BIOTROPICA
THE JOURNAL OF THE ASSOCIATION FOR TROPICAL BIOLOGY AND CONSERVATION

Towards effective monitoring of tropical phenology: Maximising returns and reducing uncertainty in long-term studies

Authors: Emma R. Bush, [1], Nils Bunnefeld[1], Edmond Dimoto[3], Jean-Thoussaint Dikangadissi[3], Kathryn Jeffery[1, 2, 3], Caroline Tutin[1], Lee White[1, 2,3], and Katharine A. Abernethy[1,2]

1. Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling, FK9 4LA, Scotland, UK
2. Institut de Recherche en Écologie Tropicale, CENAREST, BP 842, Libreville, Gabon
3. Agence Nationale des Parcs Nationaux (ANPN), B.P. 20379, Libreville, Gabon

Corresponding author: Emma Bush, e.r.bush@stir.ac.uk

Received _____; revision accepted _____.

This is the peer reviewed version of the following article: Bush ER, Bunnefeld N, Dimoto E, Dikangadissi J, Jeffery K, Tutin C, White L & Abernethy KA (2018) Towards effective monitoring of tropical phenology: maximizing returns and reducing uncertainty in long-term studies, *Biotropica*, 50 (3), pp. 455-464, which has been published in final form at https://doi.org/10.1111/btp.12543. This article may be used for non-commercial purposes in accordance With Wiley Terms and Conditions for self-archiving.

ABSTRACT

Phenology is a key component of ecosystem function and is increasingly included in assessments of ecological change. We consider how existing, and emerging, tropical phenology monitoring programs can be made most effective by investigating major sources of noise in data collection at a long-term study site. Researchers at Lopé NP, Gabon, have recorded monthly crown observations of leaf, flower and fruit phenology for 88 plant species since 1984. For a subset of these data, we first identified dominant regular phenological cycles, using Fourier analysis, and then tested the impact of observation uncertainty on cycle detectability, using expert knowledge and generalized linear mixed modelling (827 individual plants of 61 species). We show that experienced field observers can provide important information on major sources of noise in data collection and that observation length, phenophase visibility and duration are all positive predictors of cycle detectability. We find that when a phenological event lasts > 4 weeks, an additional 10 years of data increases cycle detectability by 114 percent and that cycle detectability is 92 percent higher for the most visible events compared to the least. We also find that cycle detectability is four times as high for flowers compared to ripe fruits after 10 years. To maximise returns in the short-term, resources for long-term monitoring of phenology should be targeted towards highly visible phenophases and events that last longer than the observation interval. In addition, programs that monitor flowering phenology are likely to accurately detect regular cycles more quickly than those monitoring fruits, thus providing a baseline for future assessments of change.

La phénologie est un élément clé du fonctionnement de l'écosystème et est de plus en plus incluse dans l'évaluation des changements écologiques. Nous considérons comment les programmes de surveillance de la phénologie tropicale, aussi bien courants qu'émergents, peuvent être rendus plus efficaces en étudiant les principales sources de bruit liées à la collecte de données sur un site d'étude à long terme. Les chercheurs du Parc National de la Lopé au Gabon, ont recueilli des observations mensuelles de la phénologie des feuilles, fleurs et fruits provenant de la canopée de 88 espèces de plantes depuis 1984. Pour un sous-ensemble de ces données, nous avons d'abord identifié des cycles phénologiques réguliers dominants à l'aide d'une analyse Fourier, puis testé l'impact de l'incertitude liée à l'observation sur la probabilité de détecter un cycle régulier significatif en utilisant des connaissances spécialisées et un modèle linéaire généralisé à effets mixtes (827 plantes individuelles de 61 espèces). Nous démontrons que les observateurs expérimentés peuvent fournir des informations importantes sur les principales sources de bruit liées à la collecte de données, et que la visibilité et la durée de la phénophase ainsi que la longueur de l'observation prédisent de manière positive la détectabilité du cycle. Nous constatons que lorsqu'un événement phénologique dure plus de 4 semaines, 10 années de données supplémentaires augmentent la détectabilité du cycle de 114 pour cent, et que la détectabilité du cycle est 92 pour cent plus élevée pour les événements les plus visibles par rapport aux moins visibles. Nous constatons également que la détection du cycle varie selon la phénophase, étant quatre fois plus élevée pour les fleurs que pour les fruits mûrs après 10 ans. Afin de maximiser les rendements à court terme, les ressources allouées à la surveillance à long terme devraient cibler les événements phénologiques hautement visibles dont la durée dépasse celle de l'intervalle d'observation. En outre, les programmes qui surveillent la phénologie de la floraison sont susceptibles de détecter avec précision les cycles réguliers plus rapidement que

66 ceux qui surveillent les fruits, fournissant ainsi une base de référence pour les évaluations futures
67 du changement.

68

69

70 **Key words**: Flowers; Fruits; Gabon; Leaves; Lopé national park; Observation uncertainty;
71 Tropical forest; Climate change;

72 **Word count:** 4585

WHILE THE IMPACTS OF CLIMATE CHANGE ON PHENOLOGY ARE WIDELY ACKNOWLEDGED (Chambers *et al.* 2013, Cleland *et al.* 2007), most of the evidence is geographically and taxonomically biased towards temperate regions and vertebrates (Feeley *et al.* 2016). There is little data available to assess change in tropical plant phenology and, to date, few relevant published studies (but see Chapman *et al.* this issue and Pau *et al.* 2013 for recent examples).

The lack of evidence for phenological change in the tropics should not be taken as evidence of no change, but instead reflects the paucity of long-term data records and the complexity of monitoring highly diverse tropical ecosystems. The question remains as to how to fill this evidence-gap and assess both stability and change in phenological function.

Phenology datasets that have already supported effective statistical tests of change have been either very long - for example Japanese cherry blossom records began in the 9th century (Aono & Kazui 2008, Primack *et al.* 2009) - or very widespread - for example The International Phenology Gardens network, initiated in 1957, includes 89 European sites across 28 latitudes (Humboldt-University of Berlin 2012). The most widespread contemporary phenology monitoring programs are those that involve citizen scientists, make use of accessible technology - such as smartphones apps - and observations made in everyday life (e.g. the USA National Phenology Network's "Nature's Notebook", USA-NPN 2017). From these successful temperate examples we learn that to achieve phenology datasets with strong *statistical power* (long-term, widespread etc.), data collection methods need to have real *sticking power* (cultural importance, familiarity, appeal to a large spread of people and ease of recording).

It is apparent that such "sticking power" remains a challenge in the tropics. Even among science-led monitoring programs, there is little coordination of recording effort across multiple sites (Morellato *et al.* 2016, Adole *et al.* 2016), fieldwork is often remote and logistically challenging and financial resources for long-term monitoring are extremely limited meaning that few sites can be considered long-term (e.g. >10 yr continuous monitoring; Mendoza et al, 2017; Adamescu et al, *this issue*). In addition, many of the tropical phenology studies that are now invaluable to assess global change were originally conceived for the study of resource availability and are not necessarily optimised to study climate-change impacts on plants (e.g. phenology monitoring at Lopé NP was originally set up in 1984 to study Gorilla and Chimpanzee foraging: Tutin *et al.* 1991).

IMPROVING STATISTICAL POWER IN ANALYSES OF TROPICAL PHENOLOGY.— While a complete redesign of tropical phenology monitoring programs within tightly coordinated networks would be ideal, we do not consider it to be feasible immediately, nor can it can reach into the past. Instead we ask: How can we ensure that existing science-led phenology monitoring programs are allocating limited resources most effectively for their research aims?

There are two ways to improve statistical power in analyses of data from phenology monitoring programs: (1) increase sample size; and (2) reduce noise. Sample size can be restrictive both spatially (e.g. the number of sites recording phenology data or the area / number of individuals monitored) and temporally (e.g. the length of the study). The spatial sample determines the scope of potential research questions while the length of study positively affects the detectability of regular phenological cycles (Bush et al. 2017) and phenological shifts (Chambers *et al.* 2013). Noise can be introduced to phenology data through both "process uncertainty" (how well we can predict ecological processes e.g. the regularity of phenological cycles) and "observation uncertainty" (how easily we can observe and record ecological events). Different life-cycle events and stages such as development of leaves, flower and fruits, even from the same species, may differ in regularity and/or ease of observation, leading to systematic biases in phenology recording related to the frequency and type of observations (see Regan at al. 2002 for a full description and "taxonomy" of the different uncertainties associated with ecological data).

To explore this further, we present hypothetical scenarios of crown phenology observations subject to different combinations of process and observation uncertainty and demonstrate how interpretation of the data without careful consideration of the source of noise could lead to erroneous conclusions. For species where a phenological event is easy to see (e.g. large, brightly coloured flowers that contrast with the leaf canopy or cauliflorous flowers on the trunk of the tree), most observations will be accurate and it will be straightforward to tell from the recorded data if the actual cycle is regular or irregular (Fig. 1A-C, Observation uncertainty = Low, Process uncertainty = Low: High). On the other hand, if for another species the same phenological event is difficult to see (e.g. flowers that are very small, held high in the canopy or persist for just a few days), data are likely to be recorded imperfectly and the cycle may appear irregular (Fig. 1G-H, Observation uncertainty = High, Process uncertainty = Low: High). Without quantifying the observation biases for these species, it will be impossible to differentiate if their actual cycles are regular or irregular as an inaccurately recorded regular cycle will look similar to an accurately recorded irregular cycle (Fig. 1G compared to Fig. 1C). This distinction is important as adaptive features of the phenological cycle itself and changes in predictability or synchrony will be of great interest to the global change community, whereas apparent irregularity derived from inaccurate observation will not. In this paper, we seek to quantify observation biases between species and phenophases at our study site, Lopé NP, Gabon, in order to direct precious resources where they are likely to give robust data and to include important sources of variation in future explanatory models of plant phenology and ecological change.

142 Quantifying observation uncertainty for different phenological events is not easy as there
143 are multiple sources of variation - specific to the phenology sampling method in question - that
144 lead to systematic observation biases. At Lopé, phenology monitoring takes the form of crown
145 observations, and variation in the "visibility" of phenological events and their "duration" are
146 likely to be key factors contributing to uncertainty. Visibility, however, is inherently subjective
147 from the point of view of the observers. For example, the size of a flower or fruit is likely to
148 influence how visible it is, but so will its colour, or the distance it is held from the observer (e.g.
149 a large green flower high up in the canopy may be less "visible" than a small, red flower lower in
150 the canopy or a cauliflorous flower growing from the tree trunk). In order to capture this
151 information many multiple axes of variation would need to be measured and then calibrated with
152 the observer experience. Such empirical data is not readily available and so instead, we sought to
153 describe the visibility and duration of phenology events using expert knowledge elicited from
154 long-term phenology observers at our site. These experts hold substantive knowledge of the
155 ecosystem based on their personal experience over many years of fieldwork at the site (Martin et
156 al. 2012).

Considering data for all species and phenological events (leaf, flower and fruit cycles, hereafter "phenophases") recorded as part of the Lopé long-term phenology study, alongside expert knowledge for observation uncertainty, we ask the following questions: (1) Can observation uncertainty be quantified? (2) Does observation uncertainty impact detectability of regular cycles among different species and phenophases? (3) What are the relative contributions of different sources of observation uncertainty to cycle detectability? We believe that the analysis presented here, using rare, long-term data, will help to improve resource allocation and sample design at other existing and emerging tropical phenology programs and aid robust assessments of phenological change in the future.

METHODS

THE LOPÉ LONG-TERM PHENOLOGY STUDY.— Since 1984, researchers at the Station d'Études des Gorilles and Chimpanzées (SEGC) in Lopé National Park, Gabon have recorded leaf, flower and fruit phenology monthly for 88 species of tropical trees and shrubs (>1000 individuals) spread over an area of 33km^2. The SEGC study area is situated in a tropical forest-savanna matrix with an equatorial climate characterised by two dry and two wet seasons annually (see Tutin & White 1998 for detailed site description). At the beginning of every month (usually completed within the first seven working days), SEGC researchers examine the crowns of each plant from the ground with 10 x 42 binoculars and record the proportion of the canopy covered by each phenophase (new and senescent leaves, flowers, unripe and ripe fruits) as a scale from zero to four (including half points; Tutin & Fernandez 1993b, Tutin & White 1998). The data recorded for each phenophase form multiple continuous time series for each individual tree. Data are only recorded autonomously by observers with >1 yr experience with the plant species involved and working under another observer. Data have been recorded by a total of only ten observers throughout the 387 mo (32 yr) of continuous observations, with individual observers making continuous contributions of 2-20 yr. Thus this dataset is likely to have minimal (but not zero) inter-observer biases.

DETECTING PHENOLOGY CYCLES USING FOURIER ANALYSIS.— We excluded data collected before 1986 when the project was being established and made selections for further analysis according to the following criteria; more than five years continuous data for each individual plant, no data gaps greater than three months, and no persistent records of disease (e.g. field comments referring to the ill-health of a tree consistently for more than a year). The resulting sample consisted of 4280 continuous time series for new and senescent leaves, flowers, unripe and ripe fruits from 856 individual plants of 70 species. The number of individuals per species ranged from 1 to 41, with a mean of 12, while the length of time individual plants were monitored ranged from 60 to 353 mo, with a mean of 249 mo.

To identify the dominant regular cycle for each time series in this sample we used Fourier analysis; Fourier is a form of spectral analysis based on sine and cosine waves that can be used to quantitatively describe the cyclic nature of any time series data (Bloomfield 2000). We used a confidence test, based on 95% confidence intervals and a null hypothesis of "no cyclicity", to determine if the dominant cycle was objectively different to surrounding noise. We refer to a "detected cycle" as one that can be quantified and considered significant according to this method. A full explanation of the Fourier methods used and our data selection criteria is given in Bush et al. (2017).

ELICITING EXPERT-KNOWLEDGE ON TWO MAJOR SOURCES OF OBSERVATION UNCERTAINTY.—

We gathered expert knowledge to describe the observation uncertainty associated with each phenophase for every species in our study. Following the recommendations of Martin et al. (2012), the authors of this study were assigned different (sometimes multiple) roles in the process of expert elicitation; EB, NB and KA acted as the "problem owners" defining the questions and design of the expert survey, while KA, LW, ED and CT were the "experts", each of whom had recorded phenology data at SEGC for more than 15 yr. EB and NB were the "analysts" and independently processed the expert responses and analysed the data.

EB and the station manager at SEGC facilitated the process of expert elicitation in February 2016. For ease of interpretation by all experts we chose to elicit knowledge on observation uncertainty in the form of categorical measures (Method 7, Kuhnert *et al.* 2010). The experts were independently presented with a survey listing all species monitored at SEGC and five phenophases (new and senescent leaves, flowers, unripe and ripe fruits) and asked to record their perception of both the visibility and duration for each. Phenophase visibility was presented as a score from one to three, representing events that are "Difficult to see", "Easy to see" and "Very obvious". Phenophase duration was presented as a binary category: "events lasting <= 4 wks" or "events lasting >4 wks" (the 4-wk interval corresponding to the field observation frequency). The observers were informed that they were allowed to leave an answer blank if they were unsure.

A correlation matrix for phenophase visibility showed that scores were positively correlated between all observer pairs, ranging from 0.13 to 0.38 (mean = 0.27; Fig.S1). To combine the expert judgements we took group averages (Martin *et al.* 2012) by calculating mean event visibility and modal duration category for each species-phenophase. We excluded 15 percent of species-phenophase visibility scores because fewer than three observers provided an answer, and 31 percent of species-phenophase duration scores because either fewer than three observers provided an answer or there was no clear majority (e.g. if two observers considered an event to last <= 4 wks and two observers considered an event to last > 4 wks). This may occur when the true event duration is around 4 wks and thus the phenophase cannot be easily assigned to either category.

MODELLING THE IMPACT OF OBSERVATION UNCERTAINTY ON CYCLE DETECTION AMONG PHENOLOGY DATA.— To compare how different sources of observation uncertainty contribute to variation in cycle detectability we combined the data derived from the 4280 times series used in Fourier analysis with the observer scores for phenophase visibility and duration. We only included species with more than three observed individuals and complete information on phenophase visibility and duration, resulting in a final sample of 3083 time series from 827 individuals (61 species). Before analysis, we standardized predictors by scaling them to mean = 0 and standard deviation = 2 to allow meaningful comparison of effect sizes (Schielzeth 2010; Table 1).

To test the effects of phenophase visibility (**Visibility $_{Scaled}$**) and phenophase duration (**Duration**) on the likelihood of detecting a cycle we used a Generalized Linear Mixed Model (GLMM, family = binomial, link = logit). As we already know time series length is an important positive predictor of cycle detection (Bush et al. 2017) we included it as a fixed effect in the model (**Length $_{Scaled}$**).

In our mixed model we included the grouping factors tree **ID**, **Species** and **Phenophase** as random intercepts and all continuous predictors as random slopes by **Phenophase**. First, this reflected the hierarchical nature of the data (multiple phenophases simultaneously recorded per individual tree; duration and visibility scored at the level of the species-phenophase) and second, it allowed us to take account of the biological differences (process uncertainties) between species and phenophases.

We followed a model simplification process starting with the maximal model for both fixed effects (all possible pair-wise interactions between predictors) and random effects (random slope by **Phenophase** for all continuous predictors), removing each term in a step-wise fashion and then comparing resulting models using AIC values. We used the standardised effect sizes derived from the final, most parsimonious model to compare between predictors. We temporarily modified the final model by removing terms for the intercept and the main effect of continuous predictors involved in interactions to determine if predictor effect sizes were different to zero (95% confidence intervals derived from standard errors; Schielzeth 2010).

RESULTS

OVERVIEW OF DOMINANT PHENOLOGY CYCLES.— Using all available phenology time series from the Lopé long-term study after selection criteria, we confidently detected regular cycles from 36 percent of the sample (total number of time series = 4280, 5 different phenophases from 856 individuals). However, detection differed among phenophases, being highest for flowers (59%) and unripe fruit (54%) and lowest for ripe fruit (29%) and senescing leaves (25%). Annual cycles were most commonly detected among reproductive data (75% all detected cycles for flowers, unripe and ripe fruits were annual), while sub-annual cycles were most commonly detected from vegetative data (51% all detected cycles for new and senescing leaves were sub-annual).

OBSERVATION UNCERTAINTY SCORES.— The inter-quartile ranges for the visibility scores of all phenophases overlapped (Fig. 2a) but on average, new leaves were considered the most visible (mean score = 2.42) and flowers the least visible (mean score = 2.08). In contrast, event duration scores were not evenly distributed among phenophases (Fig. 2b); Unripe fruit events were perceived as lasting > 4 wks for almost all species (65 / 66 species) while new leaf events were perceived as lasting <= 4 wks for all species.

EFFECTS OF OBSERVATION UNCERTAINTY ON CYCLE DETECTION.— After model simplification, all of the main predictors and an interaction between **Length** $_{\text{Scaled}}$ and **Duration** were retained in the most parsimonious model (Table S1). We found both **Length** $_{\text{Scaled}}$ and **Visibility** $_{\text{Scaled}}$ to have significant positive effects (95% confidence intervals different to zero; Fig. 3a and Table S2) on the likelihood of detecting a cycle from our phenology data. The relative effect of **Visibility** $_{\text{Scaled}}$ (standardised effect size = 0.79) was almost half that of **Length** $_{\text{Scaled}}$ when **Duration** <= 4 weeks (standardised effect size = 1.51), and a third of **Length** $_{\text{Scaled}}$ when **Duration** > 4 weeks (standardised effect size = 2.31; Fig. 3a). Model predictions from the final model showed that when a phenophase event lasted <= 4 weeks, the likelihood of detecting a regular cycle was 0.23 after 10 yr of data collection and 0.39 after 20 yr. If the phenophase event lasted > 4 wks, the likelihood of detecting a regular cycle after 20 yr of data collection increased to 0.47 (Fig. 3b). We also found that for the least visible phenophase events (score = 1) the likelihood of detecting a regular cycle was 0.26 when the phenophase event lasted <= 4 wks and 0.34 when the phenophase event lasted > 4 wks. For the most visible phenophase events (score = 3), this increased to 0.5 when the phenophase event lasted <= 4 wks and 0.58 when the phenophase event lasted > 4 wks (Fig. 3c).

EFFECTS OF PROCESS UNCERTAINTY ON CYCLE DETECTION.— The random intercepts for **Phenophase** and **Species** accounted for most of the variance in the data (23% and 25%, respectively) while tree **ID** accounted for the least (<0.04%; see Table S3 for variance and standard deviation). The likelihood of detecting a cycle varied by **Phenophase**, being most likely for flowers and least likely for senescing leaves and ripe fruits. While for unripe fruits and new leaves, likelihood of detecting a cycle was greater than, but very similar to, the intercept for the fixed effects model (Fig. 4 and Table S4). In the most parsimonious model, a random slope term by **Phenophase** was retained for **Length** $_{\textbf{Scaled}}$. The effect of **Length** $_{\textbf{Scaled}}$ as a predictor of cycle detectability was positive for all phenophases (Fig. 4a), however, the effect was more positive than the general trend for new leaf, and flower cycles and less positive than the general trend for unripe fruit cycles (Table S4).

DISCUSSION

We have shown that experienced field observers can provide important information on major sources of noise in phenology monitoring and that this can improve explanatory power for analyses of complex phenological data. For data derived from crown observations, we found that time series length, phenophase event visibility and duration are all good, positive predictors of finding regular phenological cycles. However, a relative increase in time series length has up to three times as large an effect on likelihood of detecting a cycle as a similar increase in phenophase visibility (comparison of standardised effect sizes).

311 The hierarchical nature of our modelling approach, including both species and
312 phenophases, also allowed us to investigate variation in cycle detectability due to biological
313 differences (process uncertainty). Species is an important predictor of cycle detectability, with
314 some species - such as *Duboscia macrocarpa, Detarium macrocarpum* and *Saccoglottis*
315 *gabonensis* - much more likely to have highly regular cycles among all phenophases than the
316 general trend. We also found that cycle detectability varies among phenophases and is highest
317 for flowers, followed by new leaves and unripe fruits and lowest for senescing leaves and ripe
318 fruit. It is interesting to note, that among reproductive phenophases, detectability is highest for
319 flowers, then unripe fruits, then ripe fruits. The fact that flowers occur first in a chain of linked
320 events is likely to contribute to this pattern, as there are fewer accumulated opportunities for
321 ecological processes to contribute noise at this stage. For instance more regular flowering than
322 fruiting could arise because trees may abort their reproductive efforts after poor pollination or
323 unfavourable weather conditions, or because of widespread removal of flowers by florivores
324 (e.g. red colobus monkeys, *Procolobus rufomitratus,* in Uganda; Chapman et al. 2013).

325 LESSSONS LEARNED FOR EFFECTIVE ANALYSIS OF LONG-TERM TROPICAL PHENOLOGY DATA.— The
326 information gained from this study can help guide us to more effective data collection and more
327 robust statistical analyses of tropical phenology; namely the best ways to increase sample size
328 and reduce noise.

Our first conclusion is that differences in observation uncertainty among species for the same phenophase should be accounted for in explanatory models of phenology data, otherwise the error associated with observational differences may lead to misleading conclusions. For example, it would be possible to erroneously link some aspect of leafing phenology to the functional group of the species (e.g. shade-tolerant, long-lived species) when in reality it could have arisen from an observation bias, such as visibility, associated with those traits (Figure 1). There have been a number of calls for more quantitative assessment of the impacts of climate on tropical phenology (Butt *et al.* 2015, Mendoza *et al.* 2017) and to correct the temperate (Northern hemisphere) bias of current climate change studies (Feeley *et al.* 2016). We have shown that even a simple assessment of observation uncertainty, undertaken by experienced field observers, can provide important information and be incorporated into and improve quantitative analyses of existing tropical phenology data.

LESSONS LEARNED FOR THE DESIGN OF TROPICAL PHENOLOGY MONITORING PROGRAMS.— Going forward, we propose that both established and new programs seek to minimise sources of noise in phenology sampling design. We have shown that the length of study is the most important predictor of cycle detectability; thus it is vitally important that resources be directed towards maintaining existing and emerging long-term monitoring programs. For all phenophases, an additional 10 yr of data collection (from 10 to 20 yr) increases likelihood of detecting a cycle by 70 percent for phenology events lasting <= 4 wks and by 114 percent for events lasting > 4 wks. Observation length is a source of uncertainty relevant to all phenology sampling methods (e.g. both crown observations and traps) and clearly, the elusive "sticking power" necessary to ensure long-term data collection needs to be addressed in the tropics. This can be achieved either through recognition of the importance of phenology research and allocation of substantial long-term resources from tropical nations and international funders, or through relevant and innovative, citizen-based initiatives.

While increasing the length of observation has the largest relative effect on cycle detectability, it is not always practicable. Often the duration of monitoring programs is outside of scientists' control, or assessments are necessary in the short-term and cannot wait an additional ten years. Therefore, for new monitoring programs looking to make meaningful assessments of cycle regularity through canopy observations over a short time, we recommend that they target species with highly visible phenological events that last longer than the monitoring interval (in our case, monthly). For example, at Lopé, flowers from species *Beilschmedia fulva*, *Milica excela* and *Mammea africana* are difficult to see (visibility score <1.5) and last < 4 wks, whereas flowers from species *Antidesma vogelianum, Mangifera indica* and *Omphalocarpum procerum* are very easy to see (visibility score >2.5) and persist in the canopy for > 4 wks. Data from the latter species are more likely to be robust and free from observation error (similar to the scenarios for "low" observation uncertainty from Figure 1). After a period of initial monitoring (at least 5 yr) it will be possible for data collectors at study sites to assess the amount of noise associated with specific species and phenophases in their sample. This information would allow project managers to select directly for the most easily observed species and target limited resources towards them by increasing sample sizes and including such species in inter-site comparisons.

Inevitably there will be occasions when it is important to monitor a noisy species or phenophase. For example, Moabi (*Baillonella toxisperma*) nuts are an important source of oil for cooking, cosmetics and rural enterprises in central Africa (Plenderleith & Brown 2004) but Moabi trees exhibit irregular phenology (random intercept for cycle detectability = -0.50) and its flowers are difficult to see, as they are small and held very high in the canopy (similar to scenarios for "high" process uncertainty and "high" observation uncertainty, Figure 1). In such cases it will be important to tailor observation programs accordingly, by investing in alternative forms of monitoring (e.g. installing cameras in tree canopies opposed to observations from the ground), increasing number of trees monitored or increasing the frequency of observations.

Any systematic biases in recording phenology data will of course be related to the sampling method, "visibility" and "duration" being key sources of uncertainty identified for crown observation sampling protocols. The duration of a phenophase may be of less concern for trapping methods, although different biases are likely to arise such as rate of decomposition and trap-checking frequency, or the relative influence of weather conditions such as strong winds on the deposition of plant material. If used concurrently, crown observations and trapping methods could prove to be complimentary, accounting for different sources of uncertainty particular to each. In particular, seed traps employed alongside canopy monitoring could be used to further quantify the duration of phenological events.

The scientific community hopes to assess climate-induced changes in tropical ecological processes with only decades-long data at their disposal (for example, the data analysed here represents the longest published continuous phenology dataset in the tropics). This expectation has been raised by the rate of change observed in temperate systems (Schwartz *et al.* 2006). However, with such limited data it is essential that variation associated with processes outside of the focal question be kept to a minimum. When allocating resources for new and ongoing research, phenologists should aim to maintain monitoring programmes for as long as possible and target species and phenophases with least inherent noise to maximise statistical power and therefore ability to assess change in future analyses.

ACKNOWLEDGEMENTS

Phenology research at SEGC, Lopé National Park was funded by the International Centre for Medical Research in Franceville (CIRMF) (1986-2010) and by Gabon's National Parks Agency (ANPN) (2010 – present). EB was supported by an Impact Studentship funded by the University of Stirling and ANPN. NB received funding from the European Research Council under the European Union's H2020/ERC grant agreement no 679651 (ConFooBio). We acknowledge significant periods of independent data collection undertaken by Richard Parnell, Liz Williamson, Rebecca Ham, Patricia Peignot and Ludovic Momont. Permission to conduct this research in Gabon was granted by the CIRMF Scientific Council and the Ministry of Water and Forests (1986 – 2010), and by ANPN and the National Centre for Research in Science and Technology (CENAREST) (2010 – present). We thank Alistair Jump, Daisy Dent, Isabel Jones, Jeremy Cusack and Irene Mendoza whose comments in preparation significantly improved the manuscript.

DATA AVAILABILITY:

All Lopé long-term phenology data are archived in Gabon within the Gabon National Parks Agency (ANPN) archive and in the UK within the University of Stirling's Online Repository for Research Data (DataSTORRE; http://hdl.handle.net/11667/103). The monthly fruit, flower and leaf phenophase scores are subject to a 10-year embargo imposed by the Government of Gabon, however, applications for access may be made to ANPN (science@parcsgabon.ga) and will be considered on a case-by-case basis. The derived Fourier outputs and observation uncertainty scores used in this paper are available immediately at DataSTORRE (http://hdl.handle.net/11667/92).

LITERATURE CITED

ADAMESCU, G.S., A. J. PLUMPTRE, K. A. ABERNETHY, L. POLANSKY, E. R. BUSH, C. A. CHAPMAN, L. P. SHOO, A. FAYOLLE, K. R. L. JANMAAT, M. M. ROBBINS, H. J. NDANGALASI, N. J. CORDEIRO, I. C., GILBY, R. M. WITTIG, T. BREUER, M. BREUER-NDOUNDOU HOCKEMBA, C. M. SANZ, D. B. MORGAN, A. E. PUSEY, B. MUGERWA, B. GILAGIZA, C. TUTIN, C. E.N. EWANGO, D. SHEIL, E. DIMOTO, F. BAYA, F. BUJO, F. SSALI, J.T. DIKANGADISSI, K. JEFFERY, K. VALENTA, L. WHITE, M. MASOZERA, M. L. WILSON, R., BITARIHO, S. T. NDOLO EBIKA, S. GOURLET-FLEURY and C. M. BEALE. Annual cycles dominate reproductive phenology of African tropical trees. Biotropica, this issue.

ADOLE, T., J. DASH, and P. M. ATKINSON. 2016. A systematic review of vegetation phenology in Africa. Ecol. Inform. 34: 117–128.

AONO, Y., and K. KAZUI. 2008. Phenological data series of cherry tree flowering in Kyoto, Japan, and its application to reconstruction of springtime temperatures since the 9th century. Int. J. Climatol. 28: 905–914.

BLOOMFIELD, P. 2000. Fourier analysis of time series: An introduction. John Wiley & Sons.

BUSH, E. R., K. A. ABERNETHY, K. JEFFERY, C. TUTIN, L. WHITE, E. DIMOTO, J. T. DIKANGADISSI, A. S. JUMP, and N. BUNNEFELD. 2017. Fourier analysis to detect phenological cycles using tropical field data and simulations. Methods Ecol Evol, 8: 530–540. doi:10.1111/2041-210X.12704

.

BUTT, N., L. SEABROOK, M. MARON, B. S. LAW, T. DAWSON, J. SYKTUS, and C. MCALPINE. 2015. Cascading effects of climate extremes on vertebrate fauna through changes to low-latitude tree flowering and fruiting phenology. Glob. Chang. Biol. 21: 3267–3277.

CHAMBERS, L. E., R. ALTWEGG, C. BARBRAUD, P. BARNARD, L. J. BEAUMONT, R. J. M. CRAWFORD, J. M. DURANT, L. HUGHES, M. R. KEATLEY, M. LOW, P. C. MORELLATO, E. S. POLOCZANSKA, V. RUOPPOLO, R. E. T. VANSTREELS, E. J. WOEHLER, and A. C. WOLFAARDT. 2013. Phenological Changes in the Southern Hemisphere. PLoS One 8: e75514.

CHAPMAN, C., T.R. BONNELL, R. SENGUPTA, T.L. GOLDBERG and J.M. ROTHMAN. 2013. Is *Markhamia lutea*'s abundance determined by animal foraging? Forest ecology and management. 308: 62-66.

CHAPMAN, C.A., K. VALENTA, T.R. BONNELL, K. A. BROWN and L.J. CHAPMAN. Solar radiation and ENSO predict fruiting phenology patterns in a 16-year record from Kibale National Park, Uganda. Submitted, 2017, this issue.

CLELAND, E. E., I. CHUINE, A. MENZEL, H. A. MOONEY, and M. D. SCHWARTZ. 2007. Shifting plant phenology in response to global change. Trends Ecol. Evol. 22: 357–365.

FEELEY, K. J., J. T. STROUD, and T. M. PEREZ. 2016. Most "global" reviews of species' responses to climate change are not truly global. Divers. Distrib. 1–4.

HUMBOLDT-UNIVERSITY OF BERLIN. 2012. International Phenological Gardens. Available at: https://www.agrar.hu-berlin.de/en/institut-en/departments/dntw-en/agrarmet-en/phaenologie/ipg/ipg_allg-e [Accessed January 5, 2017].

KUHNERT, P.M., T.G. MARTIN and S.P. GRIFFITHS. 2010. A guide to eliciting and using expert knowledge in Bayesian ecological models. Ecol. Lett. 13(7): 900-914.

MARTIN, T. G., M.A. BURGMAN, F. FIDLER, P.M. KUHNERT, S. LOW-CHOY, M. MCBRIDE, and K. MENGERSEN. 2012. Eliciting Expert Knowledge in Conservation Science. Conserv. Biol., 26:29-38.

MENDOZA, I., C. A. PERES, and L. P. C. MORELLATO. 2017. Continental-scale patterns and climatic drivers of fruiting phenology: A quantitative Neotropical review. Glob. Planet. Change 148: 227–241.

MORELLATO, L. P. C., B. ALBERTON, S. T. ALVARADO, B. BORGES, E. BUISSON, M. G. G. CAMARGO, L. F. CANCIAN, D. W. CARSTENSEN, D. F. E. ESCOBAR, P. T. P. LEITE, I. MENDOZA, N. M. W. B. ROCHA, N. C. SOARES, T. S. F. SILVA, V. G. STAGGEMEIER, A. S. STREHER, B. C. VARGAS, and C. A. PERES. 2016. Linking plant phenology to conservation biology. Biol. Conserv. 195: 60–72.

PAU, S., E. M. WOLKOVICH, B. I. COOK, C. J. NYTCH, J. REGETZ, J. K. ZIMMERMAN, and S. JOSEPH WRIGHT. 2013. Clouds and temperature drive dynamic changes in tropical flower production. Nat. Clim. Chang. 3: 838–842.

PLENDERLEITH, K., and N. BROWN. 2004. Moabi (*Baillonella toxisperma*). *In* L. E. Clark and T. C. H. Sunderland (Eds.) The key non-timber forest products of Central Africa: State of the knowledge. pp. p141-162, USAID, Bureau for Africa, Office of Sustainable Development. Washington, D.C.

PRIMACK, R. B., H. HIGUCHI, and A. J. MILLER-RUSHING. 2009. The impact of climate change on cherry trees and other species in Japan. Biol. Conserv. 142: 1943–1949.

REGAN, H.R., M. COLYVAN and M. BURGMAN. 2002. A Taxonomy and Treatment of Uncertainty for Ecology and Conservation Biology. Ecol. Appl.. 12: 618-628.

SCHIELZETH, H. 2010. Simple means to improve the interpretability of regression coefficients. Methods Ecol. Evol. 1: 103–113.

SCHWARTZ, M. D., R. AHAS, and A. AASA. 2006. Onset of spring starting earlier across the Northern Hemisphere. Glob. Chang. Biol. 12: 343–351.

SINGH, K. P., and C. P. KUSHWAHA. 2005. Emerging paradigms of tree phenology in dry tropics. Curr. Sci. 89.

TUTIN, C. E. G., and M. FERNANDEZ. 1993. Relationships between minimum temperature and fruit production in some tropical forest trees in Gabon. J. Trop. Ecol. 9: 241–248.

TUTIN, C. E. G., M. FERNANDEZ, M. E. ROGERS, E. A. WILLIMSON, and W. C. MCGREW. 1991. Foraging profiles of sympatric lowland gorillas and chimpanzees in the Lope Reserve Gabon. Philos. Trans. R. Soc. London B 334: 179–186.

TUTIN, C. E. G., and L. J. T. WHITE. 1998. Primates, phenology and frugivory: Present, past and future patterns in the Lope Reserve, Gabon. *In* D. M. Newbery, H. H. T. Prins, and N. Brown (Eds.) Dynamics of Tropical Communities: 37th Symposium of the British Ecological Society. pp. 309–338, Blackwell Science, Oxford.

USA NATIONAL PHENOLOGY NETWORK (USA-NPN). 2017. Nature's Notebook. Connecting people with nature to benefit our changing planet. Available at: https://www.usanpn.org/natures_notebook# [Accessed January 5, 2017].

506 **TABLES**

507 TABLE 1. Key summary statistics and definitions of all predictors included in the maximal
508 model.

Variable	**Definition**	**Level** [a]	**Summary statistics** [b]	**Predictor** [c]
Time series length	Length of continuous observation of tree in months.	ID	Continuous (Mean = 251; SD = 93.9; Min = 60.0; Max = 353)	$Length_{Scaled}$
Phenophase visibility	Mean observer score for visibility, 1 (Difficult to see) to 3 (Very obvious).	Sp-Ph	Continuous (Mean = 2.27; SD = 0.43; Min = 1.0; Max = 3.0)	$Visibility_{Scaled}$
Phenophase duration	Modal observer score for duration, 0: <= 4 wks, 1: > 4 wks.	Sp-Ph	Categorical (0 = 59%)	Duration

509

510 [a] Level of data-collection for variable (ID = Tree ID; Sp-Ph = Species-phenophase)

511 [b] Summary statistics for variables included in binomial GLMM pre-scaling

512 [c] Code for scaled predictor included in binomial GLMM

513

514

515 **FIGURE LEGENDS**

516 FIGURE 1. Simulated data scenarios to show the impacts of both observation and process
517 uncertainty on recorded time series (observed scores: red solid line) compared to real time series
518 (actual scores: gray dashed line). Observation uncertainty is "low" when a phenological event is
519 easy to see and the recorded time series closely matches the real time series and "high" when a
520 phenological event is difficult to see leading to many missing observations and a recorded time
521 series that does not closely match the real time series. Process uncertainty is "low" when
522 phenological events occur in clean, regular cycles and are easy to predict and is "high" when
523 phenological events occur in noisy, irregular cycles and are difficult to predict. From the
524 recorded time series alone, it is impossible to differentiate between a record with high process
525 uncertainty but low observation uncertainty (bottom left) and a record with low process
526 uncertainty but high observation uncertainty (top right).

527 FIGURE 2. Summary of mean scores calculated for phenophase event visibility and duration by
528 phenophase. (A) Distribution of mean phenophase visibility scores for each species showing
529 median scores (bold horizontal lines), the interquartile range (coloured boxes) and the 95% range
530 (vertical lines) and the breaks between each score (horizontal dashed lines). (B) Percentage of
531 species categorised according to phenophase duration: events lasting <= 4 wks (light shading)
532 and events lasting > 4 wks (dark shading).

FIGURE 3. Standardised effect sizes (A) and predictions (B and C) for all predictors retained in the final model. (A) The standardised effect of each predictor on the likelihood of detecting a regular cycle (filled black circles) and 95% confidence intervals (horizontal black lines) to show whether effect is significantly different to zero (derived from a modified final model with intercept and main effect for $\text{Length}_{\text{Scaled}}$ temporarily removed). (B-C) Model predictions for the relationship between the significant predictors - time series length, phenophase visibility (both continuous) and phenophase duration (binary) - and the likelihood of detecting a significant cycle from phenology data.

FIGURE 4. Predictions from the final model by phenophase. General model predictions (grey lines) and 95% confidence intervals (grey shades) show the relationship between both significant continuous predictors (A) time series length and (B) visibility and the likelihood of detecting a significant cycle from phenology data when phenophase events last <= 4 weeks. Predictions from the random intercept and slope terms show how model predictions vary by phenophase (coloured lines). The mean probability of detecting a cycle from binned raw data demonstrates the model fit (coloured dots, scaled by number of observations in each bin).