 DOI: 10.1086/667893

Appendix A from A. B. Phillimore et al., "Dissecting the Contributions of Plasticity and Local Adaptation to the Phenology of a Butterfly and Its Host Plants"

(Am. Nat., vol. 180, no. 5, p. 000)

The Use of First Dates in Phenology Research

Observations of first dates by natural historians go back to at least the early eighteenth century (Marsham 1789) and continue to be the mode of data collection for several large-scale phenological monitoring schemes, including the UK Phenology Network (http://www.naturescalender.org.uk). The chief advantage of first dates over other methods of data collection, such as mean dates, is the ease and accuracy of recording. However, a recent simulation study comparing different metrics of phenology found that first dates yielded biased estimates of the change in phenology between two years under a variety of scenarios, while mean dates gave unbiased estimates (Moussus et al. 2010). As most phenological research involves the estimation of slopes, with year, latitude, or temperature as a predictor, a more pertinent test is whether first dates yield unbiased slope estimates.

Methods

We sampled temperature and population mean phenology from a bivariate normal distribution with mean $= 10$, temperature variance $= 1$, phenology variance $= 72$, and temperature-phenology covariance $= -6$, meaning that the expected slope of phenology on temperature $= -6$. In the context of these simulations, population size (N) corresponds to the number of individuals sampled in a particular year or a particular locality. The distribution of phenological observations in a population was simulated by drawing N random normal deviates from a distribution with mean $= 0$ and specified variance. We specified the N of each population in three ways: (1) constant N across populations, either 10 or 30; (2) N drawn from a Poisson distribution with a mean of 10 or 30; and (3) N drawn from a Poisson distribution with the mean a multiple of temperature. The number of populations considered was 10 or 30.

For each population we obtained the mean and minimum phenological observation and then used a linear model to estimate the slope of temperature on population mean or first phenology. We conducted 10,000 simulations for each set of parameters and calculated the mean and 95% confidence interval using the resulting distribution of slopes. A departure of the estimated mean slope from the expected slope would indicate bias, and the breadth of the 95% confidence interval quantifies precision. All simulations were conducted using R (R Development Core Team 2011).

Results

Slope estimates using first dates and mean dates are unbiased when population sizes are equal or when variation in population size is random with respect to the predictor variable (tables A1, pt. A, B). However, slopes are estimated with less precision when first dates are used. Precision improves as the number of populations increases, the sample size per population increases, and the ratio of within- to between-population variance in phenology decreases. Slopes estimated using first dates show bias when population sample size is itself correlated with the predictor variable (table A1, pt. C), with slopes estimated using mean dates unaffected. The bias is reduced as the number of populations sampled increases, the mean sample size per population increases, and the ratio of within- to between-population variance in phenology decreases.

Sampling and Its Effects on Slope Estimates in This Study

Although the number of observations used in this study is highly heterogeneous across time and space (figs. B1–B3), the sample size that is most pertinent in terms of biasing slope estimates is the number of individuals that an individual recorder encounters in a single year. This number will in turn be influenced by observer effort and local species abundance. Therefore, if a correlation exists between temperature and observer effort or species abundance, over either time or space, this could bias our slope estimates. While we have no way of testing whether this exists, we note that all three species are abundant across the United Kingdom and we have no evidence to suggest that observer effort varies systematically with spring temperatures. Furthermore, for our test of whether slopes differ over space versus over time to be biased, the relationship between temperature and sample size would need to differ over space versus over time. For a more extensive consideration of the biases associated with first dates, see Clark and Thompson (2011).

Table A1. Estimates of mean slope and uncertainty in slope using mean dates and first dates, where population sample sizes are constant (A), population sample sizes come from a Poisson distribution (B), or population sizes come from a Poisson distribution with the mean equal to a specified scalar of temperature (C)

No. populations	Mean population sample size	σ^2_W/σ^2_B	Mean slope estimate using mean date (95% CI)	Mean slope estimate using first dates (95% CI)
A:				
10	10	1	−5.98 (−11.12 to −1.00)	−5.98 (−12.02 to −.20)
30	10	1	−6.00 (−8.47 to −3.54)	−6.01 (−8.93 to −3.07)
30	30	1	−6.02 (−8.34 to −3.65)	−6.03 (−8.74 to −3.25)
30	30	.1	−6.01 (−8.31 to −3.73)	−6.01 (−8.35 to −3.68)
B:				
10	10	1	−6.01 (−11.07 to −1.01)	−5.99 (−12.12 to .11)
30	10	1	−5.99 (−8.56 to −3.47)	−6.00 (−9.09 to −3.03)
30	30	1	−6.00 (−8.35 to −3.64)	−6.00 (−8.75 to −3.25)
30	30	.1	−6.01 (−8.29 to −3.72)	−6.01 (−8.37 to −3.67)
C:				
10	10	1	−5.98 (−10.94 to −.97)	−6.45 (−12.49 to −.37)
30	10	1	−6.01 (−8.51 to −3.48)	−6.47 (−9.44 to −3.47)
30	30	1	−5.99 (−8.41 to −3.60)	−6.36 (−9.21 to −3.54)
30	30	.1	−5.99 (−8.33 to −3.68)	−6.10 (−8.49 to −3.73)

Note: σ^2_W/σ^2_B = the ratio of expected within-population variance to between-population variance in phenology. CI = confidence interval.

Literature Cited Only in Appendix A

Marsham, R. A. 1789. Indications of spring. Philosophical Transactions of the Royal Society B: Biological Sciences 79: 154–156.