Methods in Ecology and Evolution

Methods in Ecology and Evolution 2017, **8,** 530–540 doi: 10.1111/2041-210X.12704

Fourier analysis to detect phenological cycles using long-term tropical field data and simulations

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

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

Summary

1. Changes in phenology are an inevitable result of climate change, and will have wide-reaching impacts on species, ecosystems, human society and even feedback onto climate. Accurate understanding of phenology is important to adapt to and mitigate such changes. However, analysis of phenology globally has been constrained by lack of data, dependence on geographically limited, non-circular indicators and lack of power in statistical analyses.

2. To address these challenges, especially for the study of tropical phenology, we developed a flexible and robust analytical approach – using Fourier analysis with confidence intervals – to objectively and quantitatively describe long-term observational phenology data even when data may be noisy. We then tested the power of this approach to detect regular cycles under different scenarios of data noise and length using both simulated and field data.

3. We use Fourier analysis to quantify flowering phenology from newly available data for 856 individual plants of 70 species observed monthly since 1986 at Lopé National Park, Gabon. After applying a confidence test, we find that 59% of the individuals have regular flowering cycles, and 88% species flower annually. We find time-series length to be a significant predictor of the likelihood of confidently detecting a regular cycle from the data. Using simulated data we find that cycle regularity has a greater impact on detecting phenology than event detectability. Power analysis of the Lopé field data shows that at least 6 years of data are needed for confident detection of the least noisy species, but this varies and is often >20 years for the most noisy species.

4. There are now a number of large phenology datasets from the tropics, from which insights into current regional and global changes may be gained, if flexible and quantitative analytical approaches are used. However, consistent long-term data collection is costly and requires much effort. We provide support for the importance of such research and give suggestions as to how to avoid erroneous interpretation of shorter length datasets and maximise returns from long-term observational studies.

Key-words: circular analysis, climate change, flowering, Gabon, Lopé National Park, phenophases, spectral analysis, time-series data, tropical forests

Introduction

Phenology concerns the timing of recurring life cycle events – such as leaf growth, flowering and fruiting in plants – and has long fascinated ecologists and evolutionary scientists. Questions range from understanding the complex environmental cues and internal mechanisms that initiate phenology events (phenophases) to the adaptive significance of their timing and duration and responses to environmental change. Phenology has wide-reaching influence within ecosystems and determines the nature of many interspecific interactions (Butt *et al.* 2015). Changes in global climate will inevitably have long-term impacts on phenology (Parmesan 2006) with knock-on effects for ecosystems and people (van Vliet 2010). It is also clear that there will be feedbacks between changing phenology and climate, but they are poorly characterised by current climate models (IPCC 2014).

TROPICAL PHENOLOGY OVERLOOKED IN REVIEWS OF CHANGE

Major reviews of phenological change to date have lent heavily on evidence from temperate, especially Northern hemisphere, regions (Parmesan 2006; Cleland *et al.* 2007; Chambers *et al.* 2013). In these regions more phenology data is available and analyses are arguably simpler. The strong seasonality in temperate regions accompanied by a dormant winter season results in broad synchronisation of phenology on the annual cycle. Years can be treated to some extent as independent repeating events and researchers are able to make use

*Correspondence author. E-mail: e.r.bush@stir.ac.uk

of a relatively simple suite of 'spring indicators' (e.g. first appearance, first lay-date, bud-burst measured in days since 1 January).

While tropical climates are often seasonal, annual variation is more limited than in temperate regions and vegetative growth and reproduction are possible at any time of the year resulting in more diverse phenology and cycles other than twelve months (van Schaik, Terborgh & Wright 1993). Use of simple spring indicators is not appropriate for tropical phenology because of the ***circularity of the data*** (e.g. 1 January is an arbitrarily low value and not meaningfully different from 31 December).

Furthermore, phenology is subject to many conflicting demands, for example an organism may receive an environmental signal to reproduce but fail to do so because it lacks critical resources (Obeso 2002). Inconsistencies and gaps in data collection due to observation error are also common in long-term studies, making quantification in many cases harder still. Thus analytical approaches for tropical phenology need to take account of the circularity of the data, be flexible, quantitative and attribute confidence to conclusions.

ANALYSES OF LONG-TERM TROPICAL PLANT PHENOLOGY

Published analyses of tropical plant phenology range from simple descriptions and correlations with environmental variables to more recent, quantitative analyses of change (Appendix S1, Supporting Information). The Newstrom, Frankie & Baker (1994) framework was an important step towards objective inter-site comparisons, however, categorisation loses analytical power and visual comparisons lack objective rigour. More computationally intensive methods have included differentiation of species-level reproductive cycles using finite mixture theory and bootstrapping methods (Cannon *et al*. 2007), modelled autocorrelation functions (Norden *et al*. 2007), sinusoid-based regression (Anderson *et al*. 2005), spectral analysis (Chapman *et al*. 1999), circular statistics (Wright & Calderon 1995; Wright *et al*. 1999; Zimmerman *et al*. 2007; Ting, Hartley & Burns 2008), generalised linear models (GLMs) (Newbery, Chuyong & Zimmermann 2013) and generalised additive mixed models (GAMMs) (Polansky & Robbins 2013). While data has often been collected at the scale of the individual plant (9/18 studies in Appendix S1), this is not always reflected in analysis where individuals are clumped into species, guilds or a percentage score of a whole community, losing power and precluding vital covariate information. The longest tropical phenology data set analysed to date is 22 years of flowering data (Pau *et al*. 2013) and 18 years of flowering and fruiting data (Wright & Calderon 2006) from Barro Colorado Island, Panama with many other studies relying on fewer than 10 years data (9/18 studies in Appendix S1).

Addressing the challenges of sample size, data quality, circularity and pseudo-replication is of paramount importance to quantify tropical phenology and compare between sites and over time. Consensus as to the most suitable way to analyse these data, what length of data is necessary to identify cycles and how to attribute confidence to results has been missing, although progress is being made (Hudson & Keatley 2010).

In this article, we apply statistical theory to both field and simulated data, to develop and demonstrate objective methods – based on ***Fourier analysis*** – to detect and quantify confidence in regular phenological cycles. We also test the likelihood of detecting cycles under different data noise and length scenarios and discuss opportunities for incorporating the resulting insights into research and policy. Explanations of technical terms related to Fourier analysis used in this paper are given in the glossary in Table 1 and their first use in the text is indicated in bold italics.

INTRODUCTION TO FOURIER ANALYSIS FOR PHENOLOGY

The Fourier transform is a mathematical method used to identify regular ***cycles*** in ***time-series*** data by comparing fluctuations in the data with ***sinusoids*** (Bloomfield 2000) and has been used extensively in disciplines such as engineering and mathematics. The Fourier transform calculates the tendency (hereafter known as ***power***) of all possible cycles to appear in the data and can therefore be used to quantify seasonal phenology data without the need for prior knowledge or hypotheses of ***cycle length***. However, it has been rarely used in the context of phenology analysis and never for long-term observational phenology data. Chapman *et al*. (1999) used Fourier to identify dominant reproductive cycles from 6 years of data for a tropical tree community, but did not use a confidence test. More recently Zalamea *et al*. (2011) used Fourier to identify flowering cycles from reconstructed 12-month series of herbarium data for a genus of neotropical tree, attributing confidence to cycles using a bootstrapping method.

Compared to other data for which Fourier has been used, phenology data are often comparatively short and collected at low resolution due to the costs and effort incurred. However, in the field of movement ecology, Wittemyer *et al*. (2008) and Polansky *et al*. (2010) successfully used Fourier to confidently identify regular cycles in animal movements by comparing outputs with a null hypothesis of random movement and 95% confidence intervals.

In this paper, we build on Wittemyer *et al*.'s (2008) analytical framework to extend the existing uses of Fourier for the field of long-term phenology research. First we demonstrate appropriate application of Fourier to phenology data by quantifying flowering cycle confidence, length, power, timing and ***synchrony*** for individuals of a single species from the Lopé long-term observational study of tropical forest plants (1986–2016). Second, we up-scale this Fourier-based approach to analyse flowering phenology using newly available data for all species from the Lopé study (856 individuals, 70 species). Third, we recognise that while the Lopé study is one of the longest and most consistent of its kind in the tropics, data is still often noisy or short for certain individuals and/or species. To apply this framework elsewhere, and to inform best practice for data collection, we test the ability of the Fourier method to

Table 1. Glossary to technical terms

Term	Definition
Bandwidth	The distance at which two peaks in the ***periodogram*** can be distinguished from each other, a quantitative measure of ***resolution***. For example a bandwidth of 0·1 means that cycles can be distinguished from each other when the difference between their frequencies is at least 0·1
Circular mean	A mean value calculated for ***circular data*** where the arithmetic mean would be inappropriate. For example the circular mean of 5° and 355° is 0°, in comparison to the arithmetic mean which is 180°
Circular standard deviation	A measure of dispersion calculated for ***circular data*** where the arithmetic standard deviation would be inappropriate
Circular data	Data from circular distributions (e.g. months, hours, directions, etc.) where there is no true zero and 'high' and 'low' values are arbitrary (e.g. Fig. 1a)
Co-Fourier analysis	Simultaneous ***Fourier analysis*** of two ***timeseries***. Additional outputs include relative ***phase difference*** between the timeseries at every possible ***cycle*** (Fig. 1e)
Cycle	A pattern of repeating events in a regular order
Cycle length/Wavelength	The time taken for a whole ***cycle*** to repeat itself (e.g. number of months between repeating flowering events)
Daniell kernel	A moving-average smoother used to eliminate fine detail from the ***raw spectral estimate*** to make the output more ***stable*** and easier to interpret (e.g. ***smoothed spectral estimate*** in Fig. 1c)
Dominant cycle	The ***cycle length*** associated with the ***dominant peak***.
Dominant peak	The point in the ***spectral estimate*** with highest ***power***
Fourier analysis	Decomposition of a ***timeseries*** into a series of ***sinusoidal*** functions. The ***power*** of each ***cycle*** in the series can be used to identify ***dominant cycles*** (Fig. 1c)
Frequency	The rate at which something occurs (e.g. number of flowering cycles per month or per year)
Null continuum	A ***spectral estimate***, derived from the data series, that has been smoothed extensively so that only the underlying shape remains, and no fine detail can be identified (Fig. 1d)
Periodogram	The visual output of the ***spectral estimate*** derived from ***Fourier analysis*** (Fig. 1c,d)
Phase difference	The distance between the peaks in two ***cycles*** of matching ***frequency*** and referenced in time (Fig. 1e)
Power	The relative tendency of all possible ***cycles*** to appear in the data. Estimated in the ***spectral estimate*** and plotted in the *y*-axis of a ***periodogram*** (Fig. 1c). Cycles not well supported by the data have low power, whereas cycles well supported by the data have high power
Radians	The standard unit of angular measures; 2π radians = 360°
Raw spectral estimate	The default output of ***Fourier analysis*** where all fine-scale structure is included, and can be overly influenced by certain segments of the data
Resolution	The ability to represent fine structure and distinguish between close peaks in the ***spectral estimate*** derived from ***Fourier***, quantified as the ***bandwidth*** (Bloomfield 2000). Spectral estimates with high resolution will show all peaks including minor ones, where as spectral estimates with very low resolution may show no peaks at all, but rather the general shape of the data (e.g. the ***null continuum*** in Fig. 1d). Increased resolution reduces stability and *vice versa*
Sinusoid/Sine wave/ Cosine wave	A smooth repeating pattern occurring every 2π radians (or 360°) (e.g. the simulated curve in Fig. 1e)
Smoothed spectral estimate	The output of ***Fourier analysis*** after a moving-average smoother is applied to the ***raw spectral estimate*** (Fig. 1c,d)
Spans	The user-specified widths of the ***Daniell kernel*** smoother, specifically how many data points are used to smooth the ***spectral estimate*** in each local window
Spectral estimate/ Spectrum	The output of ***Fourier analysis*** showing the tendency of all possible ***cycles*** to appear in the data, from twice the observation interval to the full length of the series (Fig. 1c,d)
Stability	Extent to which small fluctuations in certain segments of the data influence the ***spectral estima***te derived from ***Fourier***. Greater stability reduces ***resolution*** and *vice versa* (Bloomfield 2000)
Synchrony	The simultaneous occurrence of two or more events
Timeseries	A sequence of data points arranged in time order

detect regular phenology under different scenarios using both simulated data and field data with realistic noise.

How to detect and describe flowering cycles using Fourier analysis

THE LOPÉ LONG-TERM OBSERVATIONAL PHENOLOGY STUDY

Since 1986, researchers from the Station d'Études des Gorilles and Chimpanzées (SEGC), Lopé National Park, Gabon, have observed individual plants of 88 different species each month and noted the proportion of each canopy covered by new, mature and senescing leaves, flowers, unripe and ripe fruits. Canopy coverage for a particular phenophase is assessed from the ground using binoculars and recorded as a score from 0 to 4. The study area experiences an equatorial climate, where seasonality is determined by movements of the inter-tropical convergence zone to form two dry and two wet seasons annually. See Tutin & White (1998) for detailed site description including local climate and vegetation.

In this first section we demonstrate Fourier analysis using flowering data for tree species *Duboscia macrocarpa* Bocq. (Malvaceae, $n = 11$). Initial observation of species-level data shows no apparent seasonality in flowering (Fig. 1a,b). However, this is because the true flowering cycle for this species is 18 months long and is not synchronised between individuals. This unusual reproductive phenology is useful to demonstrate the explicitly circular basis of Fourier analysis, and how analysis at the individual-level allows for quantification of complex

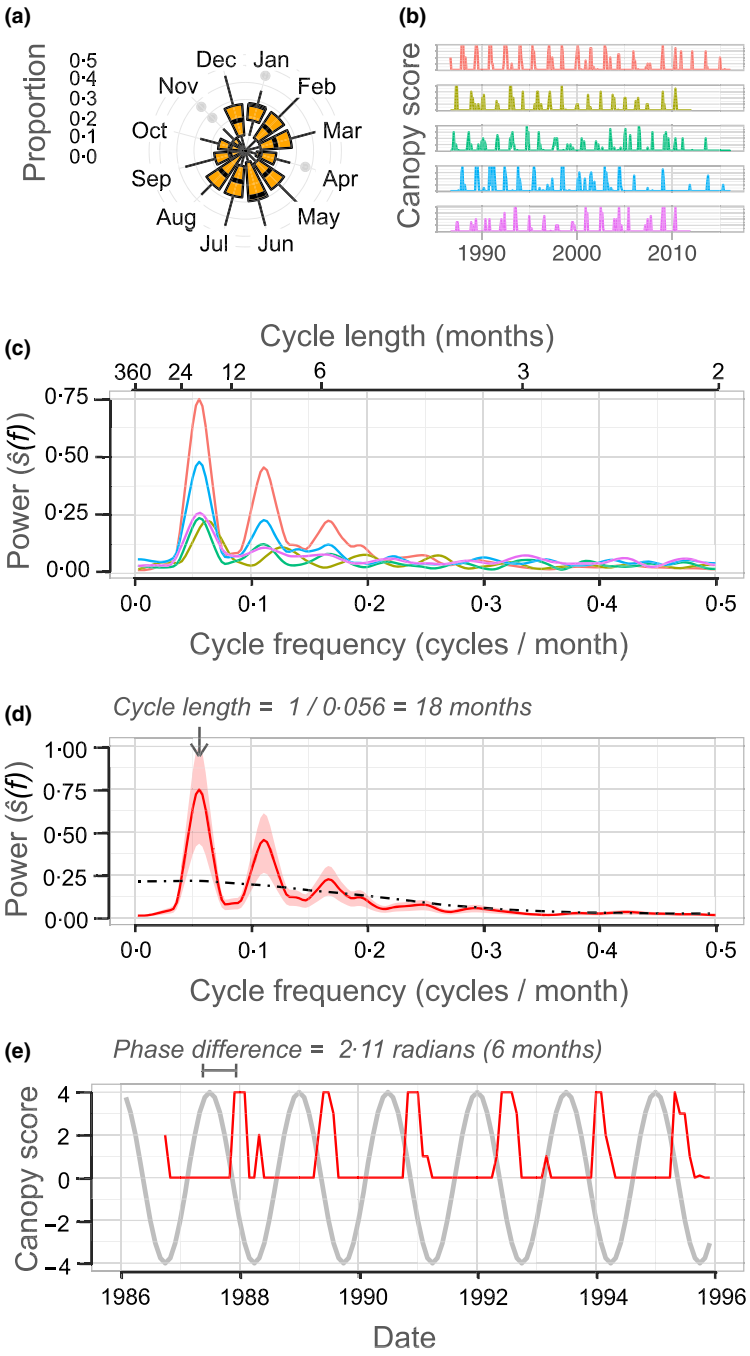


Fig. 1. Using Fourier analysis to detect flowering phenology for a single species *Duboscia macrocarpa*. (a) Boxplots showing the proportion of individuals ($n = 11$) in flower each month from 1986 to 2016. There is no obvious seasonal flowering pattern for this species. (b) ***Time-series*** plots showing flowering canopy scores every month since 1986–2016 (five individuals shown as an example). There appears to be some regular flowering cycles for individuals. (c) ***Periodogram*** displaying the ***smoothed spectral estimates*** (***bandwidth*** = 0·1) derived from ***Fourier analysis*** for each individual flowering timeseries in (b). The *x*-axis shows all possible cycle ***frequencies*** (from one cycle every 2 months to the full length of the series). The *y*-axis shows the ***power*** of each cycle. The highest peak in each ***spectrum*** occurs at a frequency of 0·056 cycles per month (indicating a flowering cycle length of 18 months). (d) Periodogram displaying smoothed spectral estimate derived from Fourier analysis for the first flowering timeseries shown in (b) (red line). The 95% confidence intervals for the spectral estimate (red shades) show that the ***dominant peak*** (grey arrow) at 0·056 cycles per month is different from the null hypothesis of no cyclicity (the null continuum: black dashed line). We can be confident that the 18-month cycle is different from surrounding noise and represents a real flowering cycle. (e) Demonstration of ***co-Fourier*** analysis to derive the relative phase of the flowering cycle identified in (d). The flowering timeseries (red line) is decomposed alongside a regular ***cosine curve***, simulated to have the same ***cycle length*** as the flowering data (18 months) and by convention for our data peaking on the 1 January 1986 (grey line). The ***phase difference*** (2·11 ***radians***) between the two timeseries can be converted to time (6 months).

tropical phenology. R scripts are provided in Appendix S6 and follow this description.

DATA INPUT REQUIREMENTS

For all Fourier analyses, we used the function *spectrum* from the R base package '*stats*' (R Core Team 2015). The method requires regular time intervals between observations, so we interpolated data for gaps up to three data points long using a simple linear estimator, *interpNA* from R package '*timeSeries*' (Rmetrics Core Team *et al.* 2015). For longer gaps we suggest analysing timeseries in separate parts but more sophisticated forms of interpolation could be used or Lamb normalised periodogram analysis (Press *et al.* 1992) which allows for unevenly spaced data.

THE PERIODOGRAM

The Fourier transform decomposes a timeseries into a series of ***sine and cosine waves*** of differing ***frequencies***, quantifying the power of each via the ***spectral estimate***, visualised in the ***periodogram*** (Fig. 1c). The shortest possible cycle for our data is 2 months long (twice the observation interval) and the longest is the full length of the data available. Cycles not well supported by the data have low power, whereas cycles well supported by the data have high power.

SMOOTHING THE SPECTRAL ESTIMATE

The ***raw*** (unsmoothed) ***spectral estimate*** shows all fine-scale structure and can be overly influenced by certain segments of data. We smooth all spectral estimates using a moving-average smoother – the modified ***Daniell kernel*** – available within function *spectrum*. The width of the Daniell kernel (known as the ***span***) is user-specified and is a compromise between ***resolution*** and ***stability***. The classic text on this method (Bloomfield 2000) recommends a trial and error approach for span-choice relying on visual observation of the periodogram. After much experimentation we found that successively applying the Daniell kernel to achieve a smoothed spectral estimate with a ***bandwidth*** close to 0·1 gave sufficient resolution to identify ***dominant peaks*** in the periodogram. For example applying a Daniell kernel with a span of seven, followed by a kernel with a span of nine to the first *D. macrocarpa* flowering timeseries of length 353 months (Fig. 1b) resulted in a spectral estimate with bandwidth 0·099. Spans to achieve this resolution vary depending on initial time-series length; we provide appropriate spans for data ranging from 24 to 360 months in Appendix S6 (line 160). ***Smoothed spectral estimates*** derived from Fourier analysis of flowering data for five example *D. macrocarpa* individuals are shown in Fig. 1c.

IDENTIFYING DOMINANT CYCLES

Interpreting the periodogram begins with observing the general shape of the ***spectrum*** (e.g. is the data influenced by short or long cycles) and then to identify the peaks with highest power, representing ***dominant cycles*** within the data. The smoothed spectral estimates derived from flowering data for *D. macrocarpa* show a similar pattern between individuals (Fig. 1c). The highest peak for each individual is near to 0·056 cycles per month (equivalent to a cycle length of 18 months).

ASSIGNING CONFIDENCE TO DOMINANT CYCLES

Tree phenology studies often rely on monthly canopy observations and are subject to both measurement error (observation uncertainty) and natural variation (process uncertainty). Because of these uncertainties a measure of confidence is needed to differentiate real cycles from the surrounding noise. Bloomfield (2000) suggests that spectral estimates approximate a chi-square distribution, and that 95% confidence intervals can be derived as follows,

$$\frac{v\hat{s}(f)}{\chi^2_v(0\cdot 975)} \leq s(f) \leq \frac{v\hat{s}(f)}{\chi^2_v(0\cdot 025)} \qquad \text{eqn 1}$$

where v is the degrees of freedom (derived from the function output), $\hat{s}(f)$ is the spectral estimate, $s(f)$ is the true spectrum and $\chi^2_v(0{\cdot}975, 0{\cdot}025)$ are the 2·5% and 97·5% quantiles of the chi-square distribution with v degrees of freedom.

There are two credible null hypotheses – representing 'no cyclicity' – with which to compare the 95% confidence intervals. The first is the ***null continuum*** of the spectrum, which is an extreme smooth of the spectral estimate such that only the underlying shape remains (dotted line, Fig. 1d). The second is simply the mean spectrum (otherwise known as the white noise spectrum; Meko 2015). We prefer the null continuum as its use results in fewer false positive results at medium to high noise scenarios (Appendix S2).

We found we could achieve sufficient smoothness for the null continuum by successively applying the Daniell kernel to give a bandwidth similar to 1 (Appendix S1 line 160). Where the lower confidence interval for a specified frequency does not overlap with the null continuum, the peak at that frequency can objectively be considered as significantly different from the surrounding noise and representing a real cycle. Bloomfield (2000) cautions against general fishing expeditions for significant peaks because the 95% confidence intervals calculated are not simultaneous. We therefore, only recommend using this method to test the dominant peak, not all local peaks. Occasionally we find that when data are highly irregular, the dominant peak is identified at the longest possible cycle length and is likely to score as 'significant' against the null continuum. To avoid these false positive results, we screen Fourier outputs and exclude dominant cycles greater than half the data length.

95% confidence intervals for the smoothed spectral estimate derived from one example *D. macrocarpa* timeseries are shown in Fig. 1d. We can be confident that the dominant peak at 18 months represents a real flowering cycle because the lower confidence interval does not cross the null continuum.

ASSESSING TIMING AND SYNCHRONY

To assess timing and synchrony within populations, we developed a method to reference the peak events of tropical phenological cycles in time using a simulated cosine curve within ***co-Fourier analysis***. Co-Fourier allows simultaneous Fourier analysis of any two timeseries and in addition to normal outputs, gives an estimate for the lag (***phase difference***) between the timeseries for every possible cycle. Once a dominant cycle has been detected in an empirical timeseries, we simulate a cosine curve with matching cycle length, by convention for our data peaking on 1 January 1986. After co-Fourier analysis of the empirical timeseries alongside the matching simulated timeseries, we then extract the phase difference associated with the dominant cycle.

In Fig. 1e we show flowering data for an example *D. macrocarpa* individual alongside a simulated cosine curve with matching cycle length (18 months) and peaking on 1 January 1986. The phase difference between these two timeseries at the dominant cycle of 18 months is 2·11 ***radians***.

Phase difference can be converted to time (an estimate of the first flowering peak, in months since 1 January) by the following,

$$\text{if } \Phi_{\text{radians}} > 0, \quad \Phi_{\text{months}} = \frac{\Phi_{\text{radians}}}{(2\pi/\lambda)}$$

$$\text{if } \Phi_{\text{radians}} < 0, \quad \Phi_{\text{months}} = \frac{\Phi_{\text{radians}} + 2\pi}{(2\pi/\lambda)} \qquad \text{eqn 2}$$

where Φ is the phase difference and λ is ***wavelength*** in months.

It is important to consider that radians are a circular unit and there are 2Π radians in a full cycle no matter how many months are in that cycle. Converting phase to months is very simple when the cycle is annual: 1 month = 2Π/12 and the first peak month will be the only peak month in a given calendar year. However, for cycle lengths other than 12 months, conversion to time will need some careful thought. For a 6-month cycle, we would expect two peaks in each calendar year, and for an 18-month cycle we would expect one peak a calendar year but in different months in alternate years.

For the *D. macrocarpa* timeseries used as an example in Fig. 1e, the phase difference of 2·11 radians converts to 6 months since 1 January, placing the first peak at the beginning of July. The next peak in flowering will occur 18 months later, at the beginning of January. We would expect this individual to have flowers in January and July in alternate years.

CALCULATING MEAN TIMING AND SYNCHRONY FOR SPECIES

Mean phenophase timing can be computed for a sample with the same dominant cycle by taking the ***circular mean*** of the phase difference (in radians) for each individual, as calculated from co-Fourier analysis. Synchrony can be quantified by taking the ***circular standard deviation*** of the mean phase [all circular values calculated using the R package '*circular*' (Agostinelli & Lund 2013)]. For the *D. macrocarpa* example, mean phase difference for all individuals with significant dominant cycle at 18 months is 0·94 ± 1·68 SD radians. Converted to time, this references a flowering peak in mid-March and mid-September in alternate years. However, synchrony between individuals is so low (SD of peak month is 4·8 months) that 'peak flowering' for the population has little biological meaning.

In Appendix S4 we include a detailed description of Fourier analysis for the flowering cycles of two additional species (*Antidesma vogelianum Muell. Arg.* flowering on a 6-month cycle, and *Pentadesma butyracea Sabine* flowering on an annual cycle) and a comparison of Fourier alongside four other commonly used methods for seasonal phenology analysis – graphical representations, circular statistics, autocorrelation analysis and GAMs.

Scaling up – quantifying flowering phenology among many individuals and species

METHODS

We used the methods developed above to quantitatively describe flowering data for all species monitored as part of the Lopé study. We preselected 856 individuals (70 species of 26 families) with the following criteria; greater than 5 years continuous data, at least one flowering event and no persistent records of disease (species list given in Appendix S3). Where we found isolated gaps longer than 3 months, we excluded data before or after (whichever was shorter) from further analysis. Linear interpolation for gaps shorter than 3 months was necessary for 95% of the individuals in the sample. Time series' length ranged from 60 to 353 months (mean = 249 months).

To quantitatively describe regular cycles, we ran Fourier analysis and a confidence test of the dominant flowering cycle for each tree. To allow comparison between individuals for the power of the dominant cycle, we normalised the spectrum so that the mean power across frequencies was equal to one (Polansky *et al.* 2010).

To summarise at the species-level we calculated the modal cycle length for species with more than five individuals with significant dominant cycles. To estimate the level of synchrony at the species-level, we ran co-Fourier analysis for each individual with a significant dominant cycle equal to the modal cycle length for that species (only including species with more than five such individuals). From the co-Fourier outputs we calculated the standard deviation of mean phase difference in radians and converted to months using eqn 2 for each species.

We present whole sample summaries for time-series length and sample size per species and compare these between all individuals and those for which we could detect significant cycles. We then present the most common flowering cycles and level of synchrony (standard deviation of mean phase difference) per species. We also tested the impact of time-series length as a predictor of detecting significant regular phenology using a binomial generalised linear mixed model (GLMM) with species as a random effect.

RESULTS

We detected significant regular flowering cycles for 509 out of 856 individuals in our sample, 79% of which were annual. Of those for which we could not confidently detect regular cycles, 22 came from five species for which no significant cycles were detected (e.g. *Baillonella toxisperma Pierre* and *Dacryodes normandii nornandii Aubr. & Pell.,* Appendix S3: Table 2).

When only trees with significant cycles were included, the sample distribution shifted towards longer timeseries (Fig. 2a), and mean sample size per species for all trees (12 individuals ±8·1 SD) was reduced (seven individuals ±5·8 SD) (Fig. 2b). We found time series' length to be a significant positive predictor (Z value = 6·42, $P < 0·001$) of the likelihood of detecting a significant regular cycle from the data (GLM outputs in Appendix S5).

To assess modal cycle length we used a subsample of 42 species (458 individuals). The modal flowering cycle for most species was annual (37 species, e.g. *P. butyracea*, Appendix S4), with others flowering on a 6-month (4 species, e.g. *A. vogelianum*, Appendix S4) and an 18-month basis (1 species, *D. macrocarpa*, Appendix S4) (Figs 2c and 3; Appendix S3: Table 2).

To assess level of synchrony between species we used a subsample of 39 species (402 individuals). The majority of species had flowering cycles well synchronised between individuals (35 species with standard deviation of mean peak <1 month) (Fig. 2d; Appendix S3: Table 2).

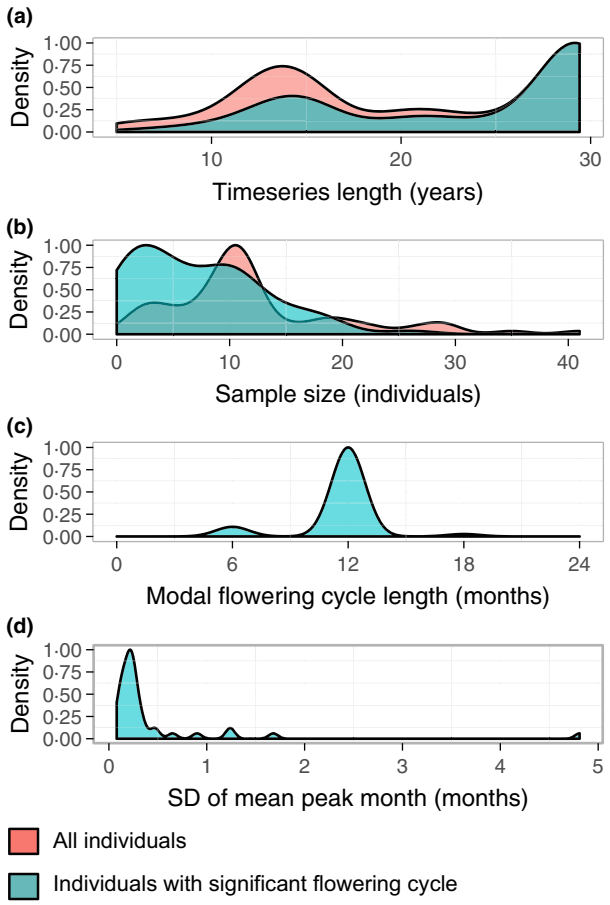


Fig. 2. Summary of flowering phenology for all tree species monitored at Lopé NP, Gabon. (a) Density plot of time series' length for all individuals analysed (red, 856 individuals) compared to individuals with significant flowering cycles (blue, 509 individuals). (b) Density plot of number of individuals per species for all individuals (red, 856 individuals, 70 species) compared to individuals with significant flowering cycles (blue, 509 individuals, 65 species). (c) Density plot of most common flowering ***cycle length*** (mode) per species, for a subsample of 42 species, each more than five individuals with significant flowering cycles (458 individuals). (d) Density plot of ***synchrony*** (standard deviation of mean peak month) per species, for a subsample of 39 species, each with more than five individuals with significant dominant cycle equal to the species modal cycle length (402 individuals).

Species showed considerable inter and intraspecific variation in flowering phenology (Fig. 3). Some species were split between different cycle length strategies; e.g. for a sample of 19 *Uapaca guieensis Muell. Arg.* trees, the dominant flowering cycle was annual for 13 trees and 6 months for six trees. Species also varied in the power of their dominant flowering cycles. Despite all individuals shown in Fig. 3 having significant flowering cycles, some species such as *Maranthes glabra (Oliv.) Prance* (mean power = 9·3 ± 1·6 SD) and *Xylopia aethiopica (Dunal) A. Richard* (mean power = 8·1 ± 2·6 SD) tended to have much stronger, less noisy cycles than others such as *Klainedoxa gabonensis Baill.* (mean power = 2·1 ± 0·4 SD) and *Pseudospondias microcarpa (A Rich.) Engl.* (mean power = 2·4 ± 0·7 SD) (Appendix S3: Table 2).

Testing Fourier under different scenarios using both simulated and field data

METHODS

To test the impact of noise and sample length on cycle detectability, we undertook a power analysis of simulated phenology data. We simulated 10 000 individual timeseries representing an annually repeating flowering cycle peaking in June, with three key parameters allowed to vary between 'individuals'; (i) the regularity of the peak month (representing process uncertainty), (ii) the detectability of flowering events (representing observation uncertainty) and (iii) the length of data recorded. For each year of data, we generated monthly flowering scores of zero and a peak of 3-month duration with positive scores randomly chosen from a distribution similar to that found in our field data. We varied levels of regularity by randomly choosing the peak flowering month each year from a truncated normal distribution (ranging from 2 to 11, with mean six and standard deviation randomly selected from 0·1 to 6). The standard deviation of the distribution was consistent between years but allowed to vary between individuals. We then varied levels of detectability by replacing a certain percentage of randomly chosen positive flowering scores with zeros (from zero to 60%). Finally, a window of data (5, 10 or 15 years) was randomly cut from each full-length timeseries prior to Fourier analysis (simulated data are plotted in Appendix S2 as an example). We assessed the dominant cycle using a 95% confidence test and whether it fell within the expected interval for an annual cycle (11–13 months).

To demonstrate the impact of data length with realistic noise we also conducted a power analysis using all individual timeseries from the Lopé study longer than 20 years, from which we had previously detected significant annual flowering cycles and for species with more than five such individuals (233 individuals of 30 different species). We randomly chose individual timeseries from this sub-sample and cut shorter windows of data (window length randomly selected from the range 2 : 20 years with randomly selected start date), repeating 10 000 times. We analysed the windowed timeseries with Fourier as described above and recorded if the dominant cycle was significant and fell within the expected interval for an annual cycle (11–13 months). We fitted binomial GLMs to compare the effect of time series' length between species.

RESULTS

The power analysis of simulated phenology data (Fig. 4) showed that as time series' length increased, from 5 to 15 years, so did likelihood of confidently detecting the annual cycle. For example for a mid-level noise scenario (cycle regularity 2 SD; zero replacement 20%) the proportion of the sample with a significant annual cycle was zero after 5 years, 57% after 10 years and 81% after 15 years. However, at relatively low-noise scenarios, (highly regular cycles <1 SD; low zero replacement <20%), the effect of time-series length saturated quickly, with 100% likelihood of detecting a significant annual cycle after

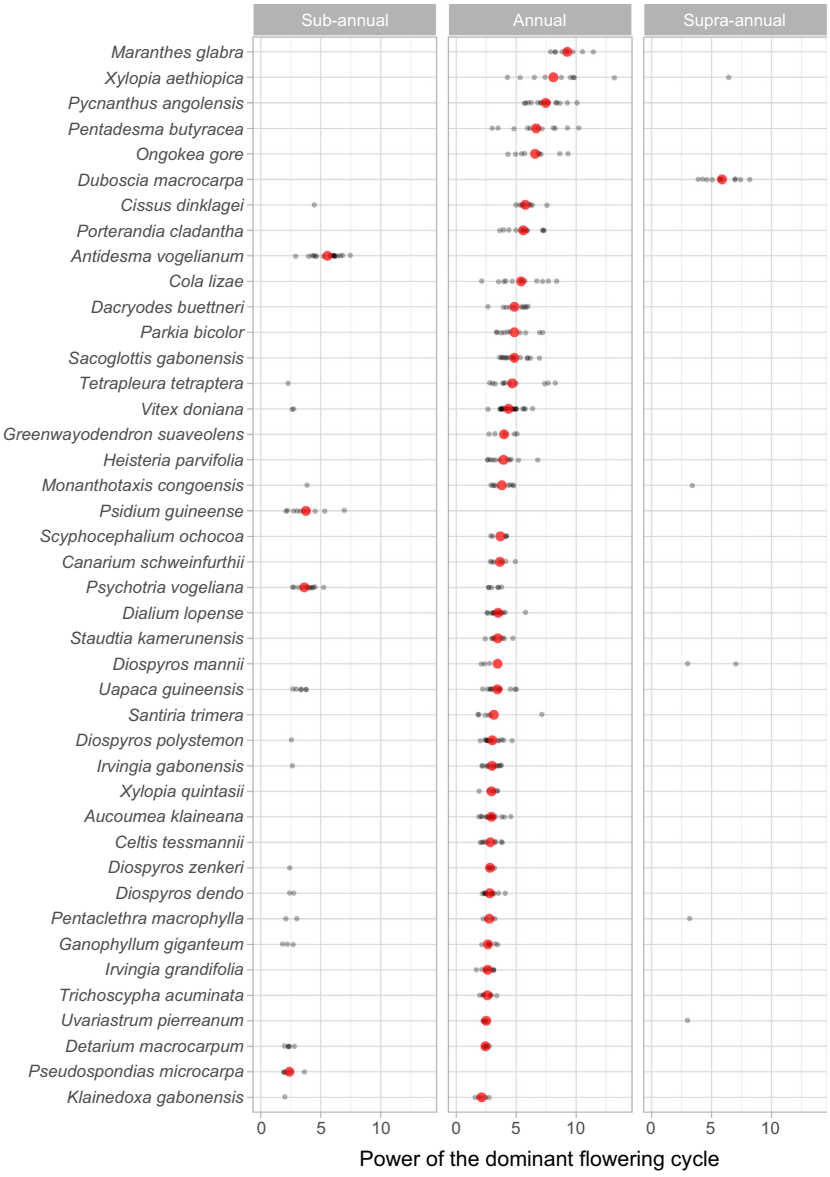


Fig. 3. Inter and intraspecific variation in flowering phenology for tree species monitored at Lopé NP, Gabon. ***Cycle length*** (sub-annual, annual and supra-annual) and ***power*** for each individual (grey dots) and modal cycle length and mean power per species (red dots) from a sub-sample of 42 species with more than five individuals with significant flowering cycles (458 individuals).

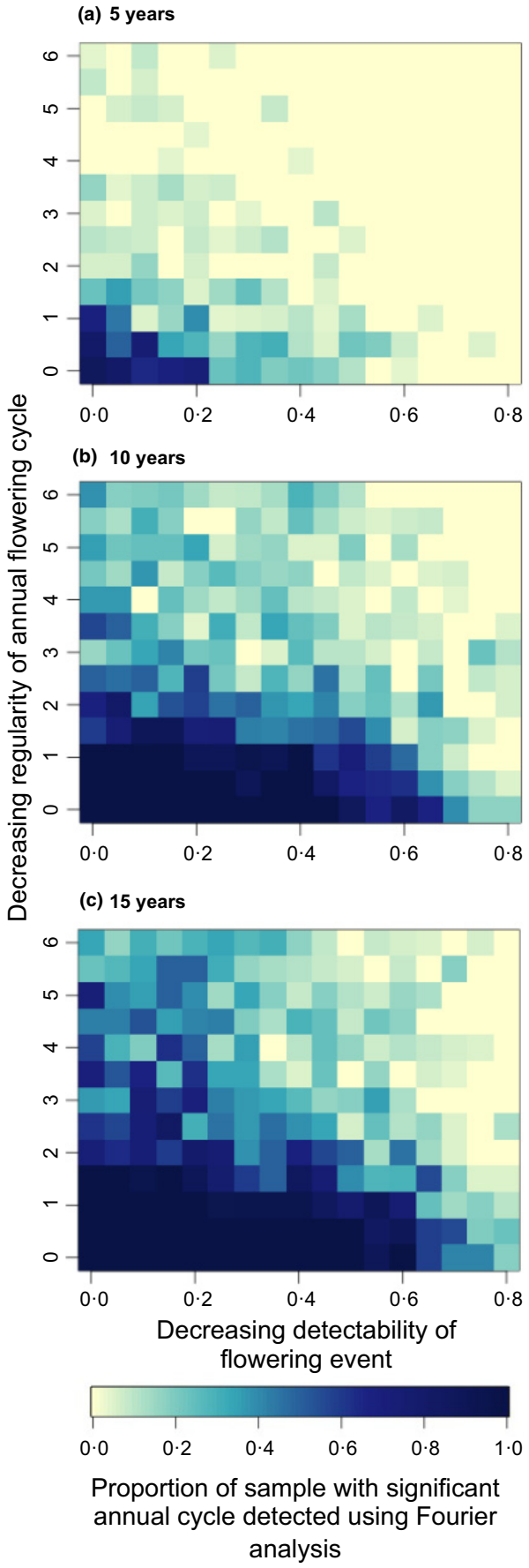
just 5 years. In contrast at high noise scenarios (highly irregular cycles >4 SD; zero replacement >60%), likelihood of detecting a significant annual cycle never rose above 20% even after 15 years. For highly regular cycles (SD < 2), even poor event detectability (zero replacement 40–60%) had little impact on likelihood of detecting the cycle.

Similar to the simulated data, we found that as time series' length increased, so did likelihood of detecting regular cyclic behaviour for our field data (Fig. 5). We found that for the species in our sample with the most positive slope estimates for time-series length [*M. glabra* and *Pycnanthus angolensis (Welw.) Warb.*, Appendix S5], just 6 and 7 years of data, respectively, were required before the annual flowering cycle could be detected with greater than 95% likelihood. However, species ranged widely, with 19 species not reaching this 95% threshold until after 20 years. The species with the least positive slope estimates were *Detarium macrocarpum Harms* and *Greenwaydodendron suaveolens Engl. & Diels*. (Appendix S5).

Discussion

DETECTABILITY AND POWER

The flowering phenology of trees observed at Lopé National Park, Gabon, is dominated by annual cycles (88% species), in contrast with forests from the neotropics that appear to be dominated by sub-annual reproductive cycles and the Dipterocarp forests of South-East Asia that are dominated by supra-annual reproductive cycles (Sakai 2001). We could not confidently describe regular cycles for many individuals in our sample (41%), where either flowering is regular but the data were too noisy or too short for detection or flowering is irregular. Observation length was shown to be a significant positive predictor of detecting regular cycles in both field data and simulations. Even when cycles were confidently described, we found that the power attributed to cycles ranged widely, meaning that the flowering phenology of some species is much



noisier than others. However, the source of this noise is difficult to differentiate for field data. To explore this further we simulated two forms of noise associated with both process and observation uncertainty and found that cycle regularity has a greater effect on ability to detect a significant cycle than event detectability: Fourier analysis can be used to detect the cycle even if the observer misidentifies 60% of flowering months. There are likely to be additional sources of noise in the field, such as false recording of non-existent phenophases, however, we consider these to occur less often.

We attributed cycle characteristics to species when we had five or more individuals with significant cycles, under the biological assumption that phenology is an evolutionarily adaptive trait and likely to be constraining con-specifics in a similar way. However, true levels of intraspecific variation are unknown. We find considerable intraspecific variation for some species (i.e. *Uapaca guineensis*) and further research may reveal that phenology is not necessarily a stable trait within a species or an individual's lifetime.

Our results can be used to inform effective collection, processing and analysis of phenological data. We have shown that where suitable data is available, objective analyses can be used to confidently detect regular phenology and that frequency-based outputs – cycle length, power, timing and level of synchrony – give a suite of indicators that could be used to quantitatively describe and compare phenology globally.

DEVELOPMENT FOR CAUSATION AND CHANGE RESEARCH

The indicators derived from Fourier analysis can be used to address research questions such as the proximate and ultimate causes of adaptive phenology and detection of change. Where data is available, analysis at the individual-level allows for inclusion of covariates (e.g. location, age, size of individuals, etc.) in subsequent statistical models, either in combination with random effects and best linear unbiased predictors to account for variation (e.g. between different sites, genera or functional groups) or as fixed effects to test hypotheses of the causes of variation between individuals' phenology. Co-Fourier analysis would allow testing of other cyclic factors (such as climate data) alongside phenology to measure synchrony. The advantage of these spectral approaches is that they explicitly model the circular nature of phenology and weather data without losing power by clumping data points into arbitrary time periods or pseudo-replication.

Detecting long-term changes in phenology is challenging and field observations (Plumptre 2011) are vital to stimulate hypotheses and further analysis. However, it will be increasingly important to measure the statistical confidence of

Fig. 4. Power analysis of simulated phenology data (n = 10 000) to show the impact of data noise and length (5, 10 and 15 years; (a)–(c)) on likelihood of detecting cycles using Fourier analysis. Noise simulated as cycle regularity (y-axis: standard deviation – 0·1 : 6 – of mean month of annual flowering event) and event detectability (x-axis: proportion – 0 : 60% – of positive flowering events replaced by zeros).

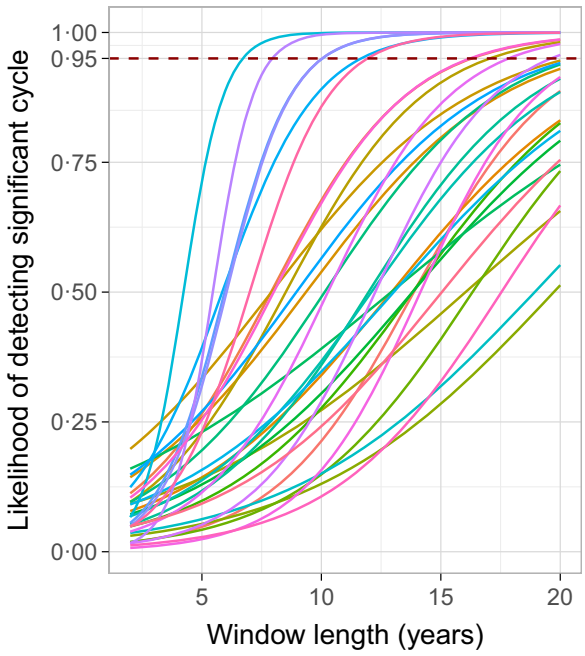


Fig. 5. Power analysis of annually flowering phenology data from Lopé NP to show the impact of time-series length (2–20 years window length) on cycle detection using Fourier analysis (10 000 random samples from 233 individuals of 30 species). Generalised linear model (GLM) predictions (family = binomial, link = logit) for each species (see Appendix S5, for species key and GLM outputs).

detected changes. To date, studies of change in tropical phenology are few (Appendix S1), due to the paucity of long-term data. Wavelet analysis is the natural extension of Fourier into the time-frequency domain (Wittemyer *et al.* 2008; Hudson, Kang & Keatley 2010; Polansky *et al.* 2010), overcoming assumptions of stationarity, to estimate the spectrum as a function of time (Cazelles *et al.* 2008). For phenology research, this could enable analysis of whether individuals or species reproduce more or less frequently (e.g. change in dominant cycle length), reproduce at the same frequency but with more or less certainty (e.g. change in the power of the dominant cycle) or shift phase and become more or less synchronised over time. The power of a cycle may be a more subtle and effective indicator for change than frequency to track increasing uncertainty over time, especially in the shorter term.

In a formal comparison of this Fourier-based method with other commonly used methods for quantifying phenology (Appendix S4), we found Fourier is flexible to diverse phenology and provides a suite of quantitative information to describe seasonal activity with attribution of variance and confidence.

STEPS FORWARD

We have shown that at least 6 years of data are necessary to confidently detect reproductive cycles amongst our species sample. For data collection scenarios resulting in noisier data – those with high likelihood of measurement error (e.g. inconspicuous flowers), systematic error (e.g. high inter-observer uncertainty) or natural variation that cannot be controlled for (e.g. diverse array of phenological responses within a population) – it will be necessary to invest in large samples of individuals over a longer time period to detect cycles confidently. To effectively monitor the response of tropical forests to global change, it will be necessary to focus efforts on suitable indicator species – those with good signal to noise ratios – to maximise analytical power over relatively short time periods.

For many phenology research questions, collecting sufficient data will be a challenge and require significant research effort. Ways to achieve this include: formation of research networks and greater coordination of methods and objectives between sites, Internet-based citizen-science data collection networks and technical solutions to data collection, such as automated canopy photography and GIS.

Conclusions

Phenology is a key adaptive trait shown to determine species distributions (Chuine 2010) and as such will shape how ecosystems respond to rapidly increasing regional and global changes including human pressure. With the emergence of long-term tropical phenology data, the need also emerges for appropriate analytical methods to improve our understanding of the functioning of ecosystems. We present a Fourier-based method that can be further developed and tested, to give simple, flexible and quantifiable indicators for phenology activity, and demonstrate the importance of consistent long-term investment in phenological research.

Authors' contributions

E.B., N.B., K.A. and A.J. conceived the ideas for this manuscript; C.T., L.W. and K.A. designed the field methodology; E.D., J.-T.D., C.T., K.A., L.W. and K.J. collected the data; E.B., N.B. and K.A. analysed the data; E.B., N.B., K.A. and A.J. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

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 is currently supported by an Impact Studentship funded by the University of Stirling and ANPN. 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 Daisy Dent, Tim Paine, Ed Mitchard and two reviewers of a previous version of this manuscript whose comments in preparation significantly improved it.

Data accessibility

R scripts: uploaded as online supporting information (Appendix S6). Individual- and species-level flowering data: University of Stirling's DataSTORRE (http://hdl.handle.net/11667/83).

References

Agostinelli, C. & Lund, U. (2013) R Package 'Circular': Circular Statistics (version 0.4-7). Available at: https://r-forge.r-project.org/projects/circular/ (accessed 13 October 2016).

Anderson, D.P., Nordhelm, E.V., Moermond, T., Bi, Z.B.G. & Boesch, C. (2005) Factors influencing tree phenology in Taï National Park, Côte d'Ivoire. *Biotropica*, **37**, 631–640.

Bloomfield, P. (2000) *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons, New York, NY, USA.

Butt, N., Seabrook, L., Maron, M., Law, B., Dawson, T.S. & Syktus, J. (2015) Cascading effects of climate extremes on vertebrate fauna through changes to low-latitude tree flowering and fruiting phenology. *Global Change Biology*, **21**, 3267–3277.

Cannon, C.H., Curran, L.M., Marshall, A.J. & Leighton, M. (2007) Long-term reproductive behaviour of woody plants across seven Bornean forest types in the Gunung Palung National Park (Indonesia): suprannual synchrony, temporal productivity and fruiting diversity. *Ecology Letters*, **10**, 956–969.

Cazelles, B., Chavez, M., Berteaux, D., Ménard, F., Vik, J.O., Jenouvrier, S. & Stenseth, N.C. (2008) Wavelet analysis of ecological time series. *Oecologia*, **156**, 287–304.

Chambers, L.E., Altwegg, R., Barbraud, C. *et al.* (2013) Phenological changes in the southern hemisphere. *PLoS ONE*, **8**, e75514.

Chapman, C., Wrangham, R.W., Chapman, L.J., Kennard, D.K. & Zanne, A.E. (1999) Fruit and flower phenology at two sites in Kibale National Park, Uganda. *Journal of Tropical Ecology*, **15**, 189–211.

Chuine, I. (2010) Why does phenology drive species distribution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 3149–3160.

Cleland, E.E., Chuine, I., Menzel, A., Mooney, H.A. & Schwartz, M.D. (2007) Shifting plant phenology in response to global change. *Trends in Ecology and Evolution*, **22**, 357–365.

Hudson, I.L., Kang, I. & Keatley, M.R. (2010) Wavelet analysis of flowering and climatic niche identification. *Phenological Research* (eds I.L. Hudson & M.R. Keatley), pp. 361–391. Springer Science and Business Media, Dordrecht, the Netherlands.

Hudson, I.L. & Keatley, M. (2010) *Phenological Research* (eds I.L. Hudson & M.R. Keatley). Springer Science and Business Media, Dordrecht, the Netherlands.

IPCC (2014) *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds, Core Writing Team, R.K. Pachauri & L.A. Meyer). IPCC, Cambridge University Press, Cambridge, UK and New York, NY, USA.

Meko, D. (2015) *Applied Time Series Analysis*. Online notes for course (Geosciences 585A) offered at the University of Arizona. Available at: http://www.ltrr.arizona.edu/~dmeko/geos585a.html (accessed 26 August 2015).

Newbery, D.M., Chuyong, G.B. & Zimmermann, L. (2013) Mast fruiting of large ectomycorrhizal African rain forest trees: importance of dry season intensity, and the resource-limitation hypothesis. *New Phytologist*, **170**, 561–579.

Newstrom, L., Frankie, G. & Baker, H. (1994) A new classification for plant phenology based on flowering patterns in lowland tropical rain forest trees at La Selva, Costa Rica. *Biotropica*, **26**, 141–159.

Norden, N., Chave, J., Belbenoit, P., Caubère, A., Châtelet, P., Forget, P.M. & Thébaud, C. (2007) Mast fruiting is a frequent strategy in woody species of eastern South America. *PLoS ONE*, **2**, e1079.

Obeso, J.R. (2002) The costs of reproduction in plants. *New Phytologist*, **155**, 321–348.

Parmesan, C. (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology Evolution and Systematics*, **37**, 637–669.

Pau, S., Wolkovich, E.M., Cook, B.I., Nytch, C.J., Regetz, J., Zimmerman, J. & Wright, S.J. (2013) Clouds and temperature drive dynamic changes in tropical flower production. *Nature Climate Change*, **3**, 838.

Plumptre, A.J. (2011) *The Ecological Impact of Long-Term Changes in Africa's Rift Valley*. Nova Science, Hauppauge, NY, USA.

Polansky, L. & Robbins, M.M. (2013) Generalized additive mixed models for disentangling long-term trends, local anomalies, and seasonality in fruit tree phenology. *Ecology and Evolution*, **3**, 3141–3151.

Polansky, L., Wittemyer, G., Cross, P.C., Tambling, C.J. & Wayne, M. (2010) From moonlight to movement and synchronized randomness: *Fourier and wavelet analyses of animal location time series data*. *Ecology*, **91**, 1506–1518.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (1992) *Numerical Recipes in C: The Art of Scientific Computing Second*. Cambridge University Press, Cambridge, UK.

R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/ (accessed 13 October 2016).

Rmetrics Core Team, Wuertz, D., Setz, T. & Chalabi, Y. (2015) timeSeries: Rmetrics – Financial Time Series Objects. Available at: https://CRAN.R-project.org/package=timeSeries (accessed 13 October 2016).

Sakai, S. (2001) Phenological diversity in tropical forests. *Population Ecology*, **43**, 77–86.

van Schaik, C.P., Terborgh, J.W. & Wright, J.S. (1993) The phenology of tropical forests: adaptive significance and consequences for primary consumers. *Annual Review of Ecology and Systematics*, **24**, 353–377.

Ting, S., Hartley, S. & Burns, K.C. (2008) Global patterns in fruiting seasons. *Global Ecology and Biogeography*, **17**, 648–657.

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

van Vliet, A.J.H. (2010) Societal adaptation options to changes in phenology. *Phenological Research* (eds I.L. Hudson & M.R. Keatley), pp. 75–98. Springer, Dordrecht, the Netherlands.

Wittemyer, G., Polansky, L., Douglas-hamilton, I. & Getz, W.M. (2008) Disentangling the effects of forage, social rank, and risk on movement autocorrelation of elephants using Fourier and wavelet analyses. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 19108–19113.

Wright, S.J. & Calderon, O. (1995) Phylogenetic patterns among tropical flowering phenologies. *Journal of Ecology*, **83**, 937–948.

Wright, J.S. & Calderon, O. (2006) Seasonal, El Nino and longer term changes in flower and seed production in a moist tropical forest. *Ecology letters*, **9**, 35–44.

Wright, S.J., Carrasco, C., Calderón, O. & Paton, S. (1999) The El Nino Southern Oscillation, variable fruit production, and famine in a tropical forest. *Ecology*, **80**, 1632–1647.

Zalamea, P., Munoz, F., Stevenson, P.R., Paine, C.E.T., Sarmiento, C., Sabatier, D. & Heuret, P. (2011) Continental-scale patterns of Cecropia reproductive phenology: evidence from herbarium specimens. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 2437–2445.

Zimmerman, J.K., Wright, S.J., Calderón, O., Aponte Pagan, M. & Paton, S. (2007) Flowering and fruiting phenologies of seasonal and aseasonal neotropical forests: the role of annual changes in irradiance. *Journal of Tropical Ecology*, **23**, 231–251.

Received 26 August 2016; accepted 30 October 2016
Handling Editor: Robert Freckleton

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Review of methods from the literature. Review of key literature analysing long-term tropical plant phenology data, detailing the phenophase of interest, site, data length, analytical methods used and the scale of data collection and analysis.

Appendix S2. Null hypothesis choice and example simulated data. Power analysis of simulated data to show the impact of null hypothesis choice (null continuum vs. white noise spectrum) for detecting periodicity.

Appendix S3. Species list from Lopé long-term phenology study. List of families (n = 26), species (n = 70) and individuals (n = 856) observed as part of the Lopé long-term phenology study included in Fourier analysis and summarised Fourier outputs at the species level.

Appendix S4. Demonstration of Fourier analysis and comparison with other methods. Demonstration of Fourier analysis for three case study species – *Antidesma vogelianum*, *Pentadesma butyracea*, *Duboscia macrocarpa* – and comparison with other common methods for quantifying flowering phenology.

Appendix S5. GLM outputs. GLM outputs for effect of time-series length on likelihood of detecting significant cycle from all available field data and from power analysis of annually cycling species.

Appendix S6. R code for Fourier analysis of phenology.