TECHNICAL COMMENT

Method matters: pitfalls in analysing phenology from occurrence records

Elise A. Larsen* (iD) and
Vaughn Shirey (iD)

Abstract

Large occurrence datasets provide a sizable resource for ecological analyses, but have substantial limitations. Phenological analyses in Fric *et al*. (2020) were misleading due to inadequate curation and improper statistics. Reanalysing 22 univoltine species with sufficient data for independent analysis, we found substantively different macroscale phenological patterns, including later onset at higher latitude for most species.

INTRODUCTION

The aggregation of large datasets from museum records and community science provides a valuable resource for macroscale ecological analyses. However, such data include spatiotemporal and taxonomic biases that must be addressed (Troudet *et al*. 2017; Ries *et al*. 2019). Given these biases, proper data curation and appropriate modelling strategies are necessary to ensure valid inferences.

Fric *et al*. 2020 used occurrence data for 100 species aggregated in the Global Biodiversity Information Facility (GBIF) in temperate regions of North America and Europe to track phenology across latitudes. Estimating phenology metrics and trends from large occurrence datasets is possible, but requires sufficient data density and appropriate statistical methods (Taylor & Guralnick 2019). The data from this study were frequently too sparse and insufficiently curated to estimate phenological patterns across latitudes. Furthermore, using regression of residuals resulted in spurious patterns; after correcting for altitude and year, onset and termination phenology appeared the same at low and high latitudes for most species, contrary to previous findings (Karlsson 2014; Matechou *et al*. 2014). We show that by applying appropriate data curation and methods, most species demonstrated later onset and shorter flight periods at higher latitudes.

Many species analysed (in 105 datasets separated by continent) in Fric *et al*. (2020) had insufficient data for independent analysis. Data were analysed with as few as 15 occurrence records across > 20° latitude and >100 years. Phenological 'onset' and 'termination' of flight periods were extracted simply as the first and last day-of-year (DOY) values within latitudinal bands, pooled across all years and altitudes. Pooling data increased spatiotemporal bias, lowered the resulting power to detect patterns and resulted in only one observation date being used as both 'onset' and 'termination' of flight periods (resulting in one-day flight periods and 'peak flight') in an average of 20% of latitudinal bands per species (Figure 1).

Fric *et al*.'s data curation was inadequate regarding spatial precision and outlier detection. Altitudes were extracted using imprecise GIS coordinates, sometimes representing sea floor or vague place names (e.g. 'Mt Shasta') and skewed left, giving high altitude observations outsized leverage in regressions. Temporal outliers were problematic; one species' onset at 68° N was in January, when the next occurrence across all latitudes was in June. No sources were cited for species traits, and we found evidence documenting additional generations in portions of their range for 22 species identified as obligate univoltine (Table S1).

Finally, the analytical approach in Fric *et al*. (2020) produced biased results. Beyond regressing individual species' phenometrics against latitude, altitude and year separately, regression of residuals was used for corrected regressions. This resulted in biased parameter estimates due to collinearity among explanatory variables and reduced statistical significance (Freckleton 2002). Results suggested most species' onset (67 datasets) and termination (71 datasets) were similar across latitudes (Figure 2). These results were surprising, considering well-documented delayed and/or shortened flight periods at high latitudes (Karlsson 2014).

We sought to validate those results with a more robust analysis, applying stricter data standards and curation. For 72 species (76 datasets) we confirmed as univoltine, we filtered data for altitude (0–500 m) and timing (March–November). We calculated phenometrics for year-latitude combinations with at least 10 observations. Only 22 datasets, all European, met these requirements in at least three latitudinal bands (Table S1). For these, onset and termination were estimated from a Weibull distribution using R package phest (Pearse *et al*. 2017) and bounded by days (60,330). To estimate

Department of Biology, Georgetown University, Washington, DC, USA
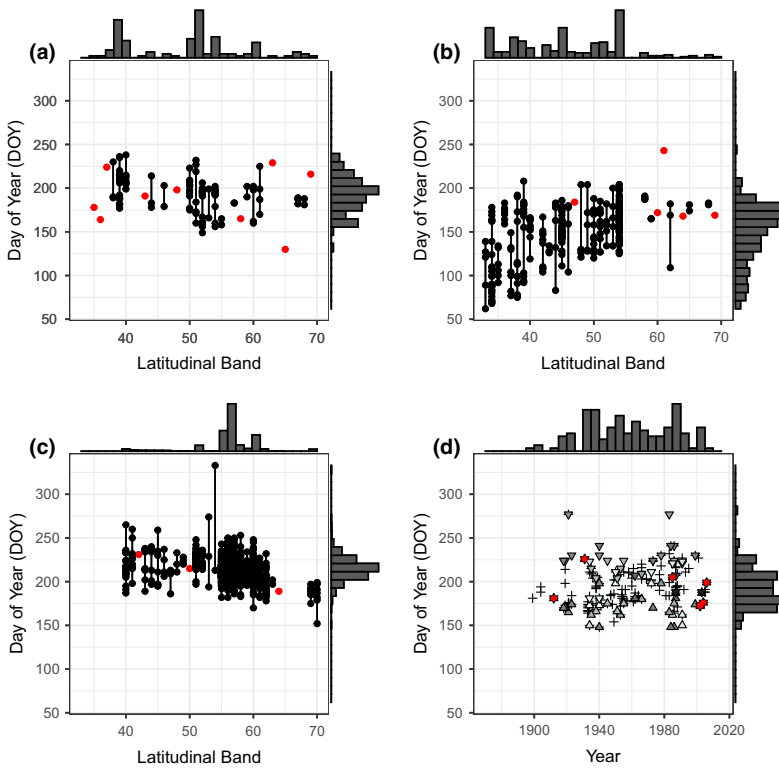***Correspondence**
E-mail: eal109@georgetown.edu

Figure 1 Raw occurrence data used in the Fric *et al.* (2020) analysis and flight period lengths by latitude for 3 datasets (a: *Agriades glandon* in N. America, b: *Glaucopsyche lygdamus* in N. America, c: *Hesperia comma* in Europe) and year for 1 dataset (d: *Parnassius smintheus* in N. America). Results for these species were presented in Fric *et al.* (2020) Figure 1 to demonstrate varied phenological patterns. Red points represent observations used as both onset and termination. Marginal histograms show data density across latitudes (above the plot) and time of year (to the right of the plot). Because onset and termination analyses used only one day of year (DOY, calculated as SuccDay) per latitudinal band, panel d identifies onset (▲) and termination (▼) observations (fill coloured by latitude where darker colours represent higher latitudes) in addition to all observations (+).
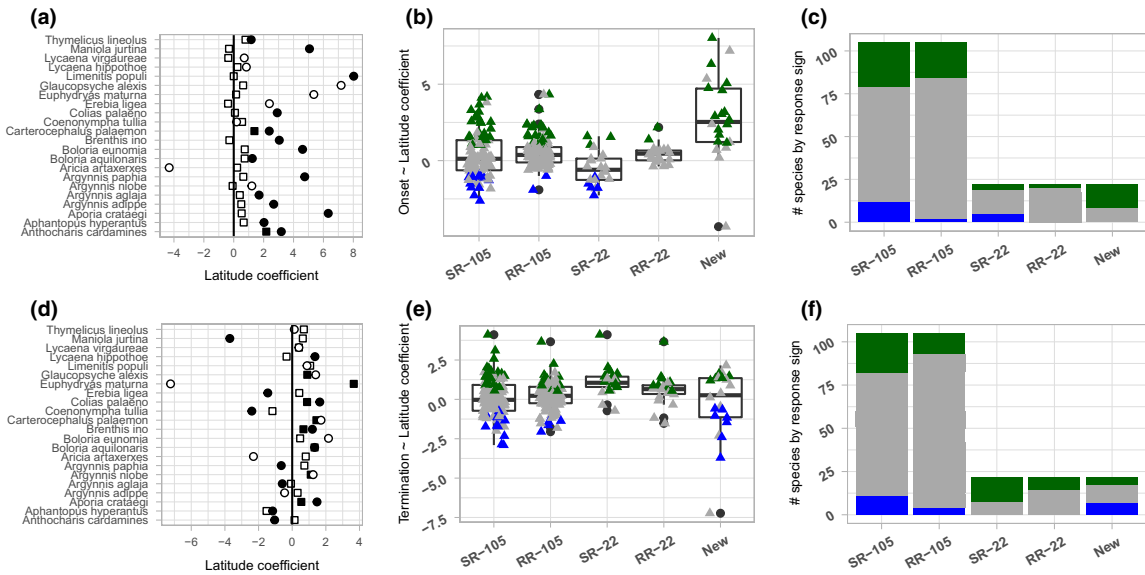

Figure 2 Comparison of model parameters for onset (panels a, b, c) and termination (panels d, e, f) dates as a function of latitude. Panels a and d show latitude coefficients for 22 species onset and termination phenology. Values extracted from Fric *et al.* (2020) Table S2 using regression of residuals against latitude, corrected for year and altitude, are shown as squares. Circles are parameters from the reanalysis. Filled symbols indicate significant slopes with latitude. Panels b and e show values and boxplots of the model coefficients for DOY ~ latitude, coloured by response: positive (green), non-significant (grey) or negative (blue) correlations. Panels c and f are stacked barplots of the response signs representing the number of datasets demonstrating positive (green), non-significant (grey) or negative (blue) correlations with latitude. Panels b, c, e and f show four sets of model results extracted from Fric *et al.* Table S2, and one set of reanalysis results. For Fric *et al.* (2020) results, 'SR' represents single regressions for DOY ~ latitude, whereas 'RR' represents regression of residuals against latitude, corrected for altitude and year. 'SR-105' and 'RR-105' show results for all 105 datasets in the original analysis, whereas 'SR-22' and 'RR-22' only include results for the 22 species-region datasets used in our reanalysis. 'New' shows results from our curation and reanalysis of those 22 species-region datasets. See Figure S1 for a comparison of species-specific data and results.

unbiased parameters, we modelled each species phenometric using multiple regression (DOY ~ latitude + year) using R version 3.6.2 (R Core Team 2019). We compared our results to those from Fric *et al.* Table S2.

We were unable to validate most patterns reported in the original study. Our results varied substantially in both onset and termination across species (Figure 2). In contrast to Fric *et al.* (2020), we found significantly later and/or shorter flight periods at higher latitudes for most species. These new results were consistent with the latitudinal gradient in climate and growing season length (Kobayashi *et al.* 2016). This evidence of inaccurate phenological patterns also discredits Fric *et al.*'s downstream trait analyses.

Despite this critique, we recognise that occurrence data have great potential to address many ecological questions. New aggregations of large datasets provide valuable inputs for macroscale ecological research, and the sheer amount of data accumulated across time and space may provide statistical power. However, 'with great power must come great responsibility' (Lee 1962); robust scientific inference requires careful data curation and robust analytical models. Other phenology metrics are less confounded by abundance and effort (Belitz *et al.* 2020); integrated community models with random species effects or informed priors better suit community phenological analyses (Ellwood *et al.* 2012). We enthusiastically support continued digitisation and use of collection data in ecological analysis, but urge researchers to exercise caution when using these data.

AUTHORSHIP

EAL conceived of the comment and designed the reanalysis. VS collated trait data and performed reanalysis. EAL and VS wrote and revised the manuscript together.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ele.13731.

DATA AVAILABILITY STATEMENT

No new occurrence data were used; only the data file provided in Fric *et al.* (2020) supplemental files (https://doi.org/10.1111/ele.13419). Voltinism data and sources are provided in Table S1. All data and code used in this reanalysis are available on github at https://github.com/RiesLabGU/Larsen-Shirey2020_EcoLettersComment.

REFERENCES

Belitz, M.W., Larsen, E.A., Ries, L. & Guralnick, R.P. (2020). The accuracy of phenology estimators for use with sparsely sampled presence-only observations. *Methods Ecol. Evolution*, 11(10), 1273–1285. https://doi.org/10.1111/2041-210X.13448.

Ellwood, E.R., Diez, J.M., Ibáñez, I., Primack, R.B., Kobori, H., Higuchi, H. & *et al.* (2012). Disentangling the paradox of insect phenology: are temporal trends reflecting the response to warming? *Oecologia*, 168(4), 1161–1171. https://doi.org/10.1007/s00442-011-2160-4.

Freckleton, R.P. (2002). On the misuse of residuals in ecology: regression of residuals vs. multiple regression. *J. Anim. Ecol.*, 71(3), 542–545. https://doi.org/10.1046/j.1365-2656.2002.00618.x.

Fric, Z.F., Rindoš, M. & Konvička, M. (2020). Phenology responses of temperate butterflies to latitude depend on ecological traits. *Ecol. Lett.*, 23(1), 172–180. https://doi.org/10.1111/ele.13419.

Karlsson, B. (2014). Extended season for northern butterflies. *Int. J. Biometeorol.*, 58(5), 691–701. https://doi.org/10.1007/s00484-013-0649-8.

Kobayashi, H., Yunus, A.P., Nagai, S., Sugiura, K., Kim, Y., Van Dam, B. *et al.* (2016). Latitudinal gradient of spruce forest understory and tundra phenology in Alaska as observed from satellite and ground-based data. *Remote Sens. Environ.*, 177, 160–170. https://doi.org/10.1016/j.rse.2016.02.020.

Lee, S. (1962) Amazing Fantasy #15, Marvel.

Matechou, E., Dennis, E.B., Freeman, S.N. & Brereton, T. (2014). Monitoring abundance and phenology in (multivoltine) butterfly species: a novel mixture model. *J. Appl. Ecol.*, 51(3), 766–775. https://doi.org/10.1111/1365-2664.12208.

Pearse, W.D., Davis, C.C., Inouye, D.W., Primack, R.B. & Davies, T.J. (2017). A statistical estimator for determining the limits of contemporary and historic phenology. *Nature Ecology & Evolution*, 1 (12), 1876–1882. https://doi.org/10.1038/s41559-017-0350-0.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ries, L., Zipkin, E.F. & Guralnick, R.P. (2019). Tracking trends in monarch abundance over the 20th century is currently impossible using museum records. *Proc. Natl Acad. Sci.*, 116(28), 13745–13748. https://doi.org/10.1073/pnas.1904807116.

Taylor, S.D. & Guralnick, R.P. (2019). Opportunistically collected photographs can be used to estimate large-scale phenological trends (preprint). https://doi.org/10.1101/794396.

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.*, 7(9132), 1–14. https://doi.org/10.1038/s41598-017-09084-6.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.