Analysis of monitoring data where butterflies fly year-round

Orr Comay1,2,3,4, Oz Ben Yehuda4, Dubi Benyamini5, Rachel Schwartz-Tzachor,6 Israel Pe’er,7 Tal Melochna,5 and Guy Pe’er1,2,3,8

1School of Zoology and the Steinhardt Museum of Natural History, Tel Aviv University, Tel Aviv 6997801 Israel
2German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5c 04103 Leipzig, Germany
3Departments of Ecosystem Services and Economics, UFZ – Helmholtz Centre for Environmental Research, Permoserstrasse 15 04318 Leipzig, Germany
4Achva Academic College, Arugot 7980400 Israel
5Israeli Lepidopterists’ Society, 4 D MicroRobotics, Leventa 91, Bet Arye 7194700 Israel
6Ramat Hanadiv, P.O.B 325, Zikhron Ya’akov 3095202 Israel
7GlueCAD-Biodiversity IT, BMS-IL web-portal, 39 Hanike Street, Hatza, Israel
8University of Leipzig, Leipzig, Sachsen Germany

Citation: Comay, O., O. Ben Yehuda, D. Benyamini, R. Schwartz-Tzachor, I. Pe’er, T. Melochna, and G. Pe’er. 2020. Analysis of monitoring data where butterflies fly year-round. Ecological Applications 30(8): e02196. 10.1002/eap.2196

Abstract. Butterfly Monitoring Schemes (BMSs) engage the public in conservation and provide data sets that cover broad geographical areas over long timescales. Most existing BMSs are in temperate climates; however, the Israeli Butterfly Monitoring Scheme (BMS-IL), established in 2009, is a notable exception as it encompasses a large climatic gradient from Euro-Siberian through Mediterranean to hyper-arid regions. Israel’s climate poses challenges in analyzing data from year-round butterfly activity, as in other tropical or arid countries. The Regional Generalized Additive Model (Regional GAM) is a butterfly phenology and abundance model based on repeat visits throughout species’ flight season. We tested the applicability of Regional GAM for species with complex flight seasonality (e.g., multivoltine) by comparing estimated abundance and seasonal indices for the full data set and rarefied subsets. We assessed the reliability of modeled flight seasons and compared abundance estimates per site resulting from biologically plausible and unreliable seasonal models. The reliability of Regional GAM rises with the number of observations, and the model tends to produce more biologically plausible models for species with simple phenologies (e.g., univoltine with a single peak in activity). Abundance estimates based on unreliable models produce values with inter-quantile ranges of 90%–153% compared with biologically plausible models, while peak time changes with an interquartile range of 0–22.5 d when comparing all rarefied models with the full data set. Regional GAM should be applied with great caution for rare species and those with a complex flight season, and the date of year start needs to be carefully chosen for species that are active year-round. We identified the key sources of error and propose an operational workflow to address them. With few adaptations, Regional GAM can support new BMSs in analyzing data where butterflies are active year-round, including tropical climates. We propose guidelines for analyzing BMS data for species or regions with long activity periods and complex phenologies.

Key words: abundance; butterflies; citizen science; modeling; phenology; warm climate.

INTRODUCTION

Butterflies and the expansion of Butterfly Monitoring Schemes (BMSs)

Evidence for the severity of insect decline has become a focus of public attention (Hallmann et al. 2017, Sánchez-Bayo and Wyckhuys 2019); however, there is a severe shortage of insect biodiversity data from large parts of the globe, particularly Africa, South America, and southern Asia (Collen et al. 2008, Sánchez-Bayo and Wyckhuys 2019). Citizen Science (volunteer-based) monitoring schemes may offer a cost-effective way of addressing this gap. Among insects, the broadest network of volunteer-based monitoring focuses on butterflies (Pe’er and Settele 2008, Van Swaay et al. 2015). However, most BMSs are located in fully humid temperate climates (categories Cfa-Cfc in the Köppen-Geiger climate classification; Kottek et al. 2006, Rubel et al. 2017), particularly in Europe, the United States, and Canada; relatively less monitoring and research have focused on other climatic areas such as Mediterranean, arid, and tropical climates (i.e., Köppen-Geiger climate categories Csa, B, and A, respectively). Addressing
biodiversity knowledge gaps in such regions is particularly important given that arid and tropical areas cover 33.3% and 22.2% of Earth’s habitable terrestrial surface, respectively (i.e., excluding polar frost and tundra regions, Köppen-Geiger climate categories EF and ET). Recently, butterfly monitoring schemes have been initiated in Texas, USA, Florida, USA (Ries et al. 2019), and Lebanon (Zorkot 2019). However, monitoring in hot climates, especially if done with volunteers, presents challenges that must be identified and addressed. Here we report on the key challenges identified when analyzing data from the Israeli Butterfly Monitoring Scheme (BMS-IL), established by the Israeli Lepidopterists Society in April 2009; BMS-IL marks the first BMS operating where adult butterflies occur year-round. The scheme was founded based on pre-existing knowledge and experts in Israel, and not as the result of a global strategic decision.

**Barriers to producing biologically plausible phenology curves**

The methodological challenge most relevant to BMSs in regions where species are active over long seasons is the strong relation between a species phenology curve (i.e., the period when adults are on the wing) and the estimated relative abundance. When no observations are conducted but butterflies are active (e.g., due to unavailability of the observer), one has to account for the missing count data to assess species abundance. This can be done using a phenology model, where the timings of seasonal peaks and lows are estimated (i.e., if a count was missing during peak season, it would have been higher than if a count was missing toward the end of the flight season). However, without a biologically plausible phenology curve (e.g., because of small or highly variable sample size), model estimates of annual abundance cannot be considered reliable, especially without proper assessment of potential bias and its sources. Otherwise, long-term abundance trends (e.g., the species is becoming rarer) could be confused with seasonal activity trends (e.g., the species starts to fly earlier in the year). Hence, it is common practice to omit species with complex phenology curves (e.g., multivoltine, defined as having more than a single generation per annum) from the analysis. In hot (Mediterranean, arid, or tropical) climates many species are multivoltine or even active year-round, and thus focusing on univoltine or bivoltine species alone would disregard most butterfly species. For example, multivoltine species can produce different numbers of generations in different areas and different years (Benyamini 2010). Therefore, simplifying assumptions about the shape of the curve cannot be made; instead, one has to ensure the curve itself is correct.

Achieving trust in phenology and abundance estimations is therefore a key task for analyzing data from any BMS where most species have complex phenologies (such as migratory species) or are active year-round. To achieve this, it is necessary to (1) identify the impacts of changes in observation frequency over the year; (2) define indicators and methods to distinguish between biologically plausible and unreliable curves; and (3) quantify the impacts of data deficiency (Schmucki et al. 2016) or typical errors in terms of phenology and abundance estimations.

Butterfly monitoring in Israel comes with a set of challenges. First, climatic and biogeographic divergence means that the country supports species that substantially differ from each other in phenology; some are active over winter or spring, many more from early spring to late autumn, while others are active year-round (Benyamini 2010). Moreover, the phenology of many species differs or even diverges across the country. For example, some species are active year-round in the desert but seasonally in the Mediterranean region (e.g., *Deu dorix livia*, *Colotis fausta*; Benyamini 2010). Additionally, the majority of Israel’s species are multivoltine, and many produce different numbers of generations in different areas and different years. Finally, the climate is highly variable, both within and between years, and this phenomenon, which is typical of arid areas, is met by species’ capacities to perform long-term (up to 15 yr) pupal diapause (e.g., Benyamini 1999, 2008). This results in strong, often enigmatic annual fluctuations in the abundance of some species.

Second, monitoring is hampered by inhospitable summer conditions, with daytime temperatures exceeding 30°C throughout the country and over 40°C in hyper-arid regions (Israel Meteorological Service 2013). This impacts on monitoring protocols; while volunteers are encouraged to continue performing observations, there is a decline in observation frequency, which can affect phenology and abundance estimations for species that are particularly active in summer (e.g., *Colotis fausta*) or those performing aestivation (i.e., lower activity during the summer, such as *Maniola teltmessia*) between spring and autumn (Benyamini 2010).

Third, for species that are active year-round, it is methodologically essential to define the ‘monitoring year’ (start and end), and identify if this can yield a bias. Using the BMS-IL as a case study, we examine these challenges and develop a workflow for overcoming them to produce biologically plausible flight curves and abundance values.

**Methods**

**Description of BMS-IL**

BMS-IL was established in 2009 by members of the Israeli Lepidopterists Society according to the Israeli protocol for butterfly monitoring (Schwartz-Tzachor et al. 2009); this protocol is based on Pollard (1977), in which volunteers mark a 300–500 m (and up to 1 km) transect, and then visit it regularly. The Dutch adjustment of the protocol, adopted by all BMSs since the...
1990s, divides the transects into 50-m sections, with all sections in the same habitat (to the extent possible). Observers walk at a slow and regular pace and count all butterflies observed within an imaginary $5 \times 5 \times 5$ m cube (2.5 m to the right and left of the observer, 5 m in front of the observer, and 5 m from the ground). Walks are not conducted during cold (less than 13°C), windy, cloudy, or rainy weather. Transect sections are chosen so that land cover change is minimal within each transect.

Voucher specimens are not collected, but observers are encouraged to capture specimens for observation, photography, and identification before releasing them. While European and North American BMS require weekly visits, in Israel volunteers are asked to visit transects twice a month from the beginning of October until the end of June (Schwartz-Tzachor et al. 2009). This monitoring year was defined (1) to cover the rainy season (summer drought normally lasts from June to September), (2) to avoid summer heat stress for volunteers, and (3) to be in line with cultural-economic activity in the country (i.e., school summer holidays in July and August, and during Jewish holidays, for which dates vary annually between September and October). While July–September is officially considered as the “summer break,” volunteers are encouraged to continue their observations, even if at lower frequency. Most transects are concentrated in the Mediterranean region of the country (more than 350 mm rainfall; Fig. 1), where most of Israel’s population resides.

BMS-IL had 12 volunteers in its first year, increasing to 182 volunteers in October 2019, considerably increasing the number of active transects (Fig. 2). An internet portal, launched during 2012, provides volunteers with standard interface forms to streamline observation data into an SQL-based database; these data were last uploaded to Global Biodiversity Information Facility (GBIF) in 2016 (Peer 2014). As of October 2019, there were 128 active transects (visited at least three times per year, although all observations are used in the analysis; see the “Study design” section). Additional elements of BMS-IL whose data are not used in the present study include (1) yearly surveys of rare species, (2) collection of data from sporadic observations (i.e., observations taken without a protocol detailing sampling method), (3) the “Big Butterfly Count” (established in April 2019), which focuses on outreach to the wider population through media and schools using a simple yet rigorous count protocol, and (4) a set of internet tools to support data collection and management, including a web portal, smartphone apps, etc.

A unique feature of BMS-IL is the establishment of volunteer communities (with the first established in 2014), in which volunteers in a defined geographical area train together to record butterflies, share experiences online and meet on a regular basis. Each volunteer community is assisted by a scientific regional mentor and a coordinator, whose role is to assist in establishing transects, address volunteers’ questions and organize activities. A national volunteer coordinator manages the activities of the community coordinators and maintains a connection with volunteers who are not grouped in communities.

As of October 2019, BMS-IL has one full-time national volunteer coordinator, five regional scientific mentors (butterfly experts assisting volunteers in identifying species in the field), nine community coordinators, and nine active communities. Active efforts are constantly made to train new volunteers (Fig. 2). Data are validated by a process in which observations that strongly deviate from expected spatial distributions and seasonal patterns are marked, discussed with the reporting observer and eventually either confirmed or edited.
To examine how sampling design (monitoring season and number of visits) affects the resulting adult flight curves and abundances, it is essential to concentrate on species with sufficient data and whose phenology is well known. The methodological insights (e.g., the number of visits sufficient to produce a reliable phenological curve) gained from such species can be later applied to other, less well-known species. Overall, 84 species were recorded in the low elevation (<700 m) Mediterranean transects of BMS-IL, 20 of which were expected to have simple phenologies while 64 were expected to have complex phenologies (i.e., multivoltine species, migrating species, where adults arrive into the country from abroad, and aestivating species; Benyamini 2010).

The Regional Generalized Additive Model (Regional GAM; Schmucki et al. 2016) is a phenological model based on butterfly observations collected using a set protocol, including repeat visits to the same sites. Regional GAM assumes that all transects share the same phenology; on this basis, in this study, we analyzed only data from Mediterranean regions with upper elevations of no more than 700 m above sea level (asl; Fig. 1); the data were collected between 2010 and 2018. The Regional GAM output is a model of species’ relative abundance of adults throughout the year, termed here as the “flight curve” because it depicts the season of adult flight in a species life cycle. Each date is given a fractional value, so that over the entire year the sum of values is equal to one. Dates in which no adults are expected are given a value of zero.

To examine how sampling design (monitoring season and number of visits) affects the resulting adult flight curves and abundances, it is essential to concentrate on species with sufficient data and whose phenology is well known. The methodological insights (e.g., the number of visits sufficient to produce a reliable phenological curve) gained from such species can be later applied to other, less well-known species. Overall, 84 species were recorded in the low elevation (<700 m) Mediterranean transects of BMS-IL, 20 of which were expected to have simple phenologies while 64 were expected to have complex phenologies (i.e., multivoltine species, migrating species, where adults arrive into the country from abroad, and aestivating species; Benyamini 2010).

We selected 12 species with complex phenologies (Benyamini 2010) and 3 species with a simple phenology (i.e., univoltine with a single peak in activity; Table 1). We fitted a Regional GAM model (Schmucki et al. 2016) independently for each species–year combination (allowing the phenology curve to vary in shape between years) first using the full data set and then using four rarefactions: (1) omitting summer visits (i.e., omitting all data from visits conducted from July to September), and randomly selecting (2) a set of 150 visits, (3) a set of 100 visits, and (4) a set of 50 visits for each monitoring year, with a visit defined as a unique combination of site and date. As the number of visits per annum was low (especially in the early years of BMS-IL; Fig. 2) we could not take different sets of 150 visits for all years. Therefore, we randomly selected a single set of visits for each rarefaction protocol (with the exception of omitting all summer observations, as this was not a random selection). The same rarefied data sets were used for all species studied (e.g., the same 150 visits in 2017, the same 50 visits in 2012, and the same 100 visits in 2015, etc.).

The regional GAM also allows the user to define the starting point of the analysis year, in order to ensure that the entire phenology is encompassed in the analysis. Thus, we tested the impact of defining the monitoring year as starting either on 1 October (considering the official start of the monitoring year and approximating the beginning of the rainy season, following a 4–5 month long drought) or on 1 January. The former definition may be useful for winter-active species (e.g., Archon apollinus, Euchloe belemia; Benyamini 2010), which first appear in November/December and continue activity into March or later, while the latter definition is more meaningful for spring, summer or aestivating species.

The resulting flight curves were first estimated on the basis of published literature on all species in Israel (Benyamini 2010) and extensive personal experience in the field. At this stage, the aim was to identify clearly erroneous flight curves even without prior knowledge of species’ natural history (categories 1 and 2; see after this paragraph), as well as those whose plausibility could only be ascertained based on such prior knowledge (categories 3 and 4). This stage of analysis was needed because the model can estimate abundance based on any phenology curve, regardless of its reliability, and to assess how often prior natural history knowledge is required to determine the reliability of phenology curves and abundance estimates generated by Regional GAM. Based on this stage of curve-assessment/validation, flight curves were classified into one of four categories:

(1). None: model produced no curve at all.
(2). Wrong: curve clearly ecologically implausible (e.g., completely linear/flat or increasing/decreasing monotonically throughout the year).
(3). Omit: model produced curve but overlapped none of the data.
(4). Omit: model produced curve that overlapped less than 50% of the data.
(5). Omit: model produced curve that overlapped more than 50% of the data.
(6). Wrong: model produced curve that overlapped 100% of the data.

FIG. 2. Number of active transects in the Israeli Butterfly Monitoring Scheme, as of October 2018. Active transects are those that were monitored at least three times in a given monitoring year (starting 1 October).
Unreliable: curve indicated low or declining abundances in a period of the year where the species typically peaks (or vice versa), or if the flight curve was too simple compared with known seasonality of the species (e.g., no aestivation in a species known to aestivate). We also noted when start-date effects could be seen (i.e., local maxima at the beginning or the end of the assumed year).

Biologically plausible: no direct peculiarities with the period of activity or time of peak or decline.

Quantitative assessment of phenology curves

To obtain quantitative indications of the reliability or stability of the curves and consequent abundance indices under rarefaction, curves were only compared within species and we inspected the following variables (indices): (1) number of peaks, (2) seasonal timing (month and day) of the highest seasonal peak, (3) seasonal timing of the start and end of flight season, defined as the earliest and latest dates of the year when relative abundance is greater than zero, and (4) abundance index.

Variables 1–3 were compared for each butterfly species, each year, and each monitoring year start date. Since our data include five different data sets (full data set, without summer visits, and 150, 100, and 50 randomly selected visits) and each species has two curves per year (beginning of the monitoring year on 1 October or 1 January), each species has up to 10 curves per year (fewer if some combinations did not produce a curve). For example, for Anthocharis cardamines, the regional GAM calculation resulted in seven phenology curves for the year 2011: three when setting the beginning of the year to January (one using the full data set, one using the data set without summer visits, and one using 150 randomly selected visits per year) and four using the same data sets but setting the beginning of the year to October (this time also successfully producing a curve with 100 randomly selected visits). Three other phenology curves could not be calculated: two when setting the beginning of the year to January (using 100 and 50 randomly selected visits) and one when setting the beginning of the year to October and using only 50 randomly selected visits. To evaluate the impacts on the abundance index, we used only abundance indices of sites in which species occurred under all rarefaction schemes; this approach was chosen because impacts on abundance index were calculated on a site by site basis (e.g., species’ abundance index of a given site using the full data set was compared with the species’ abundance index for the same site when all summer observations were removed). Consequently, one cannot compute the relative change in abundance index if the site did not appear in all data sets. For example, Papilio machaon occurred in the full data set for sites A and B, but only occurred for site A in the rarefied data sets, and so site B was removed from the analysis of this species. This process resulted in species being linked with 17–70 sites in total. The number of sites per species in a given monitoring year varied from 1

| Species | Phenology | Total no. sites | Minimum | Maximum |
|---------|-----------|----------------|---------|---------|
| Family: Papilionidae | | | | |
| Papilio machaon | multivoltine | 49 | 10 | 31 |
| Archon apollinus | simple univoltine | 27 | 7 | 19 |
| Family: Pieridae | | | | |
| Pieris brassicae | multivoltine | 66 | 15 | 46 |
| Pieris rapae | multivoltine | 70 | 17 | 54 |
| Pontia daplidice | migrating | 59 | 13 | 40 |
| Colotis fausta | over-summering | 47 | 9 | 31 |
| Anthocharis cardamines | simple univoltine | 49 | 11 | 33 |
| Colias croesus | migrating, multivoltine | 61 | 8 | 45 |
| Gonepteryx cleopatra | aestivating, multivoltine | 50 | 12 | 31 |
| Family: Nymphalidae | | | | |
| Vanessa atalanta | migrating, aestivating | 33 | 7 | 19 |
| Vanessa cardui | migrating, multivoltine | 66 | 13 | 48 |
| Family: Satyridae | | | | |
| Melanargia titea | simple univoltine | 31 | 8 | 16 |
| Hipparchia fatau | aestivating | 17 | 4 | 13 |
| Maniola telmessia | aestivating | 52 | 10 | 31 |
| Family: Lycaenidae | | | | |
| Leptotes piritous | multivoltine | 23 | 4 | 11 |

Notes: Species phenology was taken from Benyamini (2010). Minimum and maximum numbers of sites of occurrence are those found under all rarefaction schemes. Univoltine species demonstrating no phenology complications (e.g., migration) were considered to have a simple phenology. Total number of sites relate to the full data set (2009–2018).
to 48. The variance in the number of visits per site (among sites used to assess relative deviation in abundance estimate, as detailed above) was 114.8, 68.1, 95.2, 72.4, 57.2, 81.4, 110.7, and 153.1 visits in the years from 2010 to 2017, respectively. To illustrate the results, we plotted the quantitative seasonal indices against the number of visits to sites where species occurred at least once in the respective data set.

Quantifying the impact of unreliable phenology curves on abundance estimation

Abundance index.—The abundance index was computed as the rounded sum of the observed and expected counts of a given species at a given site, if a count would have been made there each Wednesday. In other words, the abundance index answers the question “How many butterflies of species X would have been observed at site Y if the site was visited every week?” For example, hypothetically consider the case of Melanargia titea, a univoltine species occurring from April to June in the Mediterranean region of Israel (Benyamini 2010). All counts of the species at site A before April and after June are zero. Counts of 1, 23, 52, 27, and 3 butterflies are recorded on 12 April, 26 April, 10 May, 24 May, and 7 June, respectively, for a total of 106 butterflies. The model imputes expected counts of 12, 34, 41, and 17 butterflies on 19 April, 3 May, 17 May, and 31 May, when no actual counts were performed. Summing up the observed and expected counts gives an abundance index of 210 Melanargia titea butterflies at site A. A detailed explanation of the statistical methods used to calculate the abundance index is given in the “Statistical analyses” section. Note that the abundance index is not a direct estimate of population size, as some individuals are likely to be observed and counted more than once, either during the same visit or owing to repeated observations of the same individuals of a territorial species (such as Vanessa atalanta or Charaxes jasius) at a later visit to the same section of a transect (Benyamini 2010). Instead, it is a relative index measuring a combination of population size, activity, and detectability. It is comparable between years, and to a lesser extent also between sites (because detectability may differ based on vegetation structure and observers’ skill) or between species (since size, color and movement behavior affect detectability; Isaac et al. 2011).

When the results indicate that the phenology is unreliable, we need to estimate the impact on abundance estimation. In this study, we divided the abundance index (of a given species, year and site) by the equivalent abundance index when the phenology was quantified as biologically plausible. The quotient was defined as the “multiplier of abundance.” Consider the hypothetical example of Melanargia titea given above: if an unreliable flight curve would lead to an abundance index of 168 butterflies instead of 210, the multiplier of abundance would be $0.8 (168/210)$, reflecting a $20\%$ underestimation of the species abundance at site A.

We calculated multipliers of abundance for species–site–year combinations for which at least one data set (the full data set or any rarefaction) produced a biologically plausible curve and at least one other data set produced an unreliable curve. When more than one data set produced a biologically plausible curve, the denominator was that produced based on the greatest amount of data (e.g., if biologically plausible curves were produced based on the full data set and when summer visits were removed, we used the abundance index calculated based on the full data set). As flight curves were based on data sets of multiple sites, and one multiplier of abundance was calculated per site, various multipliers of abundance were produced per species. We compared the distribution of the multipliers of abundance for species with simple and complex phenologies, and tested whether the multipliers of abundance showed directional trends (i.e., a tendency to underestimate when phenology was unreliable). These calculations were performed for curves based on a monitoring year of January to January, to avoid spurious increases in sample size.

Quantifying impacts of rarefaction on seasonal indicators.—Within each species, we compared the seasonal timings of the start, peak and end of the flight season between flight curves calculated based on the full data set and those based on the rarefied data sets. Differences in the date of peak abundance were calculated for all studied species, while differences in season start and end dates were only calculated for species not active year-round, as year-round active species (e.g., Vanessa cardui) have no unequivocally defined seasons. To calculate the descriptive statistics of these differences, we used the absolute value of the difference rather than the relative value (i.e., negative numbers for earlier dates and positive numbers for later dates). For example, if when using the full data set a species’ season start was calculated as 15 February, but when only 150 or 100 visits per annum were used the season start was calculated as 14 or 16 February, the difference was calculated as $+1$ d in both cases, and not as $-1$ d and $+1$ d, because this would have led to a mean difference of zero days. This resulted in a value range of whole numbers from zero (no change in peak date) to 364 (peak date almost a year later/sooner than the full data set). Note that a 364 d shift could occur if, for instance, the peak occurred at one edge of the curve in one case or the other end under rarefaction. For this calculation, we used the January to January year definition for all species except Archon apollinus; both the literature (Benyamini 2010) and our preliminary results indicate that this is the only species in our analysis that peaks in winter.

Statistical analyses

Analyses were performed using R version 3.6.0 (R Core Team 2019) through RStudio version 1.1.453 (RStudio Team 2016), as well as the following R
We used a Generalized Additive Model (GAM) to concomitantly extract phenology and abundance while imputing for missing observations (Dennis et al. 2013). This approach has been tested and extended to a Regional GAM approach where sites are clustered as long as they can be assumed to share the same bioclimatic region (Schmucki et al. 2016). We used the Regional GAM where inputs are dates, sites, and counts of each species (including zero counts for dates in which a visit was made to a site but no adults of that species were observed), and the output is the estimated relative detectability of the species for each day of the monitoring year, summed up to 1 for each entire monitoring year. Regional GAM explicitly assumes that all input sites exhibit the same phenology. Accordingly, for the analyses presented in this paper, we used only low-elevation sites (<700 m asl; Fig. 1) in regions with a Mediterranean climate (i.e., more than 350 mm rainfall per year).

We utilized functions included in the R package rbms (version 0.1.1; Schmucki et al. 2016, Schmucki 2018), with minor code modifications necessary to implement coding loops rather than analyzing each species separately. The modified functions are available in Data S1. We used the package’s default values for the minimum number of visits per site (n = 3), minimum number of sites (n = 1), and minimum number of occurrences per site (n = 2) when fitting a phenology curve. These values had a direct impact on the number of cases when no flight curve would be produced (rather than producing potentially erroneous curves when data were too scarce).

A species was considered as active at a given date if its relative abundance (as calculated by the Regional GAM) was greater than zero. Season start and end dates were set as the earliest and latest dates in the year in which the species was active according to the produced curve. The highest seasonal peak date was the date of relative abundance (as calculated by the Regional GAM) was used as an offset (i.e., a predictor with a predetermined coefficient value; in our case, it was set to 1). The resulting model was used to predict (impute) the expected number of individuals that would have been seen for all site–date combinations. Next, a table of number of butterflies for all species-site-date combinations was created using observed counts where available and model predictions otherwise. In cases where species were not observed in a given site-year combination, or when a phenology curve was not produced, we did not assign a predicted value.

We tested whether species with a simple phenology are more likely to have a biologically plausible phenology compared with those with a complex phenology. To this end, we fitted a binomial generalized linear model (GLM) with a logit link function, using species identity, numbers of occurrences and visits (counting only visits to sites in which the species occurred at least once in that data set), definition of the year start (January or October) and year (in the case of a year starting in October, based on the calendar year of the starting October) as predictors, and the curve reliability (whether it was assessed as biologically plausible according to the criteria that we outlined) as the response variable. Species identity and year definition were also tested in interaction. We omitted nonsignificant predictors and refitted the model without them.

Quantitative indices of phenology curves (season start, peak and end dates, number of active days, and number of peaks per curve) were plotted both as points (one point for each phenology curve) and as trend curves using local polynomial regression fitting through the loess function in the R stats package. This method fits a value for each X value (X being the number of visits to sites where the species occurred at least once) based on observed values weighted by their distance from X.

We modeled the deviations (in absolute number of days) of the quantitative indices of the rarefied curves from the curves produced using the full data set as the response variable in a negative binomial generalized linear mixed model (GLMM). In this model, the non-random predictors were the species, number of visits, and the index type (deviations in start, peak, and end of season dates, as well as season lengths, all expressed in days and thus potentially ranging from 0 to 364 d), while the year was a random predictor. This model design allowed us to assess which index tended to produce greater deviations from the full data set, while accounting for the effects of species, number of visits and year.

To assess the impact of obtaining an unreliable phenology on abundance estimates, we calculated the quantiles of the multipliers of abundance. We used permutation ANOVA with 999 permutations to test if the mean multiplier of abundance between species with
complex phenologies was significantly different from that of species with simple phenologies.

**RESULTS**

A data set describing the results of 1,200 attempts to fit Regional GAM to data is given in Data S2. Appendix S1: Fig. S1 depicts the evenness of dates retained under each rarefaction scheme. Appendix S1: Fig. S2 depict the computed start, highest peak and end of the flight season by number of visits in the Regional GAM data sets (rarified and complete) per species. Appendix S1: Fig. S3 demonstrate how the relative abundance of species varies within each site under rarefaction.

**Impacts of defining monitoring year start date**

In some cases, the monitoring year definition had a high impact on the resulting phenology curve. For example, *Vanessa cardui* flight curves often peak at the beginning or end of the predefined year if the starting date is set at 1 January but not when setting the year start is 1 October. This phenomenon, termed the “start-date effect,” is exemplified in curves for the year 2015 (Fig. 3) when modeling using the full data set. *Vanessa cardui*’s

![Phenology curves created by the Regional Generalized Additive Model (Regional GAM) model of (a) Vanessa cardui and (b) Maniola telmessia during the year 2015, using the full data set for the period and based on two different monitoring year start definitions: January (red) and October (cyan). Abundance is relative to a given data set (year and year definition combination), and hence is not directly comparable between months preceding and following October 2015 when defining the year start at October. (a) When creating the phenology curve using the January year start definition, the phenology curve rises rapidly (almost six-fold) during December 2015. When modeling based on the October year start definition, the rise throughout December 2015 is considerably slower (about 1.51-fold). (b) Aestivation is visible as lowered relative abundance during the peak summer months when using the full data set. However, when the summer (July–September) observations are removed from the data set, aestivation is no longer apparent. When the year start is defined in October, no gradual return to activity following aestivation is apparent, even when all observations are used.](image-url)
relative abundance was multiplied by 5.97 during December 2015 when the year start was defined as 1 January, but only by 1.51 when the monitoring year start was defined as 1 October (Fig. 3a). On the other hand, in September 2015, we found an abundance rise if the year was defined as starting in January, compared with a small decline if the analysis-year started in October. This difference is probably due to observations taken during October 2015, which were included in the data set for the model with a monitoring year start in January, but not in the data set of the model with the year start in October. These examples demonstrate that start-date effects are impossible to avoid for species that are active year-round. For *Papilio machaon*, the impact on curve reliability was significant (Table 2).

In other cases, both year definition and selective removal of summer observations qualitatively impacted the resulting phenology curve. For example, the aestivating species *Maniola telmessia* produced the anticipated phenology curve for summer 2015 when the monitoring year start was set to January and the full database was used (red unbroken line in Fig. 3b), namely, a high peak in May, decrease in June–July, and a return to activity in autumn (Benjamini 2010). The aestivating phenology

| Predictor                                                                 | Estimated coefficient ± SE | \( P \)   |
|---------------------------------------------------------------------------|----------------------------|-----------|
| *Anthocharis cardamines* (simple phenology)                               | 2.48 ± 1.06                | 1.92 × 10⁻² |
| *Archon apollinus* (simple phenology)                                     | 0.97 ± 0.56                | 0.09      |
| *Colias croeus* (complex phenology)                                       | −0.77 ± 0.45               | 0.09      |
| *Colotis fausta* (complex phenology)                                      | −1.14 ± 0.47               | 1.48 × 10⁻² |
| *Gonepteryx cleopatra* (complex phenology)                                | −0.71 ± 0.46               | 0.12      |
| *Hipparchia fatua* (complex phenology)                                    | −0.90 ± 0.55               | 0.11      |
| *Leptotes piritheus* (complex phenology)                                  | 0.86 ± 0.57                | 0.13      |
| *Maniola telmessia* (complex phenology)                                   | −0.52 ± 0.47               | 0.27      |
| *Melanargia titea* (simple phenology)                                     | 1.00 ± 0.56                | 0.07      |
| *Papilio machaon* (complex phenology)                                     | 0.66 ± 0.55                | 0.23      |
| *Pieris brassicae* (complex phenology)                                    | −1.36 ± 0.48               | 4.88 × 10⁻³ |
| *Pieris rapae* (complex phenology)                                        | −0.60 ± 0.56               | 0.28      |
| *Pontia daplidice* (complex phenology)                                    | 0.77 ± 0.56                | 0.17      |
| *Vanessa atalanta* (complex phenology)                                    | −0.55 ± 0.58               | 0.35      |
| *Vanessa cardui* (complex phenology)                                      | −2.85 ± 0.55               | <1 × 10⁻³  |
| Year start at October                                                     | −0.05 ± 1.46               | 0.97      |
| Number of occurrences                                                     | 0.04 ± 4.94 × 10⁻³         | <10⁻³     |
| Year 2011                                                                 | 0.04 ± 0.32                | 0.90      |
| Year 2012                                                                 | −0.02 ± 0.31               | 0.95      |
| Year 2013                                                                 | 0.16 ± 0.33                | 0.62      |
| Year 2014                                                                 | −0.64 ± 0.36               | 0.08      |
| Year 2015                                                                 | −1.03 ± 0.37               | 5.60 × 10⁻³ |
| Year 2016                                                                 | −0.93 ± 0.35               | 7.84 × 10⁻³ |
| Year 2017                                                                 | −0.52 ± 0.37               | 0.16      |
| *Archon apollinus* × Year start at October                                | −0.50 ± 1.62               | 0.76      |
| *Colias croeus* × Year start at October                                   | −1.34 ± 1.59               | 0.40      |
| *Colotis fausta* × Year start at October                                 | −0.08 ± 1.57               | 0.96      |
| *Gonepteryx cleopatra* × Year start at October                            | −0.27 ± 1.56               | 0.86      |
| *Hipparchia fatua* × Year start at October                               | −1.02 ± 1.64               | 0.53      |
| *Leptotes piritheus* × Year start at October                             | −2.65 ± 1.65               | 0.11      |
| *Maniola telmessia* × Year start at October                              | −2.18 ± 1.63               | 0.18      |
| *Melanargia titea* × Year start at October                               | 0.41 ± 1.68                | 0.81      |
| *Papilio machaon* × Year start at October                                | −3.43 ± 1.63               | 3.51 × 10⁻² |
| *Pieris brassicae* × Year start at October                               | 0.24 ± 1.57                | 0.88      |
| *Pieris rapae* × Year start at October                                   | −0.28 ± 1.60               | 0.86      |
| *Pontia daplidice* × Year start at October                               | 2.17 ± 1.60                | 0.18      |
| *Vanessa atalanta* × Year start at October                               | 0.96 ± 1.66                | 0.56      |
| *Vanessa cardui* × Year start at October                                 | 1.46 ± 1.58                | 0.36      |

Notes: Curve reliability was defined as the binary response variable (biologically plausible/implausible, without curves that were not produced; \( N = 793 \)), and hence binomial distribution was assumed for the model. Interaction terms between the monitoring year definition and species (lower rows) are given by \( \times \) following the species name. Significant (\( z = 0.05 \)) predictors are shown in boldface type. Significant positive coefficients imply greater tendency to obtain a biologically plausible phenology curve and vice versa.
pattern disappeared when defining the monitoring year as starting in October (cyan unbroken line in Fig. 3b), as well as when removing summer month visits to mimic a full summer break by volunteers (Fig. 3b).

Curve reliability

Of 1,200 model runs (potential phenology curves), in 407 cases, the model did not produce a curve. An expert-based assessment of the former curves led to classifying 454 curves as biologically plausible, 272 as unreliable, and 67 as wrong. The proportion of unreliable and wrong curves, as well as failure to produce a curve, was much higher for species with complex phenology and increased with the volume of data lost by rarefaction (Fig. 4). When assessing the factors determining curve reliability (for the 793 curves that could be produced), species with simple phenologies (Anthocharis cardamines, Archon apollinus, and Melanargia titea) had a greater tendency to produce biologically plausible phenotype curves, even when accounting for the number of occurrences and definition of the start of the year (Table 2); however, only in the case of Anthocharis cardamines was this tendency significant (α = 0.05). Two years in our data set (2015 and 2016) produced notably fewer biologically plausible curves, potentially indicating unusual phenology in all species studied.

Figs. 5 and 6 depict how the season start, peak, and end dates (Fig. 5), flight season duration in days (Fig. 6a), and number of peaks (Fig. 6b) of six species’ adult flight season varied with decreasing number of visits used to construct the phenology curve. Appendix S1: Figs. S4–S5 depict the same for all 15 species studied. Note that the year definition had a spurious impact on the local polynomial regression fitting of Archon apollinus’ date of highest abundance (Fig. 5), as in some curves its peak dates occurred in late December (high values on the Y-axis) and in others from January to March (low values on the Y-axis): both cases are consistent and indicate a winter peak for this species.

For most species, quantitative phenology curve indices stabilized when the number of visits used to construct the phenology curve was high (>250–300 visits to sites where the species occurred at least once; Appendix S1: Figs. S4–S5), suggesting reliable results when ample data are available. However, many curves were not produced at all when the number of visits was less than 100 (indicated by zero peaks, Fig. 6c). Moreover, for three aestivating species (Hipparchia fatua, Gonepteryx cleopatra, and Maniola telmessia), phenology curves based on more than 200 visits almost always showed two peaks (i.e., aestivation was identified), but rarely so when there were fewer than 100 visits. Flight season duration was often in line with expectations based on expert observation.
knowledge when the number of visits was over 200, but was sometimes grossly over-estimated (e.g., *Anthocharis cardamines* active for 8–12 months per annum) or under-estimated (e.g., *Colias croceus* active only for 7–8 months per annum) when the number of visits was less than 200 (Fig. 6a). Overall, curves based on more than ~200 visits/yr were generally reliable and were not considerably changed by additional visits (Appendix S1: Figs. S4–S5). For species with a simple phenology, 16 occurrences were very often sufficient for a reliable phenology, while for species with a complex phenology at least 32 occurrences were often required to obtain a reliable flight curve (Fig. 4).

**Impacts of rarefaction on seasonal indicators**

Of the 1,200 curves studied, 288 were used to quantify differences in the season peak date between curves based on the full data set and those based on rarefied data sets. Curves not used for this analysis were those based on the full data set, rarefied data sets that did not produce a curve, and curves based on a monitoring year start set to October for all species but *Archon apollinus* (for whom curves based on a monitoring year start set to January were omitted). Of those curves, 93 were used to describe deviations in season start and end dates, as these indicators only make sense for species that are not active year-round (i.e., *Anthocharis cardamines*, *Archon apollinus*, *Melanargia titea*, *Gonepteryx cleopatra*, *Hipparchia fatica*, *Maniola telmessia*, and *Vanessa atalanta*). Table 3 summarizes descriptive statistics of deviations from the dates calculated based on the full data set.

When assessing the sensitivity of various indices of curve-stability to rarefaction (change in start, peak, and end dates, flight season duration, and number of peaks; Figs. 5, 6 and Appendix S1: Fig. S4–S5), we found that season length was the most sensitive index (Table 4), with a mean deviation of approximately 1 month when comparing rarefied data sets with full data sets (Table 3). Nevertheless, season peak and season end dates were
also strongly affected by rarefaction. The most reliable index was the season start date, which often deviated by less than 10 d. Nevertheless, when small data sets (<100 visits) were used to construct flight curves, the season start date could deviate by more than 1 month compared with the full data set (Fig. 5).

Impacts of curve reliability on abundance estimates

A total of 613 multipliers of abundance were calculated from 71 different sites, from all species but Anthocharis cardamines. The lowest multiplier of abundance was 0.05 (i.e., obtaining unreliable phenology led to a 95% under-estimation of abundance) and the highest was 9.5 (i.e., abundance was over-estimated by a factor of 9.5). A 10% over-estimation bias was found in the multipliers of abundance, since the median value was 1.10 (Table 3). In other words, one-half of the abundance indices based on unreliable phenologies were between 10% lower and 53% higher than the equivalent abundance indices based on biologically plausible phenologies, and the number of cases of under-estimation (multiplier of abundance is less than 1) was almost equal to the number of cases of over-estimation (multiplier of abundance > 1). A permutation ANOVA found no significant differences between species in the multipliers of abundance ($P = 0.686$), and a Kolmogorov-Smirnov test comparing these distributions was insignificant ($P = 0.444$).

Impact of species’ phenology

When using data sets with numerous visits (more than 250–300 visits/yr), most species exhibited the phenology expected from the literature (Schwartz-Tzachor 2007, Benyamini 2010), with the notable exception of Colotis fausta (large salmon Arab). According to Benyamini (2010), Colotis fausta appears year-round in hyper-arid regions of the Dead Sea Valley and Arava Valley, while in the Mediterranean region, it only occurs from May or later, with the local population dying out the following winter and being recolonized by migrants in the next
spring/summer. Nevertheless, the Regional GAM model predicted a year-round occurrence of this species in the Mediterranean region (Figs. 5 and 6). This is in accordance with the few sporadic observations (which were not used to train the model) of this species in the Mediterranean region of Israel from December to February (see Sightings Data, available online). The predicted season of occurrence was shorter when the model relied on fewer than 200 visits per year, but even then the model predicted at least 10 months of activity per year. This prolonged flight season of *C. fausta* resulted from occurrences until the end of December or even January. First occurrences later than April were only predicted when the model was based on fewer than 200 visits/yr. In non-rarefied data sets with more than 200 visits per year, only once did the model predict that the species is missing from the Mediterranean region during winter, and in that case it was predicted to occur as early as late February (Fig. 5a). All model simulations for non-rarefied data sets concluded that *C. fausta* adults occur in the region until the end of December. However, *C. fausta*’s host plants in the Mediterranean region of Israel (*Capparis zoharyi*; Benyamini 2010) shed their leaves and green branches during the winter (Danin and Fragram-Sapir 2019), and *C. fausta* cannot establish breeding populations during this time. Hence, while *C. fausta* adults regularly occur in the Mediterranean region of Israel during the winter, they are a sink population of the species (probably migrants from nearby deserts, where other *Capparis* species are their hosts, or offspring of the autumn generations) that cannot reproduce until the leafing out of *Capparis zoharyi* in the following spring.

**DISCUSSION**

This study used a mix of expert-based evaluation and quantitative indices, combined with a rarefaction approach, to examine the robustness of phenology curves and calculated butterfly abundances. We assessed the impacts of data deficiencies on estimated seasonality and abundance in 15 species occurring in the Mediterranean region of Israel; in this region, many species are multivoltine or even active year-round, and thus often have complex phenology patterns (Benyamini 2010; Appendix S1: Fig. S4–S5). For species with complex phenology, data sets with at least ~200 visits and 16–31 occurrences/yr were sufficient to produce biologically plausible flight curves, while for species with simple phenologies, >31 occurrences/yr were generally sufficient (Figs. 4–6). Owing to the prolonged activity time of species in this region, choosing an inappropriate definition of the starting point of the year (necessary for the Regional GAM calculation) could lead to unreliable phenology curves, and as a result, a bias in abundance values. Importantly, no single year definition was suitable for all 15 species studied, and for some (year-round) species, there appears to be no simple way to completely avoid start-date effects generated by observations at the beginning or end of the analysis-year. Nonetheless, seasonal start, peak, and end dates seem relatively robust to rarefaction, with less than 25% of the rarefied curves deviating from the full data set by more than 23 d for any of these seasonal indices (Table 3). This number is lower than observed changes in phenology with respect to climate change, indicating a shift of 9.5 (Westwood and Blair 2010; flight period length), 8.1 (O’Neil et al. 2012; first flight date), 3.7 (Roy and Sparks 2000; mean first appearance), 3.6 (Karlsson 2014; mean flight date), 2.1 (Daly 2018; mean flight date), and 1.7 (Forister and Shapiro 2003; first flight date) d/decade. This suggests that long-term impacts of climate change can be obtained with fewer data and hence earlier in the lifetime of a monitoring scheme or, for relatively small monitoring schemes or data sets, even when the phenology is complex. In contrast, abundance assessments based on unreliable curves were slightly biased toward over-estimation, and ranged between 10% under-estimation to 53% over-estimations in one-half of all cases. We consider this to be a large deviation, as abundance trends used to

### Table 3. Descriptive statistics of the deviations of the seasonal indicators of flight curves and abundance estimates based on rarefied data sets from those calculated based on the full data set.

| Deviation from:       | N   | Minimum | First quartile | Median | Third quartile | Maximum | Mean  |
|-----------------------|-----|---------|----------------|--------|----------------|---------|-------|
| Season start (d)††    | 93  | 0       | 0              | 0      | 10             | 332     | 12.18 |
| Season peak (d)       | 288 | 0       | 1              | 6.5    | 22.5           | 364     | 28.25 |
| Season end (d)††      | 93  | 0       | 0              | 2      | 18             | 238     | 21.33 |
| Season length (d)††   | 93  | 0       | 0              | 8      | 42             | 310     | 32.96 |
| Site-specific abundance (%) | 613 | 4.9     | 89.7           | 110.2  | 152.9          | 950.0   | 134.3 |

**Notes:** Monitoring year start date was set to 1 January for all species except *Archon apollinus*. Descriptive statistics were calculated based on the absolute value of the deviations, so positive values (indicating a later date than that of the full data set) were not offset by negative values (indicating an earlier date than that of the full data set). Site-specific abundances were calculated only for curves that were classified as unreliable and divided by the corresponding abundance estimate (same species, site, and year) when the curve was assessed as biologically plausible (see the “Study design” section), to assess the direction of the bias (over-estimation when relative abundance >100% and vice versa). d, days.

†Based on species that are not active year-round (*Anthocharis cardamines*, *Archon apollinus*, *Melanargia titea*, *Gonepteryx cleopatra*, *Hipparchia fatua*, *Maniola telmessia*, and *Vanessa atalanta*).
estimate species conservation status use a 50% decline over 10 yr as the threshold of the least severe category of endangerment (Vulnerable; IUCN 2012). Therefore, the use of untested phenology curves (which may include misleading curves) to estimate abundance could cause a bias, and particularly a directional bias of missing negative trends.

Our results demonstrate the challenge of correctly assessing butterfly phenology and abundance in areas where adult butterflies are active year-round, with no clear seasonal break (e.g., during winter in temperate climates). This problem may be further amplified in tropical areas, where larger numbers of species are active year-round. With sufficient data and expert knowledge, a priori expectations can be developed, allowing for the assessment of phenology curves prior to analysis (see also Arfan et al. 2018). In the absence of such knowledge, our procedure for testing the robustness of phenology curves can be applied (i.e., analyzing the consistency of phenology curves produced using rarefied data sets and different year start dates). Based on our findings, we suggest the following guidelines for analyzing data from butterfly monitoring schemes operating in non-temperate climates.

Predicted flight curves are an important reference when evaluating the phenology output of Regional GAM. Expert knowledge of the species is sometimes available, although sometimes this knowledge is informal and/or semi-quantitative (e.g., Vanessa cardui numbers should rise around April due to migration from the south). Predictions should consider the life histories of the species in similar regions/climate; for example, Anthocharis cardamines peaks in May/June in Britain (Courtney and Duggan 1983), while in Israel, this species peaks in February–March (Fig. 5).

The analysis year should be defined on a per species basis, since no universal definition fits all species in regions where adult butterflies are active year-round (Wolda 1988, Hamer et al. 2005, Maicher et al. 2018). For species that occur only in certain seasons, the start of the analysis-year should be set at least 1 or 2 months prior to the expected start of the season to ensure that the season is fully captured. In our data set, this was the case for Archon apollinus, which occurs from October–May (Fig. 5). For year-round species, we recommend choosing the start point with the lowest relative abundance while also taking note of seasons for which the relative abundance trend is expected to change. For example, in aestivating species such as Maniola telemessia, more observations are expected in the autumn than in the summer because adults increase their activity following aestivation. In this case, defining the start of the monitoring season in October fails to capture this lower aestivation. In this case, defining the start of the monitoring season in October fails to capture this clear seasonal break (e.g., during winter in temperate climates). When evaluating the phenology output of Regional GAM, an inadequate definition of the monitoring year could lead to extreme values at the start or end of the monitoring year, even in curves that appear to be otherwise biologically plausible (Fig. 3a). For species that are active year-round, a possible solution could be to run the Regional GAM over a period of ~16 months and then remove the edges to obtain the 12-month phenology and abundance.

Table 4. Coefficients of a generalized linear model of the deviations of the rarefied curves from the values of the seasonal indices of the full data set by index type, species, and number of visits (years were a random predictor).

| Predictor                        | Estimated coefficient ± SE | P     |
|----------------------------------|----------------------------|-------|
| Season start date                | 2.71 ± 0.40                | <10^{-3} |
| Season peak date                 | 5.19 ± 0.44                | <10^{-3} |
| Season end date                  | 3.12 ± 0.38                | <10^{-3} |
| Season length (d)                | 3.61 ± 0.38                | <10^{-3} |
| Number of visits                 | −5.94 × 10^{-3} ± 8.5 × 10^{-4} | <10^{-3} |

*Notes:* Deviations were measured as absolute number of days. Positive coefficients indicate tendency to larger deviations and vice versa. All species are compared to Anthocharis cardamines (simple phenology). Predictor rows whose P values are lower than 0.05 are given in boldface type.
However, this solution requires further software development (e.g., to the R package rbms; Schmucki 2018). In the meantime, running the model with two starting points allows for the detection of biases and the evaluation of their impacts.

Drawing curves for visual evaluation is essential, even when no expert-based predictions are available, as it allows for the identification of clearly implausible curves (e.g., straight or monotonically increasing or decreasing curves), start-date effects or peculiar seasonal patterns. Furthermore, strong deviations from the expected phenology could depict real ecological trends rather than insufficient sampling. For instance, in this study, the patterns obtained for Colotis fausta (Figs. 5 and 6) differed from those anticipated, but yielded new understanding of the species’ ecology.

Counting the number of visits and occurrences (as well as their timing) on which each flight curve is based helps to assess robustness and differentiate between potential artifacts and real ecological phenomena. Another possible measure to indicate artifacts is patterns in the unexpected observations. For example, if only a few observers record out-of-season observations in a given species, it might indicate erroneous species identification. Rarefaction could also be used to test curve reliability: if similar phenologies are produced from randomly rarefied data sets, they are not likely to be products of few misidentifications.

**Implications for BMS design and implementation**

One question that is relevant to any BMS (or monitoring scheme) in an area where species are active year-round is whether a break in observations may bias the phenology and abundance outcomes. In Israel, the summer period (July–September) marks an official break, but volunteers are asked to continue monitoring if possible. Accordingly, in this study, we tested the impact of selectively removing all summer observations. The outcomes indicate a strong impact on the flight curves of aestivating species, eliminating the aestivation altogether (Fig. 3b); however, it had almost no impact on univoltine species that are not active during the summer (e.g., Anthocharis cardamines and Archon apollinus; Fig. 5). For species that are active year-round, removing summer observations had no clear impact on the resulting phenology curves. In summary, conducting observations throughout the year is important; however, citizen science depends on people’s long-term motivation, and it is difficult to request that volunteers monitor year-round without a break. Options to address this problem include (1) motivating volunteers to continue observations during break times, with the hope that at least some volunteers continue (an option that works successfully in Israel); (2) to assign more than one volunteer per transect (as is done in some BMSs); (3) to offer volunteers the option to select among two or three alternative sampling periods, in order to reduce the synchrony between volunteers’ breaks; and (3) to employ paid observers during relevant months (e.g., the hottest months in hyper-arid areas where temperatures are inhospitable).

**Value of the community model**

The community model of operation, where groups of volunteers support each other, seems to successfully address the barriers of volunteer-activation and coordination. To allow this, BMS-IL increasingly places an emphasis on a community model. The community coordinator can be invaluable for controlling data collection in BMS; for example, they personally connect with each volunteer and can plan volunteers’ leaves of absence to ensure that personal needs are met while simultaneously minimizing periods in which no data are collected in any given geographical area. The community model also strengthens the capacity for data validation owing to both stronger peer-to-peer support (e.g., through social networking and the sharing of pictures for assistance with species’ recognition) and by offering the opportunity to compare transects within and between communities. Notably, in each community, there are varying levels of skill in identifying butterflies in the field, and hence different transects from the same community could serve as replicates. This would also facilitate the testing of differences in phenology between smaller and more uniform ecological regions (e.g., with smaller variations in annual rainfall), as the basic assumption of all Regional GAM models is that transects have similar climates and should produce similar flight curves.

**Limitations and outlook**

Although careful data collection and analysis can mitigate a range of challenges associated with applying the Regional GAM model in non-temperate climates, some issues are more challenging. Our results for the painted lady (Vanessa cardui) present an outstanding example: while this species is very common in Israel (Table 1) and globally, and has a well-known life history (Benyamini 2010, 2017, Stefanescu et al. 2016, Talavera and Vila 2016, Talavera et al. 2018), Regional GAM consistently produces unreliable flight curves for it (Table 2). Vanessa cardui is a migrating species, with large numbers of adults conducting a huge northward migration from the Sahara or Sub-Saharan Africa to Europe during the spring, and a smaller migration southward during the autumn. Therefore, one could expect two peaks of activity, one in the spring and a smaller one in the autumn. However, Israeli Vanessa cardui butterflies observed during the migration seasons are a mixture of arrivals and locally hatched individuals (Talavera et al. 2018), and their relative abundances may vary both seasonally and geographically. Consequently, while some annual variations can indicate real differences in the life cycle (e.g., a robust peak in late January 2015; Fig. 3a), some peaks...
represent start-date effects, disappearing when the monitoring year is defined differently. This exemplifies the need for ample data and expert knowledge of species with complex phenologies; in addition, further development of methods to scrutinize the data and identify robust estimates of phenology and abundances is needed. Nevertheless, as deviations in abundance estimates rarely exceed 53% bias (Table 3), large-scale population irruptions should be interpreted as true signals.

Another limitation, inherent to the Regional GAM method, is the inability to include environmental covariates (such as latitude, mean annual temperature, etc.) when computing phenology curves. This is already implemented in other methods such as the Generalized Abundance Index, which is also suitable for assessing the phenology curves of multivoltine species (Dennis et al. 2016). However, the Generalized Abundance Index is a parametric model that regresses phenology parameters based on environmental covariates when the general shape of the curve (i.e., number of peaks) is predefined. This practice can be justified for species with a constant number of generations per year; however, the number of generations per annum can vary. The Generalized Abundance Index computes the number of splines (turns and bends of the curve, corresponding with generations, migrations, aestivation, etc.), similar to the Regional GAM, and offers no straightforward method of regressing complex phenologies with covariates. However, when evidence for the number of generations is available (e.g., sightings of worn individuals, and then later in the season, sightings of younger, less worn individuals), the Generalized Abundance Index could be used to assess environmental impacts on the number of generations.

The complexity of species’ phenology in Israel indicates both a need for, and potential for, further research into the factors that best explain inter- and intra-annual variability in abundance and distribution. In addition, the methods and analyses presented here could also be applied to BMSs already operating in Florida and Texas, USA, southern Italy, and Lebanon, as well as any future schemes established in warm climates.

**Acknowledgments**

The authors wish to express their gratitude to all BMS-IL’s volunteers for laboriously collecting the data analysed in this paper. Two anonymous reviewers provided useful feedback on an earlier draft of this paper. We are grateful to The Israeli Lepidopterists Society, and especially to its coordinator Leah Benyamini, for founding and supporting BMS-IL throughout the years. Zvi Avni and Gadi Ish-Am assisted in data validation. Elana Erez, Hen Cohen, Tal Gitman, Inbar Ktalav, Racheli Schwartz-Tzachor, and Hadas Yelinek are the members of the volunteer communities. The work of Orr Comay and Tal Melocha has been funded by The Israeli Lepidopterists Society. Guy Pe’er was supported by sDiv, the Synthesis Centre of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, funded by the German Research Foundation (FZT 118). We thank Reto Schmuki and Martin Musche for assistance in data analyses. Tips from Elisabeth Kühn, Diana Bowler, Anett Richter, and members of the Department of Ecosystem Services at iDiv/UFZ were valuable as well.

**Literature Cited**

Arfan, M., G. Pe’er, B. Bauch, J. Settele, K. Henle, and K. Klenke. 2018. Evaluating presence data versus expert opinions to assess occurrence, habitat preferences and landscape permeability: a case study of butterflies. Environments 5:1–20.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1–48.

Benyamini, D. 1999. The biology and conservation of *Iolana alferii* Witshire 1948: the Burning Bush blue (Lepidoptera: Lycaenidae). Linneana Belgica 17:119–134.

Benyamini, D. 2008. *Euchloe falloui* Allard, 1867 (Pieridae) the butterfly with the longest diapause. Nota Lepidopterologica 31:293–295.

Benyamini, D. 2010. A field guide to the butterflies of Israel, including of Mt. Hermon, Sinai and Jordan. Revised edition. Keter Publishing House, Jerusalem, Israel.

Benyamini, D. 2017. A swarm of millions of *Vanessa cardui* (Linnaeus, 1758) in winter-spring 2015–2016 in the south-east Mediterranean – The missing link (Lepidoptera, Nymphalidae). Atalanta 48:103–128.

Cullen, B., M. Ram, T. Zamin, and L. McRae. 2008. The tropical biodiversity data gap: addressing disparity in global monitoring. Tropical Conservation Science 1:75–88.

Courtney, S. P., and A. E. Duggan. 1983. The population biology of the Orange Tip butterfly *Anthocharis cardamines* in Britain. Ecological Entomology 8:271–281.

Daly, K. M. 2018. Morphological and phenological responses of butterflies to seasonal temperature increase in Alaska. Thesis. University of Alaska Fairbanks, Fairbanks, Alaska, USA.

Danin, A., and O. Fragman-Sapir. 2019. *Flora of Israel* online. http://flora.org.il/en/plants/a.org.il/en/plants/

Dennis, E. B., S. N. Freeman, T. Brereton, and D. B. Roy. 2013. Indexing butterfly abundance whilst accounting for missing counts and variability in seasonal pattern. Methods in Ecology and Evolution 4:637–645.

Dennis, E. B., B. J. T. Morgan, S. N. Freeman, T. M. Brereton, and D. B. Roy. 2016. A generalized abundance index for seasonal invertebrates. Biometrics 72:1305–1314.

Dowle, M., and A. Srinivasan. 2019. *data.table*: Extension of ‘data.frame’. R package version 1.12.2. https://CRAN.R-project.org/package=data.table

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.

Hallmann, C. A., et al. 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS ONE 10:e0185809.

Hamre, K. C., J. K. Hill, N. Mustaffa, S. Benedick, T. N. Sherrat, V. K. Chey, and M. Maryati. 2005. Temporal variation in abundance and diversity of butterflies in Bornean rain forests: opposite impacts of logging recorded in different seasons. Journal of Tropical Ecology 21:417–425.

Herve, M. 2019. RVAideMemoire: testing and plotting procedures for biostatistics. R package version 0.9.73. https://CRAN.R-project.org/package=RVAideMemoire

Isaac, N. J. B., K. L. Cruickshanks, A. M. Wedle, J. M. Rowcliffe, T. M. Brereton, R. L. H. Dennis, D. M. Shuker, and C. D. Thomas. 2011. Distance sampling and the challenge of monitoring butterfly populations. Methods in Ecology and Evolution 2:585–594.

Israel Meteorological Service. 2013. Explanation for multianual temperature values 1995–2009. February 2013. http://
Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/eap.2196/full