. Onset of faunal phenology is offset by over two weeks between wet and dry years, making the window from 14 May (dry year) and 2 June (wet year) the best potential time for maximum species richness for a single site visit at Gates Canyon. Since species attrition did not start until 5 April, potentially one additional sampling date could be used to capture early species composition. The model also confirms that visits after the middle of July are the least likely to add species to a monitoring record. This approach can potentially be applied to other monitoring projects in areas with seasonal changes. Common species can be modeled with four-to-five-years worth of monitoring data.

LITERATURE CITED

Access. 2004. Access relational database. Microsoft, Seattle, Washington, USA.

Cayan, D. R., S. A. Kammerdiener, M. D. Dettinger, J. M. Caprio, and D. H. Peterson. 2001. Changes in the onset of spring in the western United States. Bulletin of the American Meteorological Society **82**:399–415.

Chiune, I., P. Yiou, N. Viovy, B. Seguin, V. Daux, and E. Le Roy Ladurie. 2004. Summer temperature variations are reconstructed from harvest dates since 1370. Nature **423**(7015):289–290.

Diamond, J. M. 1983. Laboratory, field and natural experiments. Nature **304**(5927):586–587.

Emmel, T. C. 1963. Ecological studies of Rhopalocera at Donner Pass, California. II. Meteorological influences on flight activity. Journal of the Lepidopterists' Society **17**:7–20.

Emmel, T. C., and J. F. Emmel. 1962. Ecological studies of Rhopalocera at Donner Pass, California. I. Butterfly associations and distributional factors. Journal of the Lepidopterists' Society **16**:23–44.

Forister, M. L., and A. M. Shapiro. 2003. Climatic trends and advancing spring flight of butterflies in lowland California. Global Change Biology **9**:1130–1135.

Gurevitch, J., S. M. Scheiner, and G. A. Fox. 2002. The ecology of plants. Sinauer Associates, Sunderland, Massachusetts, USA.

Hellmann, J. J., S. B. Weiss, J. F. McLaughlin, P. R. Ehrlich, D. D. Murphy, and A. E. Launer. 2004. Structure and dynamics of *Euphydryas editha* populations. Chapter 3. Pages

34–62 *in* P. R. Ehrlich and I. Hanski, editors. On the wings of checkerspots: a model system for population biology. Oxford University Press, Oxford, UK.

Inouye, D. W., B. Barr, K. B. Armitage, and B. D. Inouye. 2000. Climate change is affecting altitudinal migrants and hibernating species. Proceedings of the National Academy of Science **97**:1630–1633.

Konvicka, M., M. Maradova, J. Benes, Z. Fric, and P. Kepka. 2003. Uphill shifts in distribution of butterflies in the Czech Republic: effects of changing climate detected on a regional scale. Global Ecology and Biogeography **12**:1195–1202.

Morisita, M. 1967. The seasonal distribution of butterflies in the suburbs of Kyoto. Pages 95–132 *in* M. Morisita and T. Kira, editors. Natural history—ecological studies. Chuo Koron, Tokyo, Japan.

Opler, P., and R. L. Langston. 1968. A distributional analysis of the butterflies of Contra Costa County, California. Journal of the Lepidopterists' Society **21**:89–107.

Pollard, E., D. Moss, and T. J. Yates. 1995. Population trends of common British butterflies at monitored sites. Journal of Applied Ecology **32**:9–16.

Preston, F. W. 1948. The commonness, and rarity, of species. Ecology **29**:254–283.

Robert, C. P., and G. Casella. 2004. Monte Carlo statistical methods. Springer, Berlin, Germany.

Roy, D. B., P. Rothery, D. Moss, E. Pollard, and J. A. Thomas. 2001. Butterfly numbers and weather: predicting historical trends in abundance and the future effects of climate change. Journal of Animal Ecology **70**:201–217.

Roy, D. B., and T. H. Sparks. 2000. Phenology of British butterflies and climate change. Global Change Biology **6**: 407–416.

Sanz, J. J. 2003. Large-scale effect of climate change on breeding parameters of pied flychatchers in Western Europe. Ecography **26**:45–50.

SAS Institute. 2000. SAS. Version 8. SAS Institute, Cary, North Carolina, USA.

SAS Institute. 2002. JMP. Version 5. SAS Institute, Cary, North Carolina, USA.

Shapiro, A. M., R. VanBuskirk, G. Kareofelas, and W. D. Patterson. 2003. Phenofaunistics: seasonality as a property of butterfly faunas. Pages 111–147 *in* C. L. Boggs, W. B. Watt, and P. R. Ehrlich, editors. Butterflies: ecology and evolution taking flight. University of Chicago Press, Chicago, Illinois, USA.

Southward, A. J., S. J. Hawkins, and M. T. Burrows. 1995. Seventy years' observations of changes in distribution and abundance of zooplankton and intertidal organisms in the western English Channel in relation to rising sea temperature. Journal of Thermal Biology **20**:127–155.

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

Venables, W. N., and B. B. Ripley. 2002. Modern applied statistics with S. Springer-Verlag, New York, New York, USA.

Walther, G., E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. Beebee, J. Fromentin, O. Hoegh-Guldberg, and F. Bairlein. 2002. Ecological responses to recent climate change. Nature **416**:389–395.

Western Regional Climate Center. 2004. Historical climate information, Station number 9200. Western Regional Climate Center, Reno, Nevada, USA. ⟨http://www.wrcc.dri.edu⟩

Whitfield, J. 2001. The budding amateurs. Nature **414**:578–579.

Wolfe, D. W., M. D. Schwartz, A. N. Lakso, Y. Otsuki, R. M. Pool, and N. J. Shaulis. 2005. Climate change and shifts in spring phenology of three horticultural woody perennials in northeastern USA. International Journal of Biometeorology **49**:303–309.

APPENDIX

Habitat types at Gates Canyon, in the California Coast Ranges, USA (*Ecological Archives* A016-061-A1).

SUPPLEMENT

Daily observation probabilities for 78 butterfly species at Gates Canyon (*Ecological Archives* A016-061-S1).