Methods in Ecology and Evolution

MICHAEL BELITZ (Orcid ID : 0000-0002-8162-5998)

The accuracy of phenology estimators for use with sparsely sampled presence-only observations

Running Header: Accuracy of presence-only phenology estimators

Michael W. Belitz[1,2,*], Elise A. Larsen[3], Leslie Ries[3], Robert P. Guralnick[1]

1. Florida Museum of Natural History, University of Florida, Gainesville, FL 32611
2. University of Florida Biodiversity Institute, Gainesville, FL, 32603
3. Department of Biology, Georgetown University, Washington DC, United States, 20057

* Corresponding author:

Michael Belitz; Florida Museum of Natural History; University of Florida; Dickinson Hall 32611; mbelitz@ufl.edu

Abstract

1. Phenology is one of the most immediate responses to global climate change, but data limitations have made examining phenology patterns across greater taxonomic, spatial, and temporal scales challenging. One significant opportunity is leveraging rapidly increasing data resources from digitized museum specimens and community science platforms, but this assumes reliable statistical methods are available to estimate phenology using presence-only data. Estimating the onset or offset of key events is especially difficult with incidental data, as lower data densities occur towards the tails of an abundance distribution.

2. The Weibull distribution has been recognized as an appropriate distribution to estimate phenology based on presence-only data, but Weibull-informed estimators are only available for onset and offset. We describe the mathematical framework for a new Weibull-parameterized estimator of phenology appropriate for any percentile of a distribution and make it available in a R package, phenesse. We use simulations and empirical data on open flower timing and first arrival of monarch butterflies to quantify the accuracy of our estimator and other commonly used phenological estimators for 10 phenological metrics: onset, mean, and offset dates, as well as the 1st, 5th, 10th, 50th, 90th, 95th, and 99th percentile dates. Root mean squared errors and mean bias of the phenological estimators were calculated for different patterns of abundance and observation processes.

3. Results show a general pattern of decay in performance of estimates when moving from mean estimates towards the tails of the seasonal abundance curve, suggesting that onset and offset continue to be the most difficult phenometrics to estimate. However, with simple phenologies and enough observations, our newly developed estimator can provide useful onset and offset estimates. This is especially true for the start of the season, when incidental observations may be more common.

4. Our simulation demonstrates the potential of generating accurate phenological estimates from presence-only data and guides the best use of estimators. The estimator that we developed, phenesse, is the least biased and has the lowest estimation error for onset estimates under most

simulated and empirical conditions examined, improving the robustness of these estimates for phenological research.

1 Introduction

Global climate change is disrupting the seasonal timing of life history events—the phenology—of numerous species worldwide (Parmesan, 2007). Developing accurate phenological metrics (phenometrics) from presence-only data has wide utility in the biological sciences, especially given the rapidly-increasing number of observations being reported on global biodiversity portals. Plant phenology is an ecologically important regulator of ecosystem processes that influences the seasonality of albedo, fluxes of water, and nutrient cycling (Richardson et al., 2013). At a community level, phenology is important because food webs depend on the developmental timing of interacting species (Chuine and Régnière, 2017). The disruption of important ecological interactions can have negative demographic effects for interacting species and disrupt ecosystem functions through nutrient cycles (Beard et al., 2019).

Historically, most phenological research has used data from long-term ecological research projects and standardized monitoring programs to estimate variation in phenology through space or time. These programs often collect presence/absence or count data, allowing researchers to use a variety of robust statistical methods to model phenology including generalized additive models (Moussus et al., 2010; Schmucki et al., 2016) and extensions of occupancy models (Chambert et al., 2015; Socolar et al., 2017). However, such standardized data are limited in space and time. Instead, the fastest growing data resources usable for documenting phenology come from incidental or opportunistic community science platforms e.g., iNaturalist, where users share and identify photos of biodiversity across the globe (Barve et al., 2020). In 2019, 538,906 digital vouchers of butterflies were added to iNaturalist, a 1.8-fold increase from 2018 (as of Feb. 2020). Additionally, initiatives to digitize natural history museum specimens have mobilized millions of historical occurrence records (Nelson and Ellis, 2018). Existing data resources will only grow as most specimen data are not yet

digitized. For instance, it is estimated that only five percent of arthropod specimens housed in North American natural history collections are both digitized and georeferenced (Cobb et al., 2019). The exponential growth of incidental or opportunistic biodiversity records plus scalable solutions for annotating phenology (Brenskelle et al., 2019) provide novel opportunities to better understand past and present phenology patterns and associated drivers across greater spatial, temporal, and taxonomic scales.

Generating accurate phenology estimates with unstructured community observations is challenging because sampling effort is unquantified, and the biases inherent to the observation process cannot be fully known (Kelling et al., 2019). This often translates into only having a series of presence records over time as an input into calculating phenometric estimates. Additionally, incidental data collection is not repeated so phenological estimates using these data must be aggregated to arbitrary spatial units, potentially altering phenology patterns across different scales (Keyzer et al., 2017).

The two most common metrics used with incidental data to estimate timing of a particular phenophase or subphase are the mean date of all observations within a year and the first observation date of an event in a year (Jones and Daehler, 2018; Ward et al., 2016). Although mean dates have consistently been found to be an accurate phenometric (Moussus et al., 2010; Bertin, 2015; Miller-Rushing et al., 2008), they may not always be the most biologically relevant phenometric of interest. For example, studies examining phenological mismatch are often interested in understanding if the start and end of seasonal abundance curves overlap, while the middle of seasonal abundance curve is less critical (Both et al., 2006; Mayor et al., 2017). Additionally, for species with multiple broods such as multivoltine insects, mean estimates may reflect a time in-between broods where no adult individuals are emerged. It is therefore critical that mathematical frameworks are developed to estimate phenometrics that more closely reflect the start and end of seasonal abundance curves. This is challenging since first and last observation dates based on incidental data are biased, as the first presence observation of a phenological event almost certainly occurs after the true event begins. Likewise, the true offset of an event likely occurs after the last observation is documented.

Recently, Pearse et al. (2017) developed a statistical estimator that fits a Weibull distribution to estimate the onset and offset of historic and contemporary phenology based on sparsely sampled,

presence-only, incidental observations. By drawing strength across many observations, not simply the earliest observation, Pearse et al. (2017) demonstrate that their method provides estimates closer to the true onset of a process than the first sample. Still, estimating the true onset or offset of a process may be more challenging than estimating a percentile of the phenology curve within the bounds, because it is notoriously difficult to model the tails of distributions as there are fewer data points to parameterize the model (Pearse et al., 2017).

Our goal was to develop a Weibull-parameterized point estimator that could provide accurate estimates of phenology for any percentile of a distribution. Here, we describe the mathematical framework used to develop our estimator, which we make available in an R package titled phenesse. Using this framework, we quantify the accuracy of our newly developed estimator and three other commonly used phenological estimators based on simulated and empirical incidental data. We use simulations to assess the accuracy of ten phenometrics estimated using four estimators while varying parameters for the underlying seasonal abundance curves and observation process. We vary the seasonal abundance curve by changing the modality, duration, and skewness of the underlying distributions that we use to draw our simulated observations. The observation process was varied by changing the number of observations used to calculate estimates and by changing if the observation process was random or skewed towards earlier observations.

We further evaluate the application of our newly developed estimator on two exemplar empirical incidental datasets. The first example compares flowering timing for two plant species, using incidental records from the citizen science platform iNaturalist (http://www.inaturalist.org/), and more structured data assembled from the National Phenology Network. The second example again uses iNaturalist observations, but focuses on the migratory monarch butterfly, in comparison to onset timing based on presence/absence reporting from the citizen science resource, Journey North, focused specifically on first sightings. This allowed us to test the phenesse estimator based on two dynamics fundamental to phenology, the development time of temperature-limited growth (common for plants and ectothermic animals) and also arrival timing of volant migrants, which can include insects, birds or bats. Our study contributes a quantitative framework to estimate phenological events more accurately using presence-only observations and informs future researchers on the phenological metrics and methods that can lead to more robust estimates.

2 Methods

2.1 Developing a phenology point estimator for any percentile of a distribution

Cooke (1979) constructed bias corrected estimators of the bounds of random variables which were better than the extreme order statistics for many classes of random variables. We adapted this statistical framework to develop a numerical solution that calculates a point estimate for any percentile. The joint distribution of the most recent observations of an event has approximately the same Weibull form, regardless of the distribution from which those sightings were sampled (Cooke, 1980). Therefore, the Weibull distribution is useful to model data integrated across multiple sources that may use different sampling regimes.

To estimate any percentile, we estimated the maximum likelihood estimation (MLE) of the shape and scale parameters of the Weibull distribution using the fitdistrplus package (Delignette-Muller et al., 2019). After estimating the shape and scale parameters of the Weibull distribution, we plotted the cumulative distribution function (CDF) of the parameterized Weibull distribution. We calculated the CDF of the Weibull as $F(x;\lambda,k) = 1 - e^{(-x/\lambda)^k}$ where x are the observations, λ is the scale parameter, and k is the shape parameter. We then solved for the observation date that corresponded to a CDF of 0.01 and 0.99 respectively to calculate approximate bounds of the Weibull distribution parameterized by the original observations. The observation date that corresponds to a CDF of 0.01 was calculated as $\lambda * (-log(1 - 0.01)^{\frac{1}{k}})$ and the observation date that corresponds to a CDF of 0.99 was calculated as $\lambda * (-log(1 - 0.99)^{\frac{1}{k}})$. To ensure a smooth CDF curve, we calculated the CDF of all values in a sequence with increment of 0.5 starting at the observation date that corresponds to a CDF of 0.01 and ending at the observation date that corresponds to a CDF of 0.99.

The metric here cannot estimate a true onset and offset (e.g. 0th or 100th percentile), we instead estimate what we refer to as the pseudo-0th and pseudo-100th percentile. This is calculated by subtracting one (day) from the observation date that corresponds to a CDF of 0.01 and adding one (day) to the observation date that corresponds to a CDF of 0.99. These new observation dates were assigned pseudo-CDF values of -0.001 and 1.001 respectively. In a future release of our R package, phenesse, we expect to remove the 0th and 100th percentile estimates, as these estimates are not based

on the Weibull distribution but rather an arbitrary cutoff. Instead, users will have to determine what quantile value represents the start or termination of a phenophase.

The Weibull corrected estimate ($\overline{\theta}$) was calculated as

$$\overline{\theta} = \hat{\theta}_{original} - \widehat{Bias}$$

$$\widehat{Bias} = (\frac{1}{B}\sum_{i=1}^{B}\hat{\theta}_i) - \hat{\theta}_{original}$$

$$\overline{\theta} = 2\hat{\theta}_{original} - (\frac{1}{B}\Sigma_{i=1}^{B}\hat{\theta}_i)$$

where $\hat{\theta}_{original}$ is the original estimate determined as the x-axis value perpendicular to the percentile of interest (y-axis) given the CDF fit to the original observations (Fig. 1, Step 2), $\hat{\theta}_i$ is the estimate at iteration i (Fig. 1, Step 4) given the CDF fit to newly sampled observations (Fig. 1, Step 3), and B is the number of iterations (500 for estimates included in this study).

2.2 Setting up the simulation experiments

We simulated biologically plausible seasonal abundance curves, where the number of individuals in a landscape varied over time to assess the accuracy of our estimator. We also compare our estimator to three commonly-used estimators: phest (Pearse et al., 2017), quantile, and mean. By knowing the distribution of the simulated taxa, we could determine how well our estimates compared to benchmark values.

We evaluated the effectiveness of our estimator under different classes of seasonal abundance distributions. The first shape was a normal, unimodal distribution and could represent a variety of biological processes e.g., flowering of plants, arrival of migratory birds. The second shape was an asymmetric bimodal distribution, where the second peak in the distribution was larger than the first (Fig. 2). Skewed bimodal seasonal abundance distributions could represent emergence of bivoltine insect species whose second flight has a larger abundance than the first.

For each seasonal abundance curve, we manipulated the length of the seasonal abundance curves by changing the standard deviations (SD). All unimodal distributions had a mean of 200, representing the same day of peak abundance, and three different standard deviations: 10, 20, and 40. The bimodal distributions had two peaks with the respective means of 150 and 220, with two-thirds of

the abundance occurring in the second peak. We did not model a bimodal distribution with a standard deviation of 40, because the two peaks overlapped such that the distribution formed a single peak with a large standard deviation. We examined the effect of observation skewness by running all our simulations under two sampling scenarios. The first assumed random sampling, where the probability of observing an individual was the same across the entire phenophase. The second assumed a positive skewed scenario, where more observers are capturing incidental reports at the beginning of the season rather than the end. Here we set weights for obtaining observations in the 10th percentile at 0.6, between the 10th percentile and the mean at 0.3, and after the mean at 0.1. These weights reflect potential skewness that may be present in incidental data. For instance, iNaturalist observation densities often peak in spring especially in relation to events such as the City Nature Challenge. For each unique modality, standard distribution, and sampling combination, we randomly selected 10, 20, or 50 individuals without replacement using the sample function in R to generate the "incidental observations" used to estimate a suite of phenometrics. We conducted 100 simulations for each unique combination of simulation parameters (Table 1).

2.3 Analyzing the accuracy and bias of phenological estimators on simulated data

We evaluated the accuracy of our newly developed estimator, phenesse, by comparing estimated values to the benchmark values generated by the simulations. For phenometrics estimating the start (onset) and end (offset) processes, we compared our estimates to those generated by the R package phest. Phest offers an analytical solution to calculate a Weibull informed estimate of the limits of a phenology curve but can only calculate the bounds of a distribution (Pearse et al., 2017). Therefore, we only tested the accuracy of the phest estimator for the start (onset) and end (offset) processes. We also compared our estimates to the accuracy of the default quantile algorithm from the R Stats package for the pseudo-0th, 1st, 5th, 10th, 50th, 90th, 95th, 99th , and pseudo-100th percentiles of a process, where the pseudo-0th and pseudo-100th percentiles represent the estimated onset and offset processes respectively. Quantile estimates are commonly used to approximate a phenometric near a tail of the distribution (Brooks et al., 2017; Jonzen, 2006). Finally, we assessed the accuracy of a

mean phenometric for each simulation experiment by comparing mean estimates to benchmark mean values.

For each simulation experiment, we calculated the root mean squared error (RMSE) and mean bias of the estimates. RMSE is calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}}$$

where P is the estimation of a phenometric and O is the observed benchmark value of the corresponding phenometric. Bias metrics can be useful in determining if an estimator consistently underestimates or overestimates a benchmark value. Mean bias was calculated as

$$MeanBias = \frac{1}{n}\sum_{n}^{1}(P_i - O_i)$$

2.4 Analyzing the accuracy and bias of phenological estimators on empirical data

We further evaluated the application of our phenesse estimator and the other estimators using empirical data. We did so by comparing estimated values generated using incidental data to benchmark values generated using more structured community science data. We analyzed the accuracy of estimating open flower phenology for eastern redbud (*Cercis canadensis*) and common milkweed (*Asclepias syriaca*), and also estimated the first arrival date of migrating monarch butterflies (*Danaus plexippus*).

Incidental plant observations recorded on iNaturalist during the year 2019 of redbud and milkweed were downloaded from GBIF (2020). We also downloaded open flower status and intensity data of redbud and milkweed for 2019 from the NPN. These data are collected by amateur and professional scientists and include records of the presence or absence of open flowers, taken repeatedly at the same site throughout a year. We generated 30-km hexagonal grids across the United States and counted the number of iNaturalist and NPN records for each species found in each grid. For each species, grid cells were filtered to those with at least 100 NPN records and 10 iNaturalist observations. Inaturalist records in these cells were scored for the presence or absence of open flowers using the software tool ImageAnt (https://gitlab.com/stuckyb/imageant). Considering only records with open flowers, sufficient data were available for redbud in two cells (one each in New York and

Indiana), and milkweed in one cell in Minnesota. Phenometrics were estimated from these incidental data using quantile, phenesse, phest, and mean estimators.

We used generalized additive models (GAMs) to generate benchmark phenometrics using the NPN presence/absence data at our three grid cells. We fit the GAMs using the R package mgcv (Wood, 2011) to a binomial distribution and specified our smooth terms to be cyclic cubic regression splines; day of year was the predictor variable and presence of open flower was the response variable. Benchmark phenometrics were estimated by extracting values from our predicted GAM models (detailed methods in Supporting Information) for comparison to estimators using incidental iNaturalist records.

Incidental monarch butterfly observations recorded on iNaturalist during the year 2018 were downloaded from GBIF (2019). Our benchmark data was collected by Journey North community scientists that report the first sighting each spring of monarch butterflies migrating from Mexico into the United States and Canada (Howard and Davis, 2009). We restricted our study area to only include monarch records collected between the longitudinal degrees -94 to -68 and the latitudinal degrees 36 to 42 to ensure documented butterflies migrated. Records within this area were annotated to verify the presence of an adult butterfly. We generated 30-km hexagonal grid cells across the study area. Grid cells were considered suitable for analysis if there were at least two Journey North records and 10 iNaturalist observations of adult monarchs, with at least one Journey north record prior to any iNaturalist observation. In total, 40 cells were kept for analysis. In each cell, the benchmark value was the earliest day of year of the Journey North records, and the estimated value was the onset estimate of the quantile, phenesse, and phest estimators applied to iNaturalist records. Onset was the only comparable phenometric using the Journey North data, as these data focused on the first monarch observed in a season. Root mean squared error (RMSE) and mean bias of the estimates compared to the benchmark values were calculated for both the monarch and open flower examples.

2.5 Code development

All functions and analyses were developed in R version 3.5.1 (R Core Team, 2018). Simulation scripts relied on the tidyverse packages (Wickham et al. 2019) and truncnorm (Mersmann et al., 2018). Figures used the additional packages cowplot (Wilke, 2018) and latex2exp (Meschiari, 2015).

The statistical estimator developed and introduced in this paper is available in the R package phenesse (Belitz, 2020).

3 Results

3.1 Accuracy and bias of phenological estimates of onset and offset

Simulation study

When estimating onset phenology under the random sampling scenario, the estimates from our R package, phenesse, always produced more accurate estimates (lower RMSE) than estimates produced using the R package phest, which itself was more accurate than the quantile estimator (Fig. 3a; Fig. 4a). Phenesse estimates of offset were also more accurate when the seasonal abundance curve was bimodal (Fig. 4), but phenesse and phest generated estimates that were similar in accuracy for offset when the seasonal abundance curve was unimodal (Fig. 3). Onset estimates generated under the skewed sampling scenario had lower RMSE than estimates generated under the random sampling scenario (Fig. 3), except when the seasonal abundance curve was shorter and bimodal (Fig. 4b). Conversely, offset estimates generated under the skewed sampling scenario had a higher RMSE than estimates generated under the random sampling scenario (Figs 3-4). Generally, an increase in the length of the phenological period (increase in SD) decreased the accuracy of onset and offset estimates for all estimators (Fig. S1). Across all simulations, neither onset or offset were ever the most accurate phenological metric, regardless of the parameters used to set up the simulation experiment or the estimator used (Figs S1-S2).

The mean bias calculations revealed that for the unimodal seasonal abundance curves, in general, the three estimators all over-estimated the benchmark onset value i.e. they predicted later than the actual onset and under-estimated the benchmark offset value (Fig. 5). For the bimodal seasonal abundance curve, estimates calculated by phest and the quantile estimator were similar in direction of bias to unimodal estimates. However, phenesse estimates underestimated the onset values except when observations were selected from the overlapping (20 SD) bimodal seasonal abundance curve and sampling was random (Fig. 6). Across all simulation experiments, estimates of onset and offset using phenesse generally had lower bias than estimates calculated using phest or the quantile

method (Figs S3-S4). The exception to this result was when observations are selected from the non-overlapping (10 SD) bimodal seasonal abundance curve and sampling was skewed (Fig. S6b).

Empirical study

Compared to other estimators, phenesse produced the most accurate estimates for both the onset of open flowers (Fig. 7A) and the arrival of the first monarch butterfly (Fig. 7C) based on comparisons to benchmark observations. Phenesse estimates also had the lowest mean bias in both cases (Fig. 7B, 7D). The difference between the estimated arrival of the first monarch butterfly and the benchmark arrival date decreased as the number of iNaturalist observations used to generate the estimate increased, with phenesse showing the least sensitivity to sample size (Fig. S5). Estimates of the termination of flowering (offset) were most accurate and least biased using the phest estimator (Fig. 7A; 7B). All estimators generally underestimated the date of the termination of flowering given our benchmark value.

3.2 Accuracy and bias of phenological estimates of the 1st to 99th percentile phenometrics

Simulation study

For the estimates generated using observations from the unimodal seasonal abundance curves and random sampling scenario, the mean estimate had the lowest RMSE (Fig. 3a) and bias (Fig. 5a). Under the random sampling scenario, RMSE showed a pattern of decay (Fig. 3a; Fig. S1a) and bias increased as estimates moved farther from the mean (Fig. 4a; Fig. S3a). Under the skewed sampling scenario, RMSE and bias were lowest with quantile estimates of the 1st, 5th, and 10th percentiles and increased as the percentile of interest approached offset values (Fig. 3b; Fig. 5b). The bias analyses showed that under the skewed sampling scenario, the estimators generally underestimated all percentiles except onset (Fig. 5b). Under most unimodal simulation scenarios, the 50th percentile estimates calculated using phenesse were more accurate than 50th percentile estimates using the quantile estimator (Fig. S1). Although, phenesse estimates were generally more accurate in estimating the 90th, 95th, and 99th percentiles of a unimodal seasonal abundance curve, quantile estimates had lower RMSE and bias than the phenesse estimator for the 1st, 5th, and 10th phenological metrics (Fig. S1; Fig. S3).

Results for estimates generated using the bimodal seasonal abundance curves were different from the unimodal results. Overall, the quantile estimator had lower RMSE and bias than phenesse

(Fig. S2; S4). The quantile estimator especially outperformed phenesse under the skewed sampling scenario and when estimating the 1st, 5th, 10th, and 50th percentiles (Fig. 4). Mean estimates were the most accurate phenometric when observations were generated under the random sampling scenario from the longer, overlapping, bimodal seasonal abundance curve (Fig. 4a). Although mean estimates under the random sampling scenario were less accurate than the 90th percentile estimates when observations were from the shorter bimodal seasonal abundance curve, mean estimates always provided unbiased estimates, with bias never being greater than one under the random sampling scenario (Fig. 6a; Fig. S4a). However, under the skewed sampling scenario, mean estimates were never the most accurate phenometric and always were biased by underestimating the benchmark mean value (Fig. 6b).

Empirical study

Our empirical results were similar to our simulation results with the RMSE and bias increasing towards the bounds (Fig. 7A; Fig. 7B). Estimates generated using phenesse had lower RMSE than estimates using the quantile estimator except for the 5th and 10th percentiles. Across all percentiles, phenesse generated estimates that were less biased than the quantile estimator.

4 Discussion

Researchers are increasingly using incidental data to estimate phenological metrics, in part due to exponential growth of digitized museum specimens and observers uploading their photos on platforms like iNaturalist. These platforms, now containing tens of millions of imaged and identified organisms, can be leveraged to document phenological signals (Taylor and Guralnick, 2019), and provide a unique opportunity to answer questions across greater spatial and taxonomic scales (Barve et al., 2020). Additionally, more than 390 million specimens of preserved plants reside in herbaria across the world (Thiers, 2019) and more than 300 million arthropod specimens are housed in North American natural history collections (Cobb et al., 2019). The increasing availability of data spanning three centuries for both plants and animals promises to allow researchers a better understanding of the tempo and mode of seasonal changes, but only if the statistical methods used in the analyses are robust and account for underlying biases. A critical question is whether the observation process and methods used to control bias in that process can approximate the true biological signal. Simulated

data, where the seasonal abundance curves are known, are an invaluable tool to assess the efficiency of phenological estimators (Moussus et al., 2010). Using this approach, we demonstrate that phenometric accuracy is context dependent, and under certain scenarios, sparsely sampled presence-only data can approximate the benchmark phenology signal. However, our results also serve as another cautionary tale that estimating certain phenometrics can be challenging (Miller-Rushing et al. 2008), perhaps especially so when using incidental data.

Our simulations indicate that the mean estimator generated the most accurate and unbiased estimates from observations drawn from unimodal seasonal abundance curves under the random sampling scenario. Mean date has consistently been found to be an accurate phenometric in other contexts (Bertin, 2015), and our example using open flower phenology also confirmed mean date to be an accurate phenometric. However, in many cases, the middle of the season is not the time of the greatest biological meaning, and it is rather the tails of the seasonal abundance curves that are most essential. In mismatch studies between interacting species, measuring the overlap of the beginning of two (or more) seasonal abundance curves may be critical to understanding potential consequences of phenological shifts. For example, if migratory birds fail to breed at a time of high caterpillar availability, populations may experience large declines (Both et al., 2006). Although the accuracy and bias of phenological estimators showed a general pattern of decay in performance moving from mean estimates to the tails, phenesse improves onset estimates compared to previous methods. Accurate onset estimates may be possible given enough observations and when the length of the phenophase or biological process is short. Additionally, the 5th and 10th quantile estimates performed well across most scenarios and provides another phenometric that can be used to approximate early periods of a seasonal phenology.

If focus is on early or late events, the possibility of biased sampling may improve metrics when in the same direction of the bias. On the other hand, seasonal sampling bias reduces the effectiveness of mean observations, usually the metric that is the most robust. In our open flower example, we did find onset to be more accurate than offset and the 5th and 10th percentiles were less biased than the mean estimator, suggesting that observations of open flowers are skewed early on iNaturalist for redbud and milkweed. Our results confirm that observation biases in incidental data can influence the accuracy of phenological estimators and that it is critical to understand the

underlying observation process as much as possible and use the best metric given that skew. When sampling is skewed early, then detecting onset is more tractable, but offset is likely to be inaccurate.

The ultimate utility of phenesse is in broad-scale assessment of phenology pattern and process. Researchers are not typically interested in phenometric estimates in just one limited area, but rather understanding phenology across species, space, and time. When using incidental or opportunistic records, this often involves spatial stratification using a gridding approach, and making estimates across those many grid cells. In most cases, the shape of seasonal abundance curves is not known *a priori* and unsurprisingly, estimators often do not perform well in cases where the underlying abundance curve is a skewed bimodal distribution. Even in such cases, useful estimates of phenology can still be generated with certain phenometrics and estimators. While more observations help reduce bias and error, even as few as 10 records generate relatively small error and bias in some cases. Our monarch arrival example suggests that phenesse may be especially useful when data is limited as its accuracy is least sensitive to sample size.

Estimators such as phenesse may best be used in conjunction with other approaches that reduce bias in estimates. In principle, if a species-region combination is known to not have a unimodal phenology, such priors could be used in a Bayesian framework to inform models and improve estimates (Chevillot et al., 2017). Due to the autocorrelation expected in phenology across grid cells, spatial autoregression models could also be developed to model autocorrelated data based on neighborhood relationships (Ver Hoef et al., 2018). Weight matrices used to develop the models could also include ecological and sampling covariates that are used to improve estimates in sparsely sampled areas and produce spatially smoothed results. Phenesse itself can directly inform such spatial or temporal weighting approaches. A critical part of our development of scientifically rigorous phenology methods is code to calculate confidence intervals for all the estimators used in this study. Calculated confidence intervals could be included as a covariate in further statistical analyses to weight additional regressions where estimates have smaller confidence intervals.

A disadvantage of the approach we developed is the computational speed of phenesse. Our solution to calculate a Weibull-parameterized estimate was a numerical, iterative approach. Therefore, the Weibull-parameterized estimates produced by phenesse are computationally more expensive than functions that have analytical solutions. This is especially true for confidence intervals given that they

are calculated using non-parametric bootstrapping. We recommend using parallel computing techniques when using the weib_percentile_ci function in the package phenesse.

5 Conclusions

We demonstrate the potential of producing more accurate phenological estimates from presence-only reporting and provide the first simulation results that can guide the best use of those estimators based on study-specific circumstances. Furthermore, our empirical examples using flowering phenology and arrival of a migratory butterfly demonstrate the utility of presence-only estimators using real-world data. Our Weibull-parameterized estimator improves accuracy of onset estimates and offers a wider array of phenometrics than previously available when using a Weibull distribution. Of equal importance, phenesse allows calculation of confidence intervals to provide a basis for understanding the relative precision and strength of estimated phenometrics.

Our results show that phenesse helps improve prediction of onset in cases where researchers are restricted to presence-only data. However, this improvement is context-dependent, as there is often a significant amount of bias and error depending on the shape of the underlying seasonal abundance curve and potential skew in observations. Additionally, RMSE and bias calculations are expected to perform well when observations are sampled from normal distributions. Our simulation study may oversimplify real-world phenology curves, inflating estimator performance. However, the results of our empirical examples were consistent with the results of our simulations, mitigating these concerns.

When the underlying seasonal abundance curve is known, these results can guide selection of the most appropriate estimators and phenometrics. Researchers should consider both the underlying curve and potential skew in observations to determine the most robust metric for analysis. In studies looking to document changes in phenology of a species over time when little or nothing is known about that underlying curve, mean estimates may provide the most robust results. Box-cox (Sakia, 1992) and other transformations offer additional methods that could address skew in variables. Some research topics such as studies examining phenological mismatch or changes in the total duration of a phenophase may necessitate the estimation of onset or offset values. Developing methods to generate

robust estimates towards the tails of the seasonal abundance curve will be important to fully leveraging presence-only data in phenological research.

Phenological estimates using incidental data will always be hindered by the unknown variance and biases in the observation process. Repeated, structured surveys can provide more accurate estimates by gathering information on absences, relative abundance, or survey effort and using informed knowledge to better sample across known seasonal abundance. Unfortunately, these standardized surveys are limited to select taxonomic groups over limited spatial extents and time scales. While our study quantifies the accuracy and bias of different estimators and phenometrics, continued research is needed to further develop statistical methodologies to leverage the strengths of different data sources into more unitary frameworks for estimating phenologies.

Authors' Contributions

MWB, EAL, LR, and RPG. conceived the ideas of the study. MWB wrote the code used for these and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data Availability

Code and data to fully reproduce the simulation and empirical analyses, results, and figures can be found on GitHub (https://github.com/mbelitz/belitz_etal_phenometrics) and is archived on Zenodo (https://doi.org/10.5281/zenodo.3565992). The raw iNaturalist observations used in this paper can be found in the following GBIF downloads (https://doi.org/10.15468/dl.fmeqni;

https://doi.org/10.15468/dl.2qjve9; https://doi.org/10.15468/dl.pbn3gq). Updated versions of the R package phenesse will be released on CRAN (https://cran.r-project.org/web/package=phenesse).

References

Barve, V.V., Brenskelle, L., Li, D., Stucky, B.J., Barve, N.V., Hantak, M.M., McLean, B.S., Paluh, D.J., Oswald, J.A., Belitz, M.W., Folk, R.A., Guralnick, R.P., 2020. Methods for broad-scale plant phenology assessments using citizen scientists' photographs. Appl. Plant Sci. 8, e11315. https://doi.org/10.1002/aps3.11315

Beard, K.H., Kelsey, K.C., Leffler, A.J., Welker, J.M., 2019. The Missing Angle: Ecosystem Consequences of Phenological Mismatch. Trends Ecol. Evol. 34, 885-888. https://doi.org/10.1016/j.tree.2019.07.019

Belitz, M., 2019. phenesse: Estimate Phenological Metrics using Presence-Only Data.

Bertin, R.I., 2015. Climate Change and Flowering Phenology in Worcester County, Massachusetts. Int. J. Plant Sci. 176, 107–119. https://doi.org/10.1086/679619

Both, C., Bouwhuis, S., Lessells, C.M., Visser, M.E., 2006. Climate change and population declines in a long-distance migratory bird. Nature 441, 81–83. https://doi.org/10.1038/nature04539

Brenskelle, L., Stucky, B.J., Deck, J., Walls, R., Guralnick, R.P., 2019. Integrating herbarium specimen observations into global phenology data systems. Appl. Plant Sci. 7, e01231. https://doi.org/10.1002/aps3.1231

Brooks, S.J., Self, A., Powney, G.D., Pearse, W.D., Penn, M., Paterson, G.L.J., 2017. The influence of life history traits on the phenological response of British butterflies to climate variability since the late-19th century. Ecography 40, 1152–1165. https://doi.org/10.1111/ecog.02658

Chambert, T., Kendall, W.L., Hines, J.E., Nichols, J.D., Pedrini, P., Waddle, J.H., Tavecchia, G., Walls, S.C., Tenan, S., 2015. Testing hypotheses on distribution shifts and changes in phenology of imperfectly detectable species. Methods Ecol. Evol. 6, 638–647. https://doi.org/10.1111/2041-210X.12362

Chevillot, X., Drouineau, H., Lambert, P., Carassou, L., Sautour, B., Lobry, J., 2017. Toward a phenological mismatch in estuarine pelagic food web? PLOS ONE 12, e0173752. https://doi.org/10.1371/journal.pone.0173752

Chuine, I., Régnière, J., 2017. Process-Based Models of Phenology for Plants and Animals. Annu. Rev. Ecol. Evol. Syst. 48, 159–182. https://doi.org/10.1146/annurev-ecolsys-110316-022706

Cobb, N.S., Gall, L.F., Zaspel, J.M., Dowdy, N.J., McCabe, L.M., Kawahara, A.Y., 2019. Assessment of North American arthropod collections: prospects and challenges for addressing biodiversity research. PeerJ 7, e8086. https://doi.org/10.7717/peerj.8086

Cooke, P., 1980. Optimal linear estimation of bounds of random variables. Biometrika 67, 257–258. https://doi.org/10.1093/biomet/67.1.257

Cooke, P., 1979. Statistical Inference for Bounds of Random Variables. Biometrika 66, 367–374. https://doi.org/10.2307/2335672

Delignette-Muller, M.-L., Dutang, C., Pouillot, R., Denis, J.-B., Siberchicot, A., 2019. fitdistrplus: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data.

GBIF., 2019. GBIF occurrence download. https://doi.org/10.15468/dl.fmeqni

GBIF., 2020. GBIF occurrence download. https://doi.org/10.15468/dl.pbn3gq

GBIF., 2020. GBIF occurrence download. https://doi.org/10.15468/dl.2qjve9

Howard, E., Davis, A.K., 2009. The fall migration flyways of monarch butterflies in eastern North America revealed by citizen scientists. Journal of Insect Conservation 13(3): 279-286. https://doi.org/10.1007/s10841-008-9169-y

Jones, C.A., Daehler, C.C., 2018. Herbarium specimens can reveal impacts of climate change on plant phenology: a review of methods and applications. PeerJ 6, e4576. https://doi.org/10.7717/peerj.4576

Jonzen, N., 2006. Rapid Advance of Spring Arrival Dates in Long-Distance Migratory Birds. Science 312, 1959–1961. https://doi.org/10.1126/science.1126119

Kelling, S., Johnston, A., Bonn, A., Fink, D., Ruiz-Gutierrez, V., Bonney, R., Fernandez, M., Hochachka, W.M., Julliard, R., Kraemer, R., Guralnick, R., 2019. Using Semistructured Surveys to Improve Citizen Science Data for Monitoring Biodiversity. BioScience 69, 170–179. https://doi.org/10.1093/biosci/biz010

Keyzer, C.W. de, Rafferty, N.E., Inouye, D.W., Thomson, J.D., 2017. Confounding effects of spatial variation on shifts in phenology. Glob. Change Biol. 23, 1783–1791. https://doi.org/10.1111/gcb.13472

Mayor, S.J., Guralnick, R.P., Tingley, M.W., Otegui, J., Withey, J.C., Elmendorf, S.C., Andrew, M.E., Leyk, S., Pearse, I.S., Schneider, D.C., 2017. Increasing phenological asynchrony between spring green-up and arrival of migratory birds. Sci. Rep. 7, 1–10. https://doi.org/10.1038/s41598-017-02045-z

mbelitz, 2020, mbelitz/belitz_etal_phenometrics: Code for manuscript resubmission (Version 1.2.0). Zenodo. http://doi.org/10.5281/zenodo.3837967

Mersmann, O., Trautmann, H., Steuer, D., Bornkamp, B., 2018. truncnorm: Truncated Normal Distribution.

Meschiari, S., 2015. latex2exp: Use LaTeX Expressions in Plots.

Miller-Rushing, A.J., Inouye, D.W., Primack, R.B., 2008. How well do first flowering dates measure plant responses to climate change? The effects of population size and sampling frequency. J. Ecol. 96, 1289–1296. https://doi.org/10.1111/j.1365-2745.2008.01436.x

Moussus, J.-P., Julliard, R., Jiguet, F., 2010. Featuring 10 phenological estimators using simulated data. Methods Ecol. Evol. https://doi.org/10.1111/j.2041-210X.2010.00020.x

Nelson, G., Ellis, S., 2018. The history and impact of digitization and digital data mobilization on biodiversity research. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 374. https://doi.org/10.1098/rstb.2017.0391

Parmesan, C., 2007. Influences of species, latitudes and methodologies on estimates of phenological response to global warming. Glob. Change Biol. 13, 1860–1872. https://doi.org/10.1111/j.1365-2486.2007.01404.x

Pearse, W.D., Davis, C.C., Inouye, D.W., Primack, R.B., Davies, T.J., 2017. A statistical estimator for determining the limits of contemporary and historic phenology. Nat. Ecol. Evol. 1, 1876–1882. https://doi.org/10.1038/s41559-017-0350-0

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Richardson, A.D., Keenan, T.F., Migliavacca, M., Ryu, Y., Sonnentag, O., Toomey, M., 2013. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. Agric. For. Meteorol. 169, 156–173. https://doi.org/10.1016/j.agrformet.2012.09.012

Sakia, R.M., 1992. The Box-Cox transformation technique: a review. Journal of the Royal Statistical Society: Series D (The Statistician), 41(2), 169-178.

Schmucki, R., Pe'er, G., Roy, D.B., Stefanescu, C., Swaay, C.A.M.V., Oliver, T.H., Kuussaari, M., Strien, A.J.V., Ries, L., Settele, J., Musche, M., Carnicer, J., Schweiger, O., Brereton, T.M., Harpke, A., Heliölä, J., Kühn, E., Julliard, R., 2016. A regionally informed abundance index for supporting integrative analyses across butterfly monitoring schemes. J. Appl. Ecol. 53, 501–510. https://doi.org/10.1111/1365-2664.12561

Socolar, J.B., Epanchin, P.N., Beissinger, S.R., Tingley, M.W., 2017. Phenological shifts conserve thermal niches in North American birds and reshape expectations for climate-driven range shifts. Proc. Natl. Acad. Sci. 114, 12976–12981. https://doi.org/10.1073/pnas.1705897114

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

Thiers, B., 2019. Index Herbariorum: A global directory of public herbaria and associated staff.

Ver Hoef, J.M., Peterson, E.E., Hooten, M.B., Hanks, E.M., and Fortin, M.J., 2018. Spatial autoregressive models for statistical inference from ecological data. Ecological Monographs, 88(1), 36-59 https://doi.org/10.1002/ecm.1283

Ward, D.H., Helmericks, J., Hupp, J.W., McManus, L., Budde, M., Douglas, D.C., Tape, K.D., 2016. Multi-decadal trends in spring arrival of avian migrants to the central Arctic coast of Alaska: effects of environmental and ecological factors. J. Avian Biol. 47, 197–207. https://doi.org/10.1111/jav.00774

Wickham et al., 2019. Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Wilke, C.O., 2018. cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2."

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B) 73(1):3-36 https://doi.org/10.1111/j.1467-9868.2010.00749.x

Tables

Table 1. Parameters used in the simulation experiments.

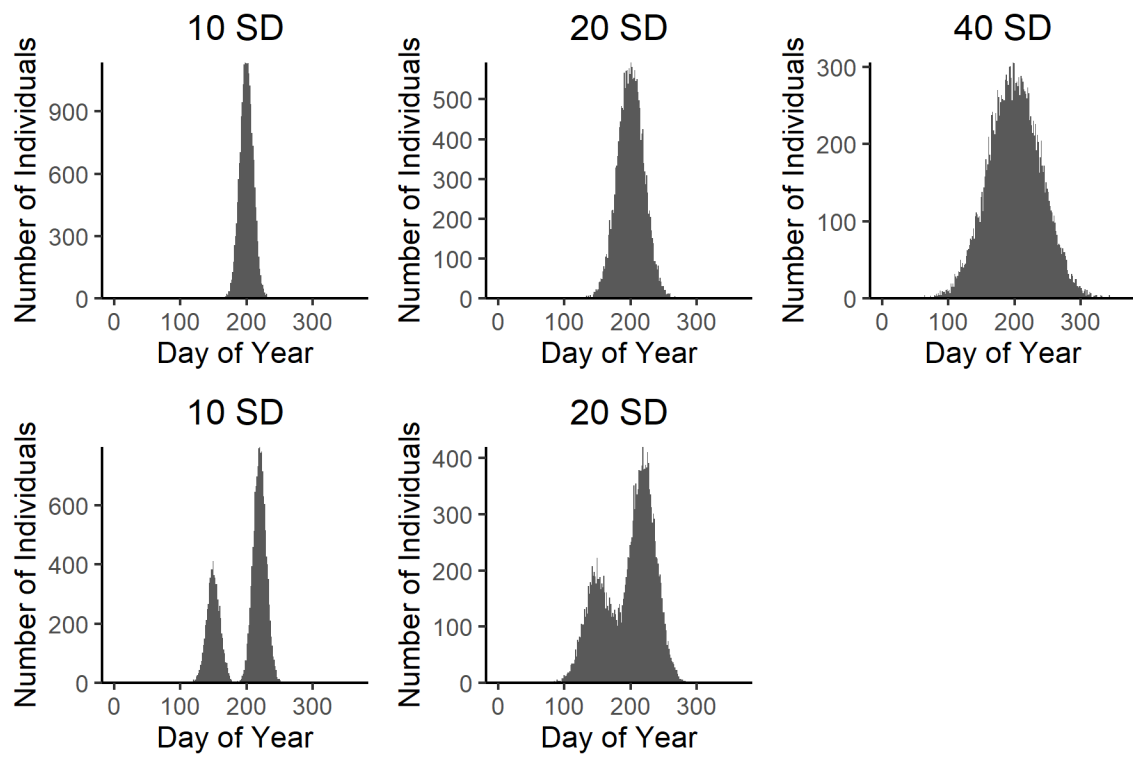| Parameter	Range of Values	Description
Standard Distribution	10, 20, 40	The standard distribution of the seasonal abundance curve. Increasing the standard distribution had the biological effect of increasing the length of the phenology period.
Observations	10, 20, 50	The number of observations used to generate our phenology estimate.

Modality	Symmetrical Unimodal or Skewed Bimodal	The shape of the seasonal abundance curve.
Sample Skewness	Random or Positive Skewed	Random sampling assumed the probability of observing an individual to be the same across the entire phenophase. Positive skewed weighted probability of sampling an individual earlier in the season.

Figure Captions

Figure 1 Workflow for developing the Weibull-parameterized point estimator. First, the CDF of the Weibull distribution is plotted, after being parameterized by the original observation points (Step 1). Next, the observation date relating to the original percentile of interest ($\hat{\theta}_{original}$; 90th in this example) is determined by calculating the observation date that is perpendicular to the percentile of interest (Step 2). To calculate bias, n number of uniform random numbers from a uniform distribution between 0 and 1 are generated, where n is the number of original observation dates. New observation dates are perpendicular to the horizontal lines determined by the n uniform numbers (Step 3). These new observation dates are used to calculate a new CDF and estimate the percentile of interest (Step 4). Step 3 and 4 is iterated i times to generate a bias value which is subtracted from the $\hat{\theta}_{original}$ to calculate the Weibull corrected estimate ($\bar{\theta}$). In the figure above, black dots represent the original observation points, and the black dashed lines show the $\hat{\theta}_{original}$. Blue dots and dashed lines represent new observation dates generated by uniform resampling.
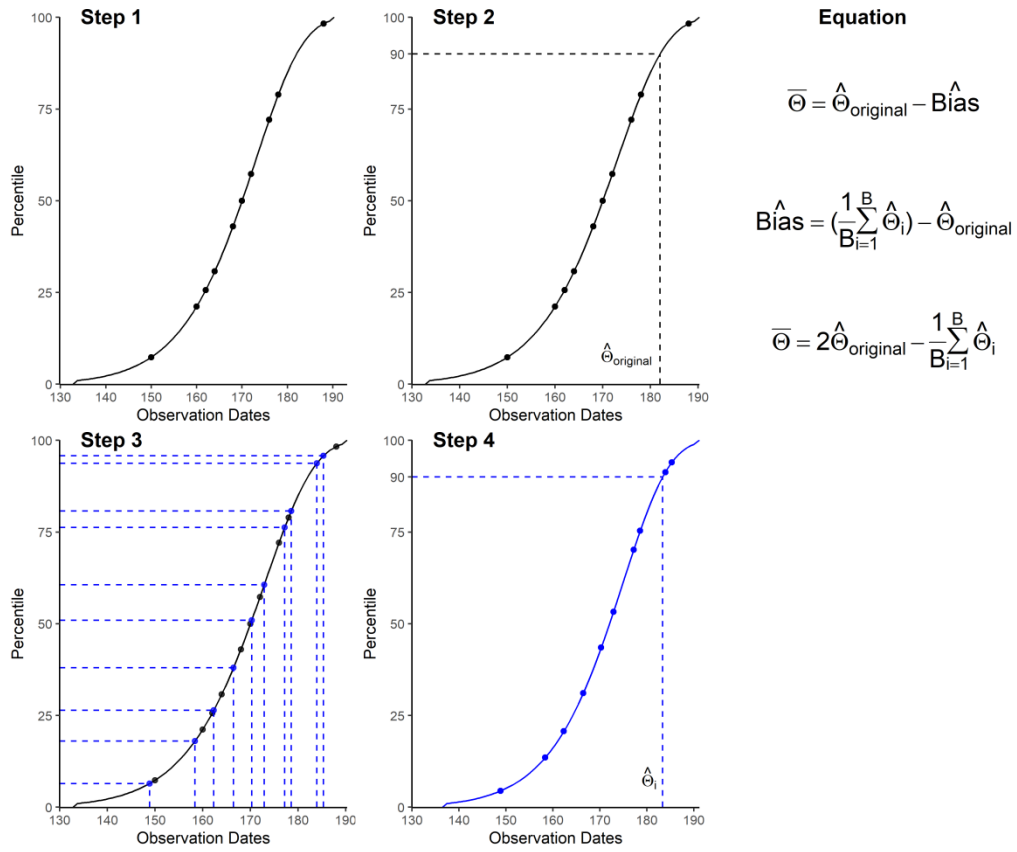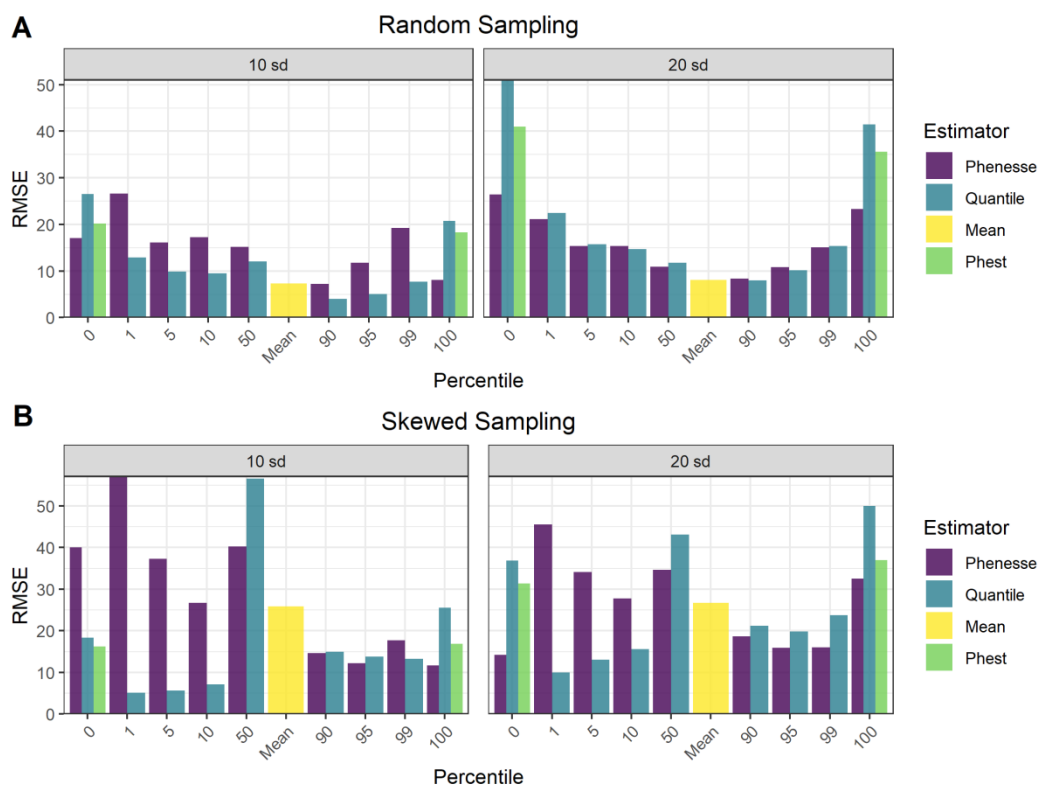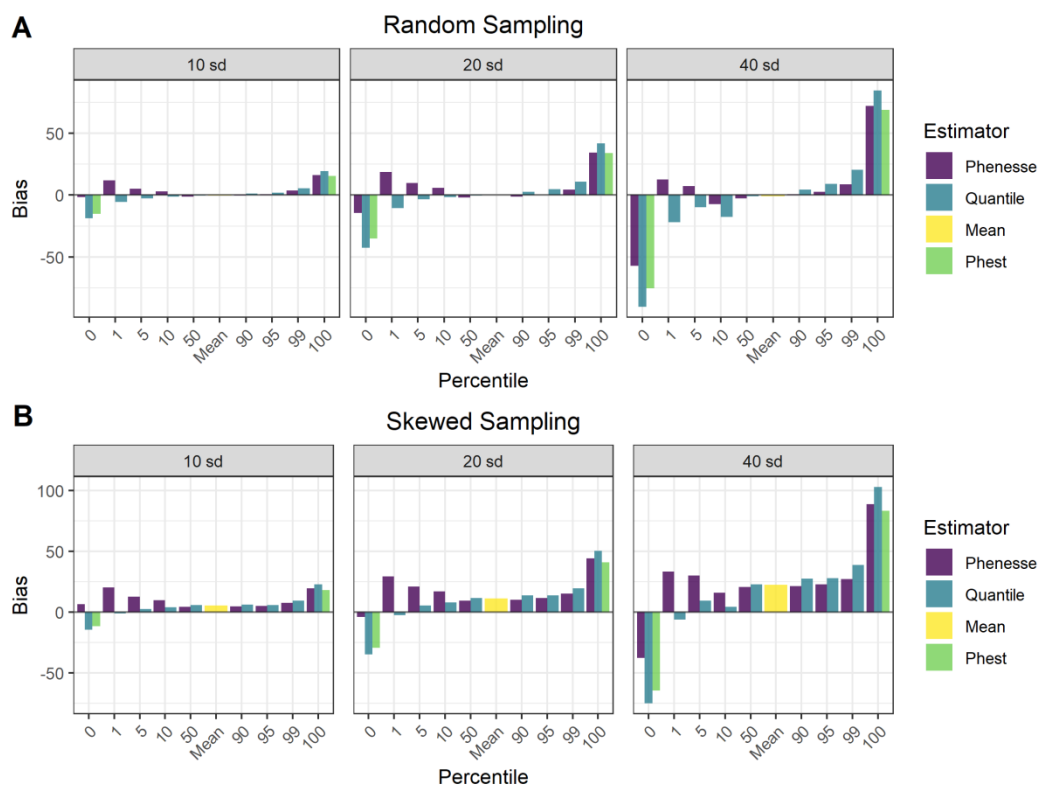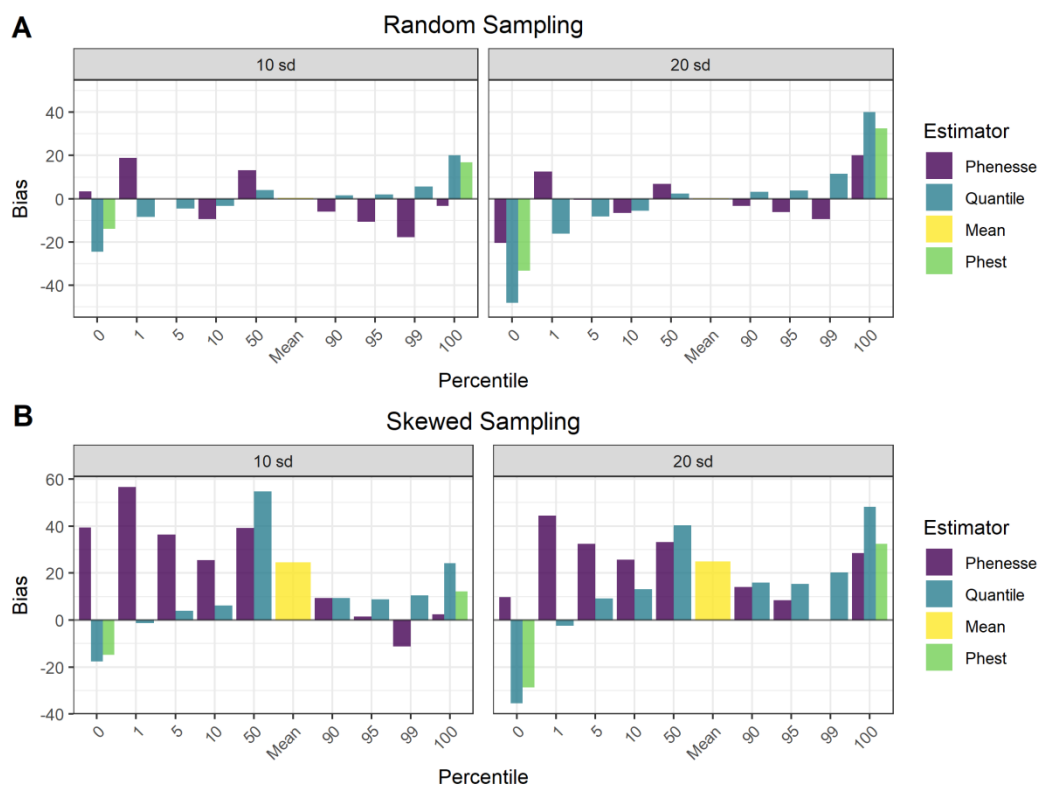
mee3_13448_f1.png

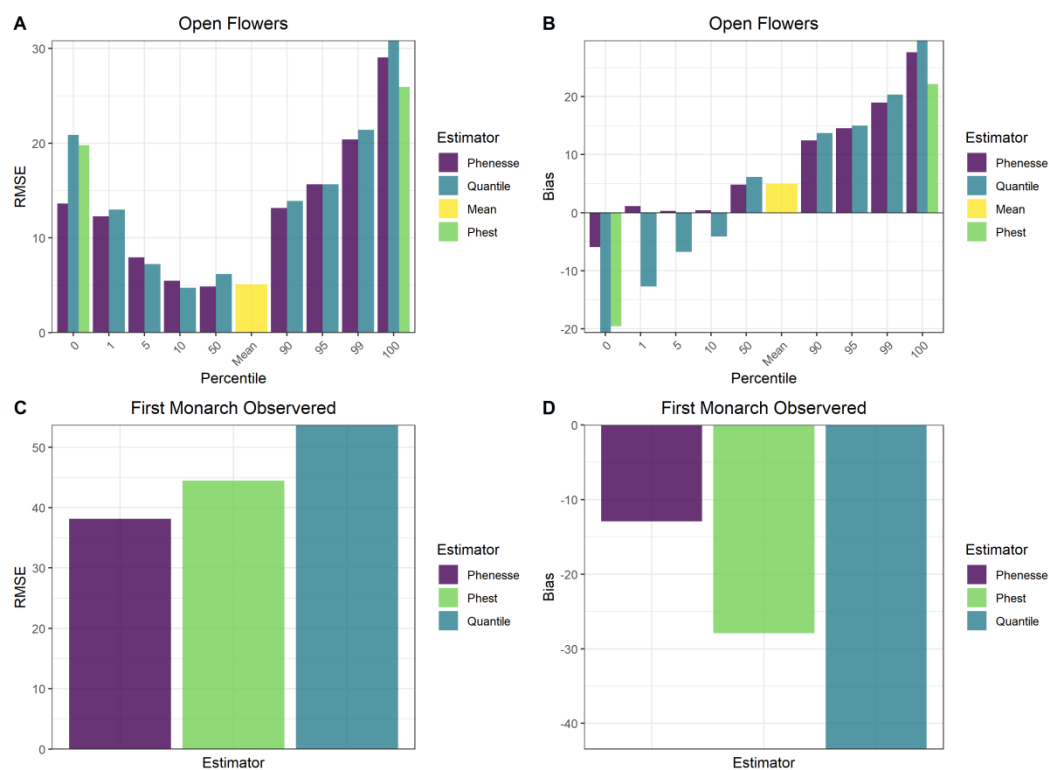mee3_13448_f2.png

mee3_13448_f3.png

A Random Sampling

B Skewed Sampling

mee3_13448_f4.png

A   Random Sampling

B   Skewed Sampling

mee3_13448_f5.png

mee3_13448_f6.png

mee3_13448_f7.png