Check for updates

Received: 20 May 2026 | Accepted: 29 September 2026

DOI: 10.1111/2041-210x.70435

Methods in Ecology and Evolution

RESEARCH ARTICLE

A new approach for detecting phenological change using Time of Emergence (ToE) methodology

Reid W. Steele[1] | **Marc S. Botham**[2] | **Michael Dowd**[3] | **Ian Middlebrook**[4] | **Derek P. Tittensor**[1]

[1]Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada

[2]UK Centre for Ecology and Hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire, UK

[3]Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, Canada

[4]Butterfly Conservation, Wareham, UK

Correspondence
Reid W. Steele
Email: reid.steele@dal.ca

Funding information
Natural Sciences and Engineering Research Council of Canada

Handling Editor: Daijiang Li

Abstract

1. Many species are responding to environmental change by shifting the timings of phenological events such as migration or reproduction. These shifts are typically reported as rates of change, which do not indicate when such shifts transcend their original states, nor consider whether species experience novel environmental conditions or phenological mismatch with other species at their phenological events.
2. We present a new, generalizable approach to detect and contextualize shifts in phenological data regardless of taxonomic group or location by applying statistical and empirical Time of Emergence (ToE) metrics derived from climate science.
3. Our approach conclusively identifies when species depart from their baseline phenological event timings and determines whether they experience novel environmental conditions or phenological mismatch with other species during their phenological event. In combination, these metrics detect emergence in timing, environment and community synchrony for species' phenological responses.
4. We explore these metrics using simulations then apply them to a diverse set of example datasets (butterflies, marine plankton and cherry blossoms). We then develop community-wide indicators of phenological emergence, environmental emergence and species mismatch, which can be applied to inform biodiversity policy at any scale and for any taxa.
5. These methods are available via the *phenometrics* R package.

KEYWORDS
applied ecology, community ecology, modelling (community ecology), monitoring (community ecology), policy, spatial or time-series, species interactions, statistics

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2026 The Author(s). *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

1 | INTRODUCTION

Many species are changing the timings of seasonal events such as migration and reproduction—their phenology—in response to environmental change. For example, higher temperatures can trigger earlier onset of spring and summer phenological events across a broad range of taxa, trophic levels and environments (Badeck et al., 2004; Inouye, 2022; Poloczanska et al., 2013). Phenology can also shift due to various other factors, such as precipitation and nutrient availability (Desmit et al., 2020; Thackeray et al., 2016). Rate of change in phenology can vary significantly between species, which can impact both individual populations and entire communities (Poloczanska et al., 2013; Thackeray et al., 2016; Visser & Both, 2005). For example, phenological desynchronization between predators and prey may lead to match-mismatch dynamics, negatively affecting predator populations (Cushing, 1990; Møller et al., 2008; Visser & Both, 2005; Visser & Gienapp, 2019). Such impacts necessitate detection, quantification and monitoring of phenological responses to ecological and environmental change.

This task requires evaluating *environmental context*, such as the temperature at which a bird migrates (e.g. Keogan et al., 2018), and *community context*—that is, the phenological synchrony with other species in the same ecosystem (Visser & Both, 2005; Visser & Gienapp, 2019), such as between a bird's migration and the emergence of its prey. Environmental context is necessary to determine whether a species' phenology is maintaining its environmental optima (Figure 1). Community context is necessary to identify timing mismatches with other interacting species (Edwards & Richardson, 2004; Visser & Both, 2005).

Phenological studies generally evaluate annual *rates* of change in either an environmental variable (usually temperature) or event timing (Keogan et al., 2018; Poloczanska et al., 2013; Thackeray et al., 2016). While useful, these do not conclusively identify when species depart previous phenological timings or experience novel environmental conditions. In contrast, the Time of Emergence (ToE) metrics frequently used in climate science apply empirical or statistical techniques to detect when the climate transcends historical norms and establishes a new regime. ToE metrics complement rate of change analyses; rate analyses detect whether a species' phenology is *changing*, while ToE metrics detect whether it has conclusively *changed*. Through this dynamic, ToE metrics could pose and answer novel research questions in phenology.

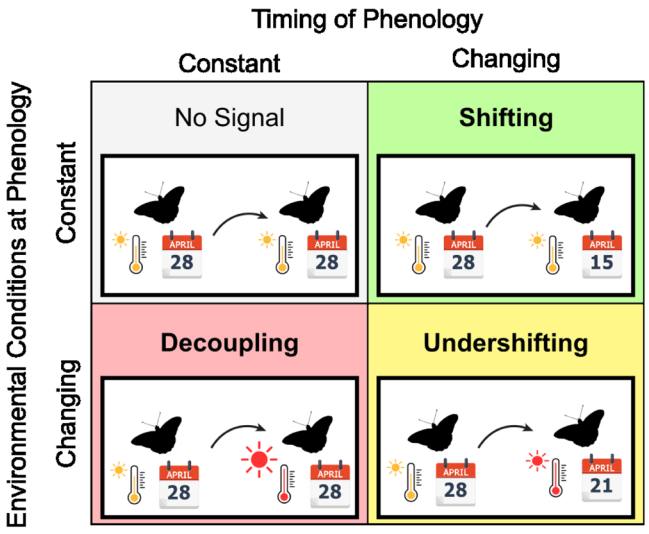


FIGURE 1 Classification of phenological responses based on phenological timing and environmental conditions.

ToE metrics also provide useful indicators for climate and adaptation policy through quantifying and comparing climate risk between locations (Chadwick et al., 2019; Mora et al., 2013; Zhang & Gao, 2023) and providing specific time horizons for adaptation (Hawkins & Sutton, 2012; Zhang & Gao, 2023). They are simple to interpret and calculate, and relatively standardizable across space, time and taxa (Mora et al., 2013). Climate ToE metrics have been used to detect and predict the establishment of new climatic regimes unfamiliar to the organisms and/or habitats experiencing them (Mora et al., 2013; Zhang & Gao, 2023). However, these are *indirect* metrics calculated from climate and not biological data, which do not account for how species *respond* to novel climates. We propose applying ToE metrics *directly* to biological data to capture this response.

Identifying differing species responses to a changing environment remains crucial. New environmental conditions may be fully offset by a species shifting its phenological event timing (e.g. migrating earlier), or there may be no phenological shift but consequently an environmental shift (e.g. migration timing unchanged but occurring at a warmer temperature). Additionally, species-level variation can lead to ecological mismatches (e.g. between predator and prey emergence from hibernation).

Here we develop ecological ToE metrics to detect conclusive changes in a species' phenological event timing, the environment under which it occurs, and phenological mismatch with other species. These metrics are designed to be simple and broadly applicable to maximize their relevance across different taxa, ecosystems, time periods, sampling methodologies and phenological events. We then develop three community-level indicators for monitoring phenological desynchrony and the proportion of species in a community that have transcended climatic and/or phenological baselines. These indicators can then be plotted over time, akin to other aggregate policy-relevant indicators such as the Red List Index or Living Planet Index (Butchart et al., 2004; Collen et al., 2009).

We apply these indicators and metrics to both real and simulated datasets to demonstrate their utility on a diverse set of taxa and time series. To ensure widespread accessibility, we have combined these methods, metrics and indicators into an R package titled *phenometrics*.

2 | METHODS

2.1 | Time of Phenological Emergence (ToPE) metric calculation

The Time of Phenological Emergence (ToPE) metric identifies the year at which phenological event timing has conclusively shifted. Ecological time series are often short, and pre-industrial baselines

2041210x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

are rare. To combat this, we compare annual time series of phenological event timing to a null time series with constant phenological timing, rather than setting aside a historical period as is typical for climate ToE studies (Mora et al., 2013; Zhang & Gao, 2023).

First, a linear regression is fit to an annual phenological time series of event timing $p_{t,r}$ in day of year, where t is time in years and r represents sample replicates (e.g. sites), to obtain the slope (β) (Figure 2a). This slope is subtracted from the original time series, creating a detrended counterfactual $d_{t,r} = p_{t,r} - \beta t$, representing a null condition under which the species' phenological timing remains flat from the first year while retaining the variability and sample size of the original time series (Figure 2b). This methodology assumes that the time series describes a single phenological event (e.g. only the first seasonal peak each year), and that any trend in the time series is linear and constant (see Supporting Information S2: Sections 2.3 and 2.4 for alternatives when these assumptions are violated).

Next, an annual test is applied to determine whether each year of the original time series exceeds the detrended counterfactual (Figure 2c). We provide both statistical and empirical methods for this, and run simulations to determine when best to apply each. To calculate statistical emergence, each year of event timing data is compared to $n_b = 100$ bootstrap samples of the counterfactual ($d_{t,b}$) for that same year using one-sided Kolmogorov–Smirnov (KS) tests, under the null hypothesis (H_0) that the empirical cumulative distribution function of event timings in the original dataset is less than or equal to that of the counterfactual bootstrap samples (this requires sufficient sample size within each year). We define:

$$\mathrm{KS}_b = \begin{cases} 1 & \text{if } H_0 \text{ is rejected} \\ 0 & \text{if } H_0 \text{ is not rejected} \end{cases} \quad (1)$$

where KS_b is the binary indicator of the KS test result for bootstrap sample b. If there is an advancing phenological trend in the original time series, the difference in event timing between the original and counterfactual will grow each year until the signal in the original time series may exceed the noise in the counterfactual. We define a positive test result ($T_t = 1$) when 60% of the bootstrapped KS tests for an individual year are statistically significant at $\alpha = 0.05$, as per Chadwick et al. (2019):

$$T_t(\text{statistical}) = \begin{cases} 1 & \text{if } \frac{1}{n_b}\sum_{b=1}^{n_b} \mathrm{KS}_b \geq 0.6 \\ 0 & \text{if } \frac{1}{n_b}\sum_{b=1}^{n_b} \mathrm{KS}_b < 0.6 \end{cases} \quad (2)$$

For empirical emergence, we compare the annual mean event timings of the original time series (p_t) to a threshold value derived from the quantiles of the detrended time series (Mora et al., 2013; Zhang & Gao, 2023). We define a positive test result as the annual mean event timing of the original time series dropping below the 25% quantile ($Q_{0.25}$) of mean event timings across all years in the counterfactual (d)

$$T_t(\text{empirical}) = \begin{cases} 1 & \text{if } p_t < Q_{0.25}(d) \\ 0 & \text{if } p_t \geq Q_{0.25}(d) \end{cases} \quad (3)$$

As it is based on annual means, empirical emergence testing is agnostic to sample size within years. To guard against stochasticity, we apply an emergence threshold of five consecutive years for defining emergence (Figure 2d) for both the statistical and empirical approaches (Chadwick et al., 2019; Zhang & Gao, 2023). The ToPE is thus the first year of five or more consecutive test results.

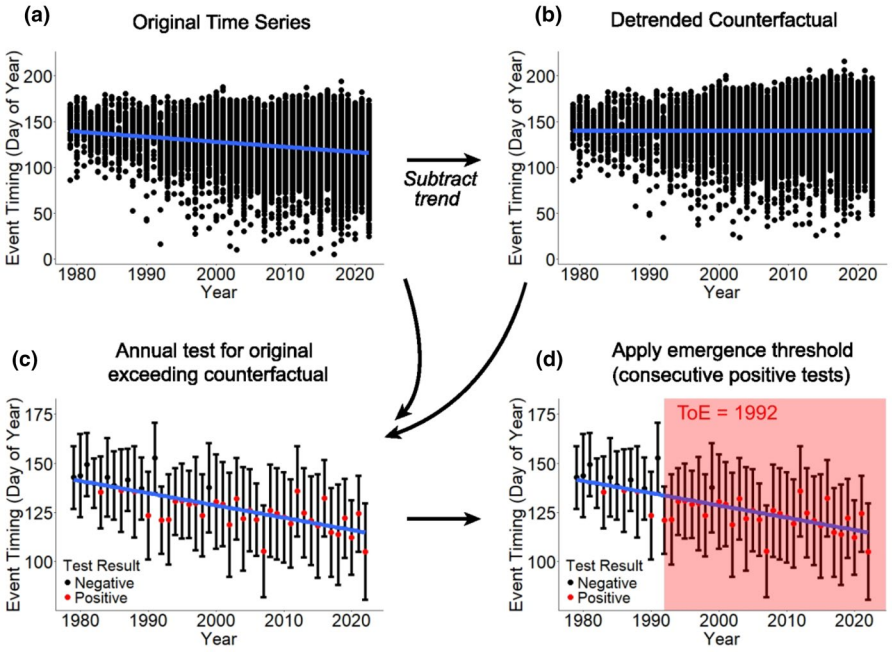


FIGURE 2 Conceptual workflow describing the Time of Phenological Emergence (ToPE) test. The linear trend of the original time series (a: phenological event timing against year) is subtracted to generate a detrended counterfactual (b). For each year, a statistical or empirical test is used to determine whether the original time series exceeds the counterfactual (c). Emergence is defined as the first point at which five *consecutive* positive test results (i.e. emergence threshold = 5) are observed (d).

20412108x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2 | Time of Environmental Emergence (ToEE) metric calculation

The Time of Environmental Emergence (ToEE) metric defines the year when a species' phenology begins occurring under novel climatic conditions. The methodology is identical to the ToPE metric, except $p_{t,r}$ is replaced with the value of an environmental variable of interest related to the timing of the phenological event $v_{t,r}$ (henceforth we use temperature on the day of the phenological event, for example, though any variable at or across any time period of interest could be used). For the ToEE metric, the counterfactual scenario represents a null condition under which the value of the environmental variable at the time of the phenological event has no trend, while again retaining the variability and sample size of the original time series. We also reverse the directionality of the test, to identify the year in which the species' phenological temperature (i.e. temperature when its phenological event occurs) rises above the null condition. For the empirical test:

$$T_t(\text{empirical}) = \begin{cases} 1 & \text{if } v_t > Q_{0.75}(d) \\ 0 & \text{if } v_t \leq Q_{0.75}(d) \end{cases} \quad (4)$$

The criteria for KS_b and T_t(statistical) remain unchanged, but H_0 is reversed. ToEE is the first year beginning five or more consecutive test results.

2.3 | Classification

Species phenological responses to environmental change can be classified into descriptive categories based on these two metrics (Figure 1). If a species shifts its phenology through time while maintaining the same environmental conditions, it will emerge in ToPE but not ToEE (herein 'shifting'). Conversely, if a species has maintained its phenological timing but experiences a novel climate, it will emerge in ToEE but not ToPE ('decoupling'; i.e. its phenology is decoupled from the environment). If a species shifts its phenology through time, but insufficiently to account for environmental change, resulting in both a new phenological time and climate, it emerges in both ToEE and ToPE (herein 'undershifting'). Finally, if no phenological response is detected due to some combination of weak or no trend, high noise, small sample size or insufficient time-series length, the species will not emerge in either ToEE or ToPE ('no signal'). Hypothetically, a species may also 'overshift' its phenology beyond its environmental preferences, but this response is rare in our example data and we do not discuss it further here (see Supporting Information S2: Section 2.1 for more).

2.4 | Time of Mismatch (ToM) calculations

The Time of Mismatch (ToM) metric is designed to detect mismatches between two phenologies. After fitting linear regressions to time series of phenological event timing for both a focal ($f_{t,r}$) and comparison ($c_{t,r}$) species, the linear trend of the focal species (β_f) is replaced with the trend of the comparison species (β_c). This generates a counterfactual scenario ($m_{t,r}$) where the focal species shares the same linear trend in event timing as the comparison species, but maintains its original time-series variance and sample size:

$$m_{t,r} = f_{t,r} - \beta_f t + \beta_c t \quad (5)$$

This is preferable to testing against the comparison species' time series directly because comparison species sampling error is not transferable to the focal species, and the variability around the focal species' phenological trend may be ecologically relevant. For example, high variance in a species' phenology may indicate a bet-hedging strategy to guard against mismatch (e.g. Hočevar et al., 2021).

For the statistical approach, annual testing is performed as previously described, and ToM is defined as the first year of five consecutive annual positive test results. Unlike ToPE and ToEE, two-sided tests are employed for ToM calculations to detect emergence regardless of which test species has the greater trend in its phenology.

For empirical testing, we calculate the annual means of both $f_{t,r}$ (f_t) and $m_{t,r}$ (m_t) and fit a linear model to m_t to generate predicted values for each year ($\hat{m}_t$). The residuals of m_t for this model ($e_{m_t} = m_t - \hat{m}_t$) serve as a baseline to calculate empirical test quantiles, and the residuals of the original data to this model ($e_{f_t} = f_t - \hat{m}_t$) serve as test values. Empirical test values are calculated as:

$$T_t(\text{empirical}) = \begin{cases} -1 & \text{if } e_{f_t} > Q_{0.75}(e_{m_t}) \\ 0 & \text{if } Q_{0.25}(e_{m_t}) \leq e_{f_y} \leq Q_{0.75}(e_{m_t}) \\ 1 & \text{if } e_{f_t} < Q_{0.25}(e_{m_t}) \end{cases} \quad (6)$$

with values of −1 and 1 indicating positive test values in opposing directions. In the statistical method, KS_b and T_t(statistical) remain unchanged, but H_0 is now two-sided. ToM is defined by five consecutive positive test values in the same direction. This method can be used to detect mismatch between two ecologically linked species (e.g. predator and prey), or applied en-masse to quantify phenological desynchrony within a community.

2.5 | Community level indicators—Phenological and environmental emergence

To create community-wide indicators of phenological change, we aggregate individual species metrics to calculate the proportion under each Figure 1 classification. We define the Phenological Shift Index (PSI) as the proportion of species shifting their phenology (shifting or undershifting in Figure 1) and the Phenological Decoupling Index (PDI) as the proportion of species experiencing a new phenological temperature (decoupling or undershifting

2041210x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in Figure 1). These indicators quantify and classify changes in community-level phenological disruption over time.

2.6 | Community level indicator—Desynchrony

The Community Desynchrony Index (CDI) represents the mean proportion of comparable species pairs which have become desynchronized in their phenological event timing, defined as emergence in ToM. Comparable species have a mean phenological event timing within one standard deviation of the focal species in the first 10 years of data (i.e. have a similar initial phenological event timing). We then apply the ToM methodology to calculate the mean annual proportion of comparison species with which each focal species is desynchronized.

Table 1 lists all species-level metrics and community-level indicators of phenological change.

2.7 | ToPE sensitivity simulations

The number of years (and thus time-series length) required to detect emergence depends on the strength of the phenological trend, the noise around the trend and the number of within-year replicates (e.g. number of sites, individuals, etc.). We performed simulations testing years required for phenological emergence detection for various combinations of trend, noise and replication (Figure 3). For each combination of parameters, annual mean phenological event timing (p_t) was drawn from a normal distribution with mean $B + \beta t$ and standard deviation e, where B is the baseline mean day of year value (i.e. the intercept of event timing against year), which was fixed at 150, β is trend strength, represented as the linear slope of event timing against year, and e is the variability around the trend across years (i.e. the annual mean error in the linear regression of p_t against t), implemented as the standard deviation of p_t.

TABLE 1 Abbreviations and descriptions for metrics and indicators in this study.

Metrics	Acronym	Description
Time of Phenological Emergence	ToPE	Year at which phenological event timing has conclusively changed
Time of Environmental Emergence	ToEE	Year at which an environmental condition experienced at phenological event has conclusively changed
Time of Mismatch	ToM	Year at which phenological event timing of a focal species has conclusively desynchronized from a comparison species
Phenological Shift Index	PSI	Proportion of species in a community emerged in ToPE
Phenological Decoupling Index	PDI	Proportion of species in a community emerged in ToEE
Community Desynchrony Index	CDI	Mean proportion of comparison species with which each focal species has become mismatched in a community

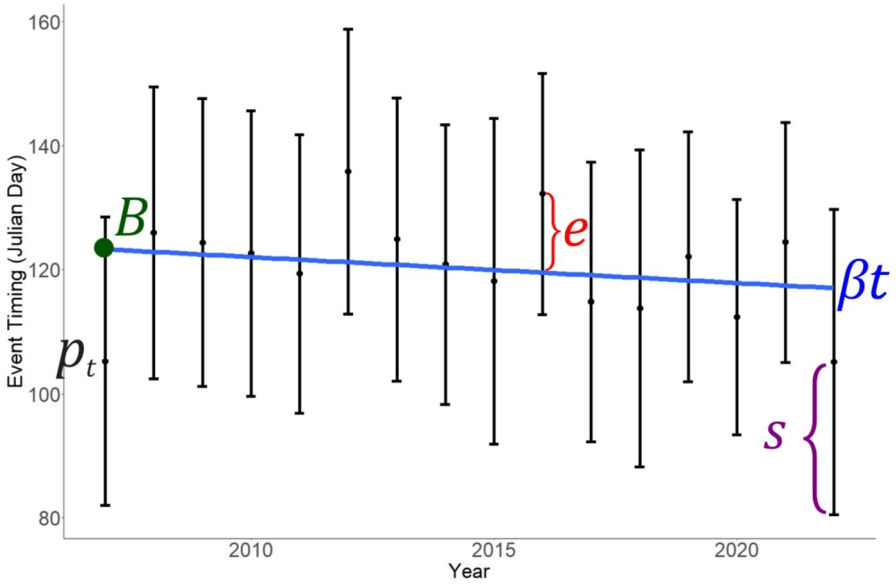


FIGURE 3 Example time series of phenological event timing against year (*Vanessa atalanta*, UKBMS), showing aspects of the time series represented by each simulation parameter: Annual mean phenological event timing p_t, baseline event timing B, across-year standard deviation e, within-year standard deviation s and predicted value at year βt, calculated from trend β and year t.

20412108x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Event timing values for each year ($p_{t,r}$) were then drawn from a normal distribution with mean p_t and standard deviation s, where $p_{t,r}$ is simulated event timing for year t and replicate r and s is sample standard deviation of event timings around p_t within each year. To reduce computational requirements, s was fixed at $1.5e$ (see Supporting Information S1 for details and rationale). We generated 50 test values for trend strength (β), variability (e) and sample size (r) across ranges informed by example datasets and performed simulations on every combination of two variables while fixing the third at its median value. For each combination , 100 50-year replicate datasets were simulated, both statistical and empirical ToPE calculated and mean first year of emergence plotted.

By default, we use entire time series to test the effect of baseline length on emergence detection, meaning the baseline grows as the time-series lengthens. To test sensitivity to baseline length disconnected from time-series length, we generated 1000 simulated 100-year time series using median simulation values of trend strength, sample size and variability, and calculated mean ToPE for baseline lengths ranging from 5 to 50 years, using both statistical and empirical methodology. To determine the sensitivity of ToE to the emergence threshold, we also calculated mean ToPE separately for emergence thresholds of 5, 10 and 15 years.

2.8 | Example datasets

We applied our methodology to three example datasets to demonstrate the utility of our metrics and indicators across a range of taxa, ecosystems and sampling methodologies.

2.8.1 | United Kingdom Butterfly Monitoring Scheme (UKBMS)

The UKBMS has measured phenological metrics for 59 species of butterflies from 2500 sites across the United Kingdom from 1976 to 2023 ($n = 903{,}399$) using a variety of transect-based surveys (Botham et al., 2023). Due to the large within-year sample sizes, we used this dataset to test statistical ToE metrics and indicators. Single observation combinations of species, site and year, as well as years prior to 1979 were removed ($n = 430{,}413$), as multiple surveys within a year are required to identify phenological timing and data coverage prior to 1979 was limited. The phenological event analysed for each species was the mean flight date for the first flight period of the year (except for *Aglais io* and *Gonepteryx rhamni* for which the second flight period was used to exclude overwintering adults), representing the emergence of the first generation of adult butterflies from overwintering each year (or first arrival in the case of migratory species *Vanessa atalanta*, *Vanessa cardui* and *Colias croceus*). To ensure we analysed only a single phenological event, data succeeding the first generation of the year was excluded for multivoltine species. For environmental context, we matched the mean flight date for each butterfly species at each site and year ($n = 472{,}986$) to that day's temperature from the Hadley Center Central England Temperature (HadCET) dataset (Met Office, 2025). We then calculated statistical ToPE, ToEE and ToM metrics for every species, as well as the PSI, PDI and CDI indicators, with site as the unit of replication. Results were divided into primarily univoltine (single generation) and primarily multivoltine (multi-generation) species to determine whether their phenological responses differ.

2.8.2 | Continuous plankton recorder (CPR)

The CPR survey is a long-running (1958–present), large-scale North Atlantic marine plankton survey, conducted via ships of opportunity towing near-surface plankton samplers on their normal routes (Edwards & Richardson, 2004). We analysed CPR records of phytoplankton and zooplankton abundance from the North Sea (51°N–61°N, 2.5°W–10°E) from 1985 to 2021 (Helaouët, 2024). We divided this area into 4°×4° grid cells, and calculated the phenological event for each species and grid cell as the day on which 25% of the cumulative abundance for that year was achieved to measure of the start of the peak abundance period, which should correspond with the spring bloom (Greve et al., 2005). To eliminate small sample size effects, only species with at least 1 month averaging ≥200 samples over the full study area and period were included. Grid cells with <25 individual samples in a year or with event timing outliers (i.e. below $Q1 - 1.5\text{IQR}$ or above $Q3 + 1.5\text{IQR}$) were also removed to ensure the event timings represented actual phenological timings, rather than statistical artefacts resulting from low sample sizes. To ensure we analysed only a single phenological event (i.e. first peak of the year) and to prevent event timing misspecification, 27 species with multimodal annual abundance patterns were removed (P. Hélaouët, pers. comm., 12 May 2025). The final data set consisted of 3395 records of 16 species. For environmental context, event dates and locations were matched to mean daily 5 km resolution satellite sea surface temperature averaged over each grid cell area, sourced from NOAA Coral Watch (NOAA Coral Reef Watch, 2025). Due to small, but highly informed, within-year sample size (i.e. 4°×4° grid cells), we calculated metrics and community-level indicators using the empirical methodology.

2.8.3 | Kyoto cherry blossom

To serve as a unique long-term example for the *phenometrics* package, we also applied our methodology to Kyoto cherry tree (*Prunus jamasakura*) full-flowering timing from 895 to 2025 A.D. (Aono, 2025) collected from historical records and diaries (Aono & Kazui, 2008). This dataset lacks within-year replication, meaning we could only apply the empirical approach. To reduce noise and focus on long-term trends, we used a 20-year recursive moving average of day of year of full-flowering. Due to missing data, years before 1000 A.D. were removed. Given the potential for significant climatic regime changes over the >1000 year time series, we applied two-sided

2041210x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ToPE tests to a maximum of 50 years of the moving average at a time using a sliding window. When emergence did occur, the time series was considered emerged until five consecutive negative test results, representing an equivalent period of 'unemergence'.

3 | RESULTS

3.1 | Simulations

Years to emergence correlated positively with variability (*e*) and negatively with trend strength (β) for both empirical and statistical approaches (Figure 4). Sample size (*r*) was negatively correlated with years to emergence for statistical ToPE but did not affect empirical ToPE (Figure 4). Increasing sample size and trend strength had diminishing effects on statistical ToPE at high values (>~100 replicates and >~0.5 days/year, respectively, Figure 4). Generally, the statistical approach detected emergence more quickly than the empirical approach with greater sample sizes and vice-versa, with inflection occurring around a sample size of 20 (Figure 4).

Baseline length simulations showed that statistical ToE increases with baseline length until reaching an asymptote, while empirical ToE increases very slowly with baseline length (Figure 5). Emergence threshold had no effect on the timing of statistical emergence but a large effect on empirical ToE.

3.2 | Example datasets

Most species in the UKBMS dataset were classified as undershifting by time-series end (Figure 6), meaning their phenologies established both novel timing and temperature (Figure 1). Shifting was more prevalent than decoupling, and pure decoupling responses were rare (Figure 6). Most species with no signal were missing multiple years of data and were less frequently observed (Figure 6). Univoltine and multivoltine species exhibited similar patterns in

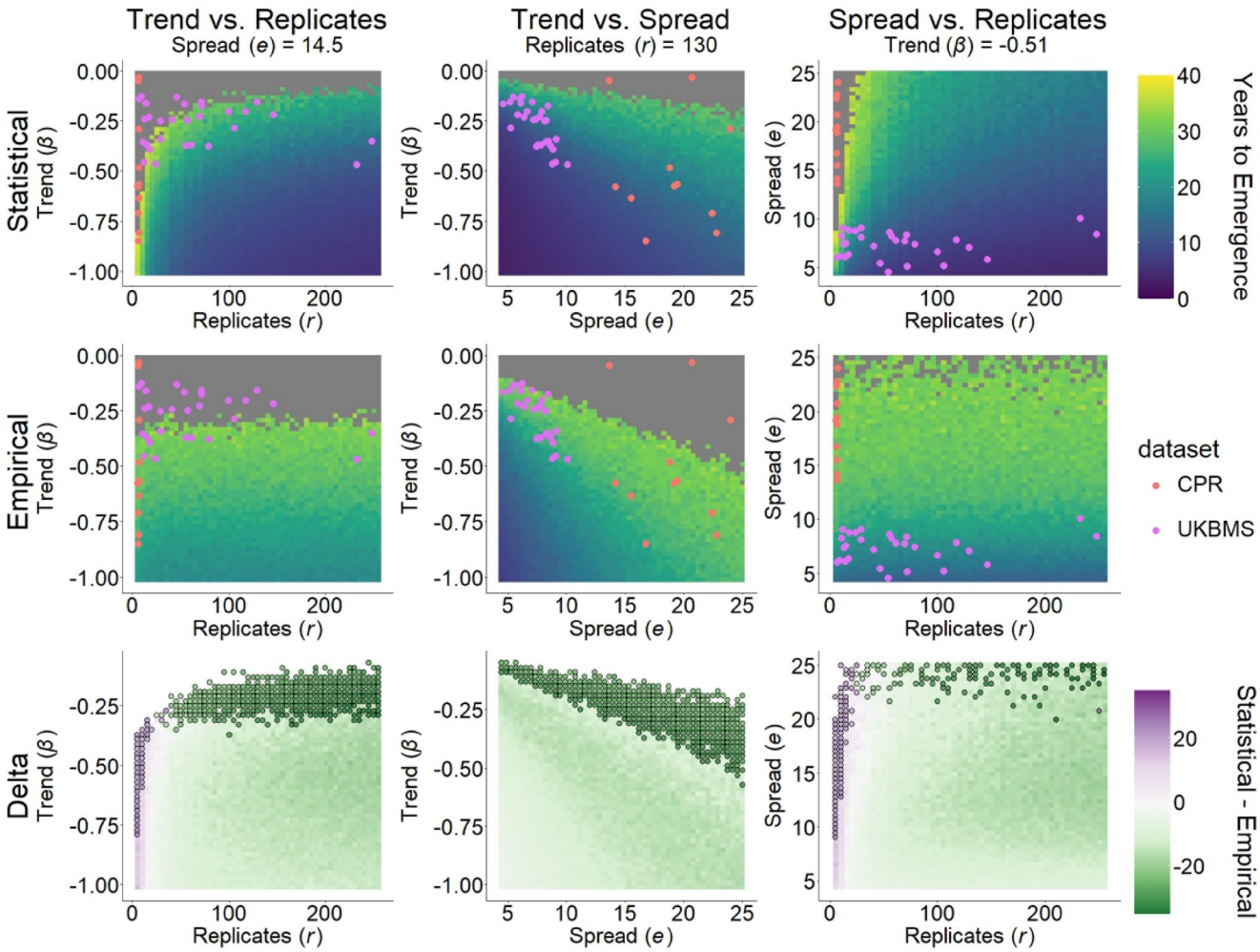


FIGURE 4 Results of simulation testing, showing mean years to emergence from statistical (top) and empirical (middle) Time of Phenological Emergence (ToPE) testing for different combinations of trend β, replicates *r* and variability *e*. Grey pixels represent cases where >50% of simulations did not emerge within 50 years. Points on the top two rows indicate empirical β, *e* and *r* values observed from individual species in the UKBMS (purple) and CPR (red) test datasets. Points on the bottom row indicate cases where emergence occurred in only one test.

20412108x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

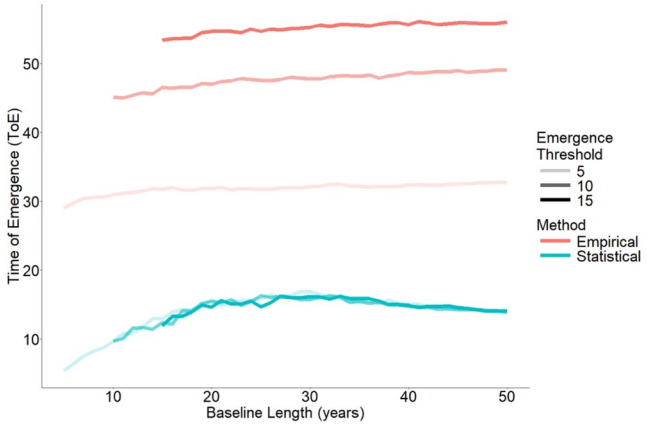


FIGURE 5 Sensitivity analysis for Time of Emergence against baseline length, that is, the number of years used to generate the detrended counterfactual and empirical test quantiles. Baselines begin in initial time-series year. Emergence thresholds are consecutive positive test results required to define emergence. Results are means of 1000 simulations generated using median values of trend strength, variability and sample size.

their community desynchrony (Figure 6). Species with no signal also appeared desynchronized with few or no species, likely due to limited data coverage (Figure 6).

Most species in the CPR dataset were classified as shifting, with only one classification each for decoupling and undershifting. Three exhibited no signal (Figure 6), which, unlike the UKBMS dataset, did not appear linked to missing years of data or smaller sample sizes, indicating possible true null responses. Mismatch proportions for individual species were lower than in UKBMS and seemed largely unrelated to final phenological classifications and trophic levels (Figure 6).

In both the CPR and UKBMS datasets, PDI was lower than PSI (Figure 7). The difference was more pronounced in the CPR dataset, where only two species exhibited environmental emergence. Both datasets achieved similar PSI values, though phenological emergences were detected more quickly in the CPR dataset (Figure 7). The rate of emergences detected in the UKBMS dataset appeared to slow over time, while CPR PSI still appeared to be increasing linearly (Figure 7). CDI increased linearly through time in both datasets, more quickly in the UKBMS dataset, with neither appearing to slow (Figure 7).

Eleven periods of emergence were detected in the cherry blossom time series, alternating between advancing and delaying phenology (Figure 8). Of these periods, six were associated with earlier full-flowering dates, indicating warmer periods, while five were associated with delays in full-flowering timing, indicating cooler periods (Aono & Kazui, 2008). Emergence period length varied from 35 to 125 years. The final period exhibited the greatest number of consecutive positive test results, likely due to a combination of anthropogenic climate change causing a stronger trend in bloom timing and more reliable data sources near the end of the time series which noticeably reduced spread (Figure 8).

4 | DISCUSSION

Our ToE-based phenology metrics and *phenometrics* R package provide a standardized methodology to identify when the timing or environmental conditions associated with species' phenological events deviate from their original states. Metrics can be calculated using both statistical and empirical approaches, depending on the dataset and user preference. The ToPE and ToEE metrics are simple to calculate and interpret, and broadly applicable across taxa, ecosystems, environments, time periods, sampling methods and phenological event types. Unlike other ToE metrics, they do not require sacrificing data to create a historical baseline (Mora et al., 2013; Zhang & Gao, 2023), and directly account for the actual biological responses of species to changes in their environments. When summarized into indicators, they can be used to compare (and potentially project) aggregate phenological responses across disparate ecosystems, classify phenological responses, compare species' phenologies to evaluate mismatches and quantify overall phenological desynchrony within communities. The simplicity and comparability of these indicators across ecosystems (Figure 7) suggests their potential utility as policy and monitoring tools for tracking potential climate impacts.

4.1 | Simulation framework

The simulation framework (Figure 4) enabled evaluation of time-series length required to detect phenological emergence for a range of data-informed scenarios. For real-world applications, this can help estimate the monitoring period required to detect shifts in a species' phenology and corresponds to real-world data showing that at least a decade (and preferably several) of observations are required to definitively detect a change in phenology with temperature (Inouye, 2022). Long-term time series are extremely valuable in ecology (Lindenmayer et al., 2012), and these findings emphasize the importance of long-term phenological monitoring programmes. However, such time series are rare, so the *phenometrics* package includes functionality to simulate future phenological event timings. This allows users to project future emergence timings: a common use case for climate ToE metrics (Chadwick et al., 2019; Mora et al., 2013; Zhang & Gao, 2023) but novel for ecology.

Baseline length simulations showed the statistical methodology was more sensitive to increasing baseline length than the empirical methodology, but neither was sensitive with sufficient time-series length (Figure 5). When baseline length is short, it likely does not contain sufficient representation of system variability, leading to increases in ToE as more data are collected, particularly in the statistical case. Conversely, once sufficient time-series length has been reached and emergence detected, the baseline may be fixed to prevent ToE from shifting as future data collection slightly shifts the baseline (however, note that once sufficient time-series length is reached, shifts in ToE due to the addition

2041210x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

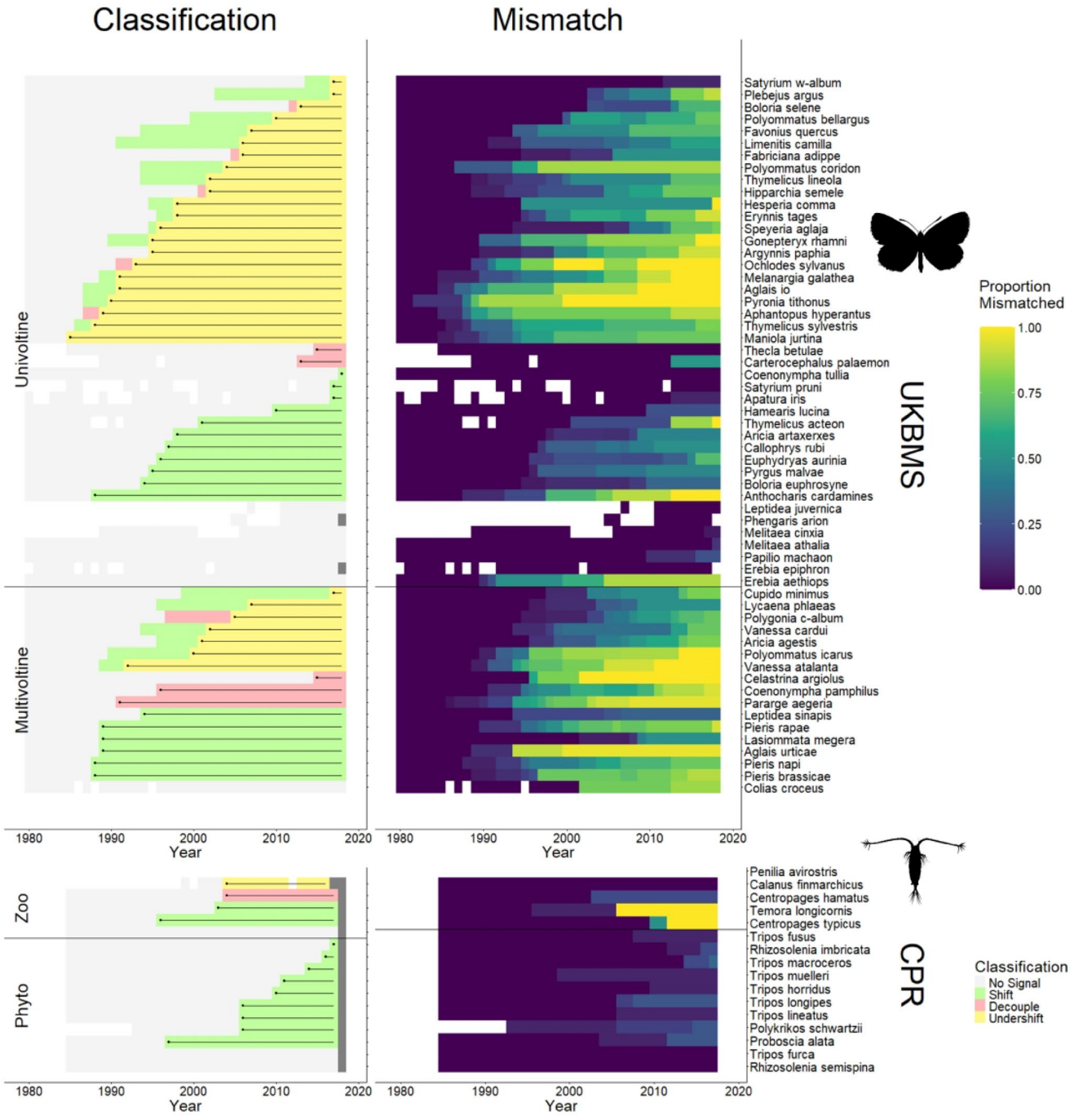


FIGURE 6 Left Panels: Classification by species and year from Time of Phenological Emergence (ToPE) and Time of Environmental Emergence (ToEE) metrics for mean flight date in the United Kingdom Butterfly Monitoring Scheme (UKBMS) dataset, and 25% annual abundance peak in the North Sea Continuous Plankton Recorder (CPR) dataset. Right Panels: Time of Mismatch (ToM) metric by species and year for both datasets. ToM is the proportion of comparison species with which each focal species is mismatched.

of future data tend to be slight as in Figure 5). We discuss this in Supporting Information S2: Section 2.3.2.

The empirical methodology was also much more sensitive to emergence threshold (Figure 5), because it tests each year against quantiles calculated from past years, whereas the statistical method tests each year against an adjusted version of the same year. This means stochasticity in the opposite direction to the trend can prevent emergence in empirical, but not statistical testing. However, this trade-off also allows empirical testing to work at any sample size, while statistical testing is ineffective at low sample sizes (~<20). Situationally, the delay in empirical emergence due to increasing the emergence threshold may be undesirable if it significantly delays emergence due to a single extreme year, or desirable if it prevents emergence due to low-frequency variability followed by a return to previous conditions. We use an emergence threshold of 5 as common in climate ToE studies (Chadwick et al., 2019; Zhang &

2041210x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

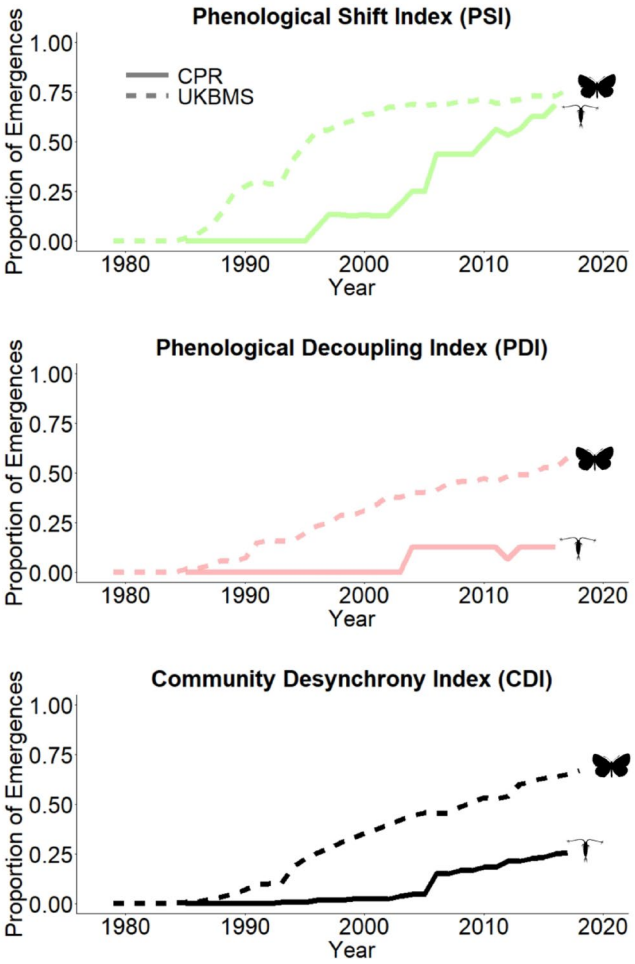


FIGURE 7 Community-level Phenological Shift Index (PSI, Top), Phenological Decoupling Index (PDI, Middle) and Community Desynchrony Index (CDI, Bottom) indicators by year for the North Sea Continuous Plankton Recorder (CPR, solid line) and United Kingdom Butterfly Monitoring Scheme (UKBMS, dashed line) datasets. CPR indicators calculated using empirical metrics; UKBMS indicators calculated using statistical metrics.

Gao, 2023). We discuss how this value may be changed in Supporting Information S2: Section 2.2.3.

4.2 | Species-level metrics

Our metrics are useful for detecting and identifying conclusive shifts in the timing of (ToPE) and environment experienced at (ToEE) a phenological event. Specifically, ToEE separates null responses due to a stable climate from decoupling responses resulting from a lack of phenological plasticity. Bringing attention to decoupling responses is important as the difference between no signal and decoupling responses (Figure 1) can be ecologically profound (Keogan et al., 2018; Møller et al., 2008), yet may be undetected or ignored if environmental context is not considered (e.g. Poloczanska et al., 2013). Furthermore, both decoupling and undershifting may challenge the common modelling assumption that species phenologies shift linearly with temperature (Roy & Sparks, 2000; Thackeray et al., 2016), which could lead to overprediction of phenological shifts.

In the UKBMS dataset, most species ended as undershifters (Figure 6), indicating a partial capability to shift phenological timing in response to climate change, which concurs with previous research (Diamond et al., 2011; Roy & Sparks, 2000). Conversely, the change in phenological temperature is novel and may indicate a limit to phenological plasticity in response to climate change, suggesting other constraints on species' phenologies and/or non-phenological adaptations to climate change (Roy et al., 2015).

For the CPR dataset, the onset of the peak abundance period shifts earlier for most species (Figure 6), which corresponds with North Sea plankton phenology literature (Desmit et al., 2020; Edwards & Richardson, 2004; Mackas et al., 2012). The notable exception was *C. finmarchicus*, which we classified as decoupling, despite literature identifying it as shifting (Mackas et al., 2012). This appears related to the time-periods of the respective studies. We did not consider years before 1985 due to the lack of high temporal resolution temperature data. However, there is a clear phenological shift between 1958 and 1984, and phenological emergence is detectable in the full dataset (1958–2021) despite the lack of phenological shifting from 1985 to 2021, suggesting the species truly is undershifting. This exemplifies the importance of shifting baselines and the availability of high temporal resolution environmental data, though neither of these issues is unique to our methodology.

Most phenological emergence periods identified in the Kyoto cherry blossom dataset were congruous with warm and cool periods identified from temperature reconstructions (Aono & Kazui, 2008), although generally corresponding to periods after phenological shifts occurred, rather than during (Figure 8). This is likely due to the effect of pre-trend data in the 50-year moving window, as well as the time and trend required to reach emergence and de-emergence thresholds.

The mismatch (ToM) metric quantitatively tests for phenological mismatches between two species (Inouye, 2022; Visser & Both, 2005). By summarizing across a community (the CDI indicator), phenological desynchrony within an ecosystem can be quantified, and species most desynchronized from their phenological contemporaries identified (Figure 6, right). Such species may be at greater (e.g. predator mismatched with its prey) or lesser (e.g. prey mismatched with its predators) risk of negative ecological impacts. Interestingly, desynchrony did not always correlate with phenological response classification (Figure 1); for example, *C. finmarchicus* did not mismatch with any species despite being the only decoupler (Figure 6).

4.3 | Community-level indicators

Both the UKBMS and CPR datasets exhibited similar patterns in PSI and CDI indicators, but the CPR dataset exhibited far less

20412108x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

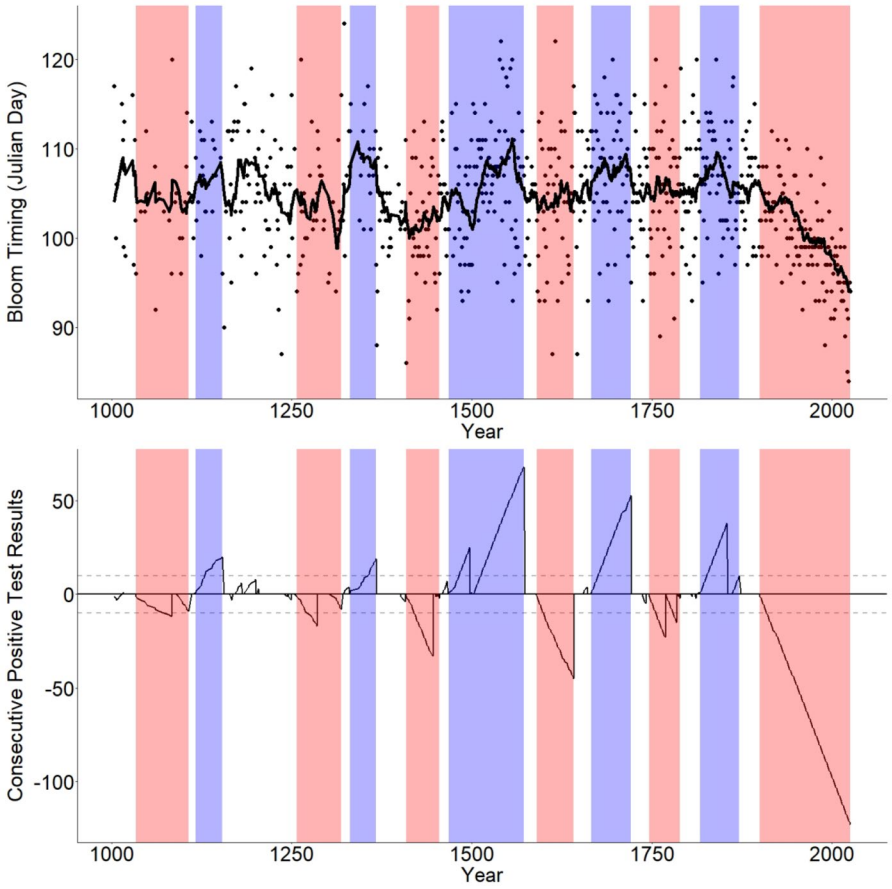


FIGURE 8 Time series of cherry blossom full-flowering timing with 20-year moving average (Aono, 2025) (top), and consecutive positive test results from empirical Time of Phenological Emergence (ToPE) testing on the 20-year average (bottom). Emergence periods are overlaid in red (advancing phenology) and blue (delaying phenology).

decoupling (PDI). Plausibly, plankton could respond more quickly to climate change than butterflies (Thackeray et al., 2016), but this could also result from better spatial matching of temperature data, lower temperature change or differing baselines. Interestingly, the similarity in PSI and CDI between datasets indicates similar variability and directionality in phenological responses within both ecosystems, as supported by previous research (Thackeray et al., 2016).

4.4 | Empirical versus statistical methodology

The primary difference between empirical and statistical methods is that the former is agnostic to within-year attributes (sample size and standard deviation) because testing uses annual means. This is preferable when sample size is low (<20), as statistical tests have little power at small sample sizes (CPR data; Figure 4). When sample size is sufficient and the time series noisy (high variability within and/or across years), such as species across a wide area or a large temporal range, then the statistical approach is preferable. In such cases, the empirical methodology misses the context of within-year variability and may be excessively conservative due to elevated across-year variability, causing it to miss statistically significant shifts. This is visible in the middle panels of Figure 4, where emergence is detected more quickly and frequently by the statistical method, particularly as variability increases.

4.5 | Assumptions and limitations

It is important that potential *phenometrics* end users understand the assumptions and limitations of our methodology. The two most important assumptions are (i) that time series constitutes a single phenological event and (ii) that changes in phenological timing and temperature are linear across the years used to calculate the detrended counterfactual (this assumption can be relaxed; see Supporting Information S2: Section 2.3). The first assumption is true of any phenological study, and explains why we constrain the UKBMS time series to only the first generation each year and only analyse unimodal abundance patterns in the CPR time series; otherwise, these time series could contain multiple phenological events rendering further analysis and interpretation meaningless. As to the second assumption, non-linear patterns in phenological time series may lead to patterns in the detrended counterfactual time-series and thus incorrect emergence timings.

A limitation of both our study and the broader climate ToE literature is the use of arbitrary thresholds to quantify emergence (Chadwick et al., 2019; Hawkins & Sutton, 2012; Zhang & Gao, 2023). We use these to define emergence from test values for both empirical (test quantiles) and statistical (proportion of significant KS tests) metrics. Compared to the climate ToE literature, we also use more liberal (but equally arbitrary) thresholds of 0.25 and 0.75 for empirical testing, as opposed to 0.025 and 0.975 (Zhang & Gao, 2023)

2041210x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

or 0 and 1 (Mora et al., 2013). Climate ToE studies have the benefit of long historical time series and future projections (Hawkins & Sutton, 2012; Mora et al., 2013; Zhang & Gao, 2023), allowing them to mitigate the effects of shifting baselines and achieve greater signal to noise ratios than most ecological data. While liberal thresholds seem appropriate in such circumstances, an alternative would be to use more conservative climate ToE thresholds and simulate future time series to project future emergence dates. Though we do not showcase it here, we do include methodology for simulating future years of phenological time series in the *phenometrics* package.

Our community indicators provide a novel means of summarizing and comparing community-level phenological responses. However, comparisons of ToE results from different ecosystems with different data sources may be imperfect, though this is commonplace in the world of biodiversity policy indicators. For example, the Living Planet Index (Collen et al., 2009) summarizes population metrics across a wide range of vertebrate species into single comparable values, despite vast differences across taxa, environment and sampling methodology. Such indicators are imperfect, but informative. To maximize indicator value comparability, we recommend applying the same methodology and threshold values when comparing ecosystems (we do not in this study, but only to showcase both statistical and empirical methodologies). Where methodologies must differ, the validity of comparing ToE results may be further informed by comparing emergence testing results and simulations to the time-series properties of each system. For example, the UKBMS and CPR datasets have similar signal-to-noise ratios despite using different ToE methodologies and data sources, indicating they should (Figure 4, middle) and do (Figure 7, top) have a similar range in years to emergence, justifying comparability.

Lastly, it is important that end users understand that the ToE functions in *phenometrics* detect, but cannot attribute either the causes (e.g. climate change) or effects (e.g. demographic consequences) of phenological change. Relationships between a species' phenology, the environment and other species are complex, and associations between time series detected by ToE metrics should not be interpreted as a causative relationship. Such associations may result from collinearity with an actual driving force, interactions between multiple relevant drivers or random chance. Further analyses are required in order to attribute the causes or effects of phenological change to upstream drivers or downstream ecological effects (e.g. Gonzalez et al., 2023).

4.6 | Package options

Within the *phenometrics* package, we include a range of optional parameters and methodologies to allow end users to best match their research questions and study systems, and to address some limitations and assumptions of the default settings. For example, test and emergence thresholds can be changed, as can test directionality, testing can be applied across a range of static or moving windows and detrending can be performed using a curvilinear model. These, and methods for handling bias and input data uncertainties, are discussed in Supporting Information S2: Section 2.

Furthermore, while the default setting of the ToE metrics in *phenometrics* is to detect advancing spring phenology and increasing temperature of phenological events, this is only one specific use case. By modifying input data and/or arguments of the ToE functions, users may unlock a much greater range of applications. For example, by modifying input data, users could compare year of emergence for differing phenological events (e.g. first flowering vs peak or end flowering), calculate emergence for environmental events (e.g. date of sea ice melt) or change the time period (e.g. 1 week before event timing), summary (e.g. weekly vs daily temperature) or variable of interest (e.g. precipitation, nutrients, etc.) used to calculate ToEE. Our methods can be applied to time series with different replication structures, such as individuals (e.g. individual arrival/departure), dedicated sites (e.g. UKBMS), spatial summaries (e.g. CPR) and so on. Similarly, in our first two examples, we aggregate data across large spatial areas, but could also calculate empirical emergence for individual sites (UKBMS) or grid cells (CPR). We recommend end users consider the most appropriate and interesting formulations for their study systems and research questions.

5 | CONCLUSIONS

The metrics and indicators in the *phenometrics* package provide a suite of methods to answer phenological questions for both single species and whole ecosystems in a unique, simple and broadly applicable manner. We provide quantitative methods to compare species' phenologies to environmental conditions and to the phenologies of other species in their communities. We also develop aggregate indicators to quantify and compare community-wide phenological responses across ecosystems. Our hope is that these may provide a practical methodology to integrate phenological monitoring and indicators into local, regional and global-scale biodiversity-climate policy.

AUTHOR CONTRIBUTIONS

Reid W. Steele contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, software, validation, visualization and writing both the original draft, and review and editing. Marc S. Botham and Ian Middlebrook contributed to data curation, investigation, methodology, project administration, and writing review and editing. Michael Dowd contributed to conceptualization, formal analysis, investigation, methodology, and writing review and editing. Derek P. Tittensor contributed to conceptualization, formal analysis, funding acquisition, investigation, methodology, resources, software, supervision, visualization, and writing review and editing.

ACKNOWLEDGEMENTS

We thank P. Hélaouët for sharing Continuous Plankton Recorder data and providing advice on its use. RWS and DPT acknowledge

2041210x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

funding by the Natural Sciences and Engineering Research Council of Canada (NSERC). DPT also acknowledges funding from the Jarislowsky Foundation. The UK Butterfly Monitoring Scheme is organized and funded by Butterfly Conservation, the UK Centre for Ecology & Hydrology, British Trust for Ornithology and the Joint Nature Conservation Committee. The UKBMS is indebted to all volunteers who contribute data to the scheme.

CONFLICT OF INTEREST STATEMENT

The authors report no conflicts of interest for this study.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210x.70435.

DATA AVAILABILITY STATEMENT

The current version of the phenometrics package is available on GitHub at https://github.com/reidsteele2/phenometrics. Code used in this study is archived at https://doi.org/10.5281/zenodo.23161200 (phenometrics package) (Steele & Tittensor, 2026), and https://doi.org/10.5281/zenodo.23187512 (publication analyses and figures) (Steele, 2026). Data used in this study are available via https://doi.org/10.5285/31cf8abb-892a-48fc-98e9-62312f51a6fb (United Kingdom Butterfly Monitoring Scheme) (Botham et al., 2023), https://www.metoffice.gov.uk/hadobs/hadcet/data/download.html (Met Office Hadley Centre Central England Temperature) (Met Office, 2025), https://doi.org/10.17031/67a0cc543b66b (Continuous Plankton Recorder) (Helaouët, 2024), https://coastwatch.noaa.gov/erddap/griddap/noaacrwsstDaily.html (Coral Watch Sea Surface Temperature) (NOAA Coral Reef Watch, 2025) and https://ourworldindata.org/grapher/date-of-the-peak-cherry-tree-blossom-in-kyoto (Kyoto Cherry Blossom) (Aono, 2025).

ORCID

Reid W. Steele https://orcid.org/0000-0002-0886-7560

REFERENCES

Aono, Y. (2025). *—with major processing by Our World in Data. Day of the year with peak cherry tree blossom in Kyoto, Japan, and 20-year average*. [Dataset]. https://ourworldindata.org/grapher/date-of-the-peak-cherry-tree-blossom-in-kyoto

Aono, Y., & Kazui, K. (2008). Phenological data series of cherry tree flowering in Kyoto, Japan, and its application to reconstruction of springtime temperatures since the 9th century. *International Journal of Climatology*, *28*(7), 905–914. https://doi.org/10.1002/joc.1594

Badeck, F.-W., Bondeau, A., Böttcher, K., Doktor, D., Lucht, W., Schaber, J., & Sitch, S. (2004). Responses of spring phenology to climate change. *New Phytologist*, *162*(2), 295–309. https://doi.org/10.1111/j.1469-8137.2004.01059.x

Botham, M. S., Middlebrook, I., Harrower, C., Conway, R., & Roy, D. B. (2023). *United Kingdom butterfly monitoring scheme: Phenology 2022*. NERC EDS Environmental Information Data Centre. https://doi.org/10.5285/01c38130-d039-4af6-82a0-b32c1290d46a

Butchart, S. H. M., Stattersfield, A. J., Bennun, L. A., Shutes, S. M., Akçakaya, H. R., Baillie, J. E. M., Stuart, S. N., Hilton-Taylor, C., & Mace, G. M. (2004). Measuring global trends in the status of biodiversity: Red list indices for birds. *PLoS Biology*, *2*(12), e383. https://doi.org/10.1371/journal.pbio.0020383

Chadwick, C., Gironás, J., Vicuña, S., & Meza, F. (2019). *Estimating the local time of emergence of climatic variables using an unbiased mapping of GCMs: An application in Semiarid and Mediterranean Chile*. https://doi.org/10.1175/JHM-D-19-0006.1

Collen, B., Loh, J., Whitmee, S., McRAE, L., Amin, R., & Baillie, J. E. M. (2009). Monitoring change in vertebrate abundance: The living planet index. *Conservation Biology*, *23*(2), 317–327. https://doi.org/10.1111/j.1523-1739.2008.01117.x

Cushing, D. H. (1990). Plankton production and year-class strength in fish populations: An update of the match/mismatch hypothesis. In J. H. S. Blaxter & A. J. Southward (Eds.), *Advances in marine biology* (Vol. *26*, pp. 249–293). Academic Press. https://doi.org/10.1016/S0065-2881(08)60202-3

Desmit, X., Nohe, A., Borges, A. V., Prins, T., De Cauwer, K., Lagring, R., Van der Zande, D., & Sabbe, K. (2020). Changes in chlorophyll concentration and phenology in the North Sea in relation to de-eutrophication and sea surface warming. *Limnology and Oceanography*, *65*(4), 828–847. https://doi.org/10.1002/lno.11351

Diamond, S. E., Frame, A. M., Martin, R. A., & Buckley, L. B. (2011). Species' traits predict phenological responses to climate change in butterflies. *Ecology*, *92*(5), 1005–1012. https://doi.org/10.1890/10-1594.1

Edwards, M., & Richardson, A. J. (2004). Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature*, *430*(7002), 881–884. https://doi.org/10.1038/nature02808

Gonzalez, A., Chase, J. M., & O'Connor, M. I. (2023). A framework for the detection and attribution of biodiversity change. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *378*(1881), 20220182. https://doi.org/10.1098/rstb.2022.0182

Greve, W., Prinage, S., Zidowitz, H., Nast, J., & Reiners, F. (2005). On the phenology of North Sea ichthyoplankton. *ICES Journal of Marine Science*, *62*(7), 1216–1223. https://doi.org/10.1016/j.icesjms.2005.03.011

Hawkins, E., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, *39*(1). https://doi.org/10.1029/2011GL050087

Helaouët, P. (2024). *CPR data request—FONI—university Dalhousie—30/04/2024*. [Dataset]. (Marine Biological Association of the United Kingdom). The Archive for Marine Species and Habitats Data (DASSH). https://doi.org/10.17031/67a0cc543b66b

Hočevar, S., Hutchings, J. A., & Kuparinen, A. (2021). Multiple-batch spawning as a bet-hedging strategy in highly stochastic environments: An exploratory analysis of Atlantic cod. *Evolutionary Applications*, *14*(8), 1980–1992. https://doi.org/10.1111/eva.13251

Inouye, D. W. (2022). Climate change and phenology. *WIREs Climate Change*, *13*(3), e764. https://doi.org/10.1002/wcc.764

Keogan, K., Daunt, F., Wanless, S., Phillips, R. A., Walling, C. A., Agnew, P., Ainley, D. G., Anker-Nilssen, T., Ballard, G., Barrett, R. T., Barton, K. J., Bech, C., Becker, P., Berglund, P.-A., Bollache, L., Bond, A. L., Bouwhuis, S., Bradley, R. W., Burr, Z. M., … Lewis, S. (2018). Global phenological insensitivity to shifting ocean temperatures among seabirds. *Nature Climate Change*, *8*(4), 313–318. https://doi.org/10.1038/s41558-018-0115-z

Lindenmayer, D. B., Likens, G. E., Andersen, A., Bowman, D., Bull, C. M., Burns, E., Dickman, C. R., Hoffmann, A. A., Keith, D. A., Liddell, M. J., Lowe, A. J., Metcalfe, D. J., Phinn, S. R., Russell-Smith, J., Thurgate, N., & Wardle, G. M. (2012). Value of long-term ecological studies. *Austral Ecology*, *37*(7), 745–757. https://doi.org/10.1111/j.1442-9993.2011.02351.x

Mackas, D. L., Greve, W., Edwards, M., Chiba, S., Tadokoro, K., Eloire, D., Mazzocchi, M. G., Batten, S., Richardson, A. J., Johnson, C., Head, E., Conversi, A., & Peluso, T. (2012). Changing zooplankton seasonality in a changing ocean: Comparing time series of zooplankton

2041210x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

phenology. *Progress in Oceanography, Global Comparisons of Zooplankton Time Series*, *97–100*, 31–62. https://doi.org/10.1016/j.pocean.2011.11.005

Met Office. (2025). *Hadley Centre Central England Temperature (HadCET) dataset* (Version 2.1.0.0) [Dataset]. https://www.metoffice.gov.uk/hadobs/hadcet/

Møller, A. P., Rubolini, D., & Lehikoinen, E. (2008). Populations of migratory bird species that did not show a phenological response to climate change are declining. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(42), 16195–16200. https://doi.org/10.1073/pnas.0803825105

Mora, C., Frazier, A. G., Longman, R. J., Dacks, R. S., Walton, M. M., Tong, E. J., Sanchez, J. J., Kaiser, L. R., Stender, Y. O., Anderson, J. M., Ambrosino, C. M., Fernandez-Silva, I., Giuseffi, L. M., & Giambelluca, T. W. (2013). The projected timing of climate departure from recent variability. *Nature*, *502*(7470), 183–187. https://doi.org/10.1038/nature12540

NOAA Coral Reef Watch. (2025). *NOAA Coral Reef Watch version 3.1 daily 5 km satellite SST*. NOAA Coral Reef Watch. https://coastwatch.noaa.gov/erddap/griddap/noaacrwsstDaily.html

Poloczanska, E. S., Brown, C. J., Sydeman, W. J., Kiessling, W., Schoeman, D. S., Moore, P. J., Brander, K., Bruno, J. F., Buckley, L. B., Burrows, M. T., Duarte, C. M., Halpern, B. S., Holding, J., Kappel, C. V., O'Connor, M. I., Pandolfi, J. M., Parmesan, C., Schwing, F., Thompson, S. A., & Richardson, A. J. (2013). Global imprint of climate change on marine life. *Nature Climate Change*, *3*(10), 919–925. https://doi.org/10.1038/nclimate1958

Roy, D. B., Oliver, T. H., Botham, M. S., Beckmann, B., Brereton, T., Dennis, R. L. H., Harrower, C., Phillimore, A. B., & Thomas, J. A. (2015). Similarities in butterfly emergence dates among populations suggest local adaptation to climate. *Global Change Biology*, *21*(9), 3313–3322. https://doi.org/10.1111/gcb.12920

Roy, D. B., & Sparks, T. H. (2000). Phenology of British butterflies and climate change. *Global Change Biology*, *6*(4), 407–416. https://doi.org/10.1046/j.1365-2486.2000.00322.x

Steele, R. W. (2026). Future-of-marine-ecosystems/Phenology_Indicators: Publication release. *Zenodo*. https://doi.org/10.5281/zenodo.23187512

Steele, R. W., & Tittensor, D. P. (2026). reidsteele2/phenometrics: Phenometrics publication release. *Zenodo*. https://doi.org/10.5281/zenodo.23161200

Thackeray, S. J., Henrys, P. A., Hemming, D., Bell, J. R., Botham, M. S., Burthe, S., Helaouet, P., Johns, D. G., Jones, I. D., Leech, D. I., Mackay, E. B., Massimino, D., Atkinson, S., Bacon, P. J., Brereton, T. M., Carvalho, L., Clutton-Brock, T. H., Duck, C., Edwards, M., … Wanless, S. (2016). Phenological sensitivity to climate across taxa and trophic levels. *Nature*, *535*(7611), 241–245. https://doi.org/10.1038/nature18608

Visser, M. E., & Both, C. (2005). Shifts in phenology due to global climate change: The need for a yardstick. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1581), 2561–2569. https://doi.org/10.1098/rspb.2005.3356

Visser, M. E., & Gienapp, P. (2019). Evolutionary and demographic consequences of phenological mismatches. *Nature Ecology & Evolution*, *3*(6), 879–885. https://doi.org/10.1038/s41559-019-0880-8

Zhang, M., & Gao, Y. (2023). Time of emergence in climate extremes corresponding to Köppen-Geiger classification. *Weather and Climate Extremes*, *41*, 100593. https://doi.org/10.1016/j.wace.2023.100593

SUPPORTING INFORMATION

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

Table S1: Empirically derived mean spread (s), intraannual standard deviation (σ), intraannual sample size (r) and linear trend (t) values for all species in the United Kingdom Butterfly Monitoring Scheme (UKBMS) and Continuous Plankton Recorder (CPR) datasets, used to inform simulation parameter ranges.

Figure S1: Linear regressions of mean intraannual standard deviation (s), against spread (e) for all species in the United Kingdom Butterfly Monitoring Scheme (UKBMS) and Continuous Plankton Recorder (CPR) datasets. Based on the correlation between s and e in both datasets, we fix s at 1.5 in simulation testing.

Data S1. Phenometrics: An R package for calculating Time of Emergence (ToE) metrics and indicators on phenological data.

How to cite this article: Steele, R. W., Botham, M. S., Dowd, M., Middlebrook, I., & Tittensor, D. P. (2026). A new approach for detecting phenological change using Time of Emergence (ToE) methodology. *Methods in Ecology and Evolution*, *00*, 1–14. https://doi.org/10.1111/2041-210x.70435

2041210x, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210x.70435 by UK Centre For Ecology & Hydrology, Wiley Online Library on [29/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License