

METHODOLOGY ARTICLE

Open Access



Uncovering periodic patterns of space use in animal tracking data with periodograms, including a new algorithm for the Lomb-Scargle periodogram and improved randomization tests

Guillaume Péron^{1*}, Chris H. Fleming^{1,2†}, Rogerio C. de Paula³ and Justin M. Calabrese^{1,2}

Abstract

Background: Periodicity in activity level (rest/activity cycles) is ubiquitous in nature, but whether and how these periodicities translate into periodic patterns of space use by animals is much less documented. Here we introduce an analytical protocol based on the Lomb-Scargle periodogram (LSP) to facilitate exploration of animal tracking datasets for periodic patterns. The LSP accommodates missing observations and variation in the sampling intervals of the location time series.

Results: We describe a new, fast algorithm to compute the LSP. The gain in speed compared to other R implementations of the LSP makes it tractable to analyze long datasets ($>10^6$ records). We also give a detailed primer on periodicity analysis, focusing on the specificities of movement data. In particular, we warn against the risk of flawed inference when the sampling schedule creates artefactual periodicities and we introduce a new statistical test of periodicity that accommodates temporally autocorrelated background noise. Applying our LSP-based analytical protocol to tracking data from three species revealed that an ungulate exhibited periodicity in its movement speed but not in its locations, that a central place-foraging seabird tracked moon phase, and that the movements of a range-resident canid included a daily patrolling component that was initially masked by the stochasticity of the movements.

Conclusion: The new, fast algorithm tailored for movement data analysis and now available in the R-package `ctmm` makes the LSP a convenient exploratory tool to detect periodic patterns in animal movement data.

Keywords: Periodicity, Circadian, Central place foraging, Autocorrelation, Activity cycles, Behavior

Background

Periodic patterns of space use, such as daily routines [1] and annual migrations [2], are a particular form of temporal autocorrelation in animal tracking data that has been largely ignored so far by movement modeling efforts. Periodic behaviors nevertheless constitute manifestations of complex, finely-tuned suites of behaviors that have been shaped by natural patterns of spatiotemporal variation in

resources and risk [3, 4], so that their occurrence, or their absence, is often central to the ecology of species and individuals [5]. In particular, animals' responses to lunar cycles [6] and resource depletion/recovery cycles [7] remain poorly described in most cases, and much insight could potentially be gained from analyzing animal movements with respect to those cycles. Memory effects, cognitive abilities [8, 9], and social interactions [10] represent other intriguing pathways through which periodicities could emerge in animal movement. In addition, periodic patterns of space use represent a major violation of the assumptions of commonly used stochastic movement

* Correspondence: peron_guillaume@yahoo.fr

†Equal contributors

¹Smithsonian Conservation Biology Institute, National Zoological Park, Front Royal, VA 22630, USA

Full list of author information is available at the end of the article

models [11, 12]. Consequently, neglecting periodic behaviors when they occur can bias analyses that condition on fitted movement models, including home range size estimation and resource utilization.

The phrase “periodic pattern of space use” is, however, not to be confused with “periodicity in activity levels”. Periodic patterns of space use, as defined here, correspond to the existence of a characteristic, expected time that separates repeated visits to any given area of the home range of the individuals. This is sometimes also called “movement path recursion” [7, 13], albeit the latter phrase does not explicitly convey the notion of a characteristic revisitation time. In contrast, periodicity in activity levels refers to cycles in a derived quantity such as movement speed, turning angle, or in a separately acquired time series such as acceleration, physiological state, or behavioral state. Periodicity in activity levels is expected in most species (rest/activity cycles) and has been the focus of many recent studies [4, 5, 14–19]. Periodic patterns of space use, by contrast, have received little attention so far, possibly due to a lack of an appropriate statistical framework.

In this study, we describe an analytical protocol to uncover periodic patterns of space use in animal tracking data. Animal movement paths (or trajectories, in the vernacular sense of the suite of positions occupied by an individual through continuous time) are the result of multiple co-occurring processes, some stochastic and some deterministic. Periodic patterns in animal tracking data can thereby be obscured by aperiodic noise, i.e., the non-periodic stochastic component of the animal path (not to be confounded with observation noise or telemetry error). The presence of this noise makes it necessary to employ signal processing methods. The proposed protocol in this study is based on periodograms, and in particular the Lomb-Scargle periodogram (LSP) [20]. This periodogram is particularly relevant for movement data because such data often feature missing observations or variation in the duration of sampling intervals (see Methods), a situation in which more standard and widely used periodograms do not apply [20, 21]. First, we introduce and fully describe a new, fast algorithm to estimate the LSP. This new algorithm makes it tractable to analyze long time series ($>10^6$ records), for which alternative software would need hours or days of computing time. Second we give a detailed primer about periodicity analysis with periodograms, focusing on the specificities of movement data relative to other uses of periodograms (in astronomy, climate science, physiology, and genetics [22–25]). In particular, we lay out a framework to detect artefactual periodicities created by patterns in the way missing data occur or sampling intervals vary. This issue is commonplace in movement ecology (e.g., if the devices are duty cycled to alternate

between fine and coarse sampling rates), but is much less prevalent in other fields and has thus been largely ignored so far. We also emphasize that the background noise of movement data is typically colored, i.e., features temporal autocorrelation in the location, the velocity, or both [26]. This renders existing statistical tests of periodicity (e.g., [27]) inappropriate and requires innovative approaches. Lastly, we highlight the biological insights that can be gained from analyzing periodicity in an animal’s locations instead or in addition to its activity level. We illustrate the approach with three examples from African buffalo (*Syncerus caffer*), waved albatrosses (*Phoebastria irrorata*), and maned wolves (*Chrysocyon brachyurus*).

Methods

Periodograms come from the field of signal processing. The approach works by decomposing the signal (here, animal locations) into a sum of sinusoids of fixed frequencies¹. Periodograms facilitate visual identification of the frequency or frequencies that contribute most to the signal [21], and are the basis for nonparametric statistical tests of periodicity.

The discrete Fourier transform

The basic issue at hand is to identify periodicity in a signal $X = \{X_j; j = 1, \dots, N\}$ where X_j is an animal’s recorded location at time t_j . In animal movement applications, X_j is typically two-dimensional (latitude, longitude) but might also have an altitude component. Assuming stationarity, i.e., that the animal does not change its periodic behavior during the course of the study, the discrete Fourier transform of X , denoted $DFT\{X\}$, yields an estimate of the contribution of any given frequency f to the signal [21]:

$$DFT\{X\}(f) = \sum_{j=1}^N X_j \cdot e^{-ijf} \quad (1)$$

with i the imaginary unit such that $i = \sqrt{-1}$. Importantly, the DFT requires that the interval Δt between subsequent samples is constant, i.e., $t_j = t_1 + (j - 1)\Delta t$.

The DFT-periodogram of a multidimensional signal is $P_{DFT}(f) = 1/K \cdot \sum_k (1/N \cdot |DFT\{X\}(f)|_k|^2)$ where $\cdot|_k$ denotes the k^{th} dimension ($k = 1 \dots K$). The DFT-periodogram reaches a local maximum if frequency f is in phase with the frequency of X . The smallest detectible period in the signal depends on the sampling interval Δt . Frequency $F = 1/2\Delta t$ is known as the Nyquist frequency and is the largest detectible frequency in a signal sampled at interval Δt [21]. There is also a direct relationship between the sample size N and the default frequency resolution of the periodogram: $\Delta f = 1/(N \cdot \Delta t)$. It is possible to refine the frequency resolution (increase the

number of points in the periodogram) up to $\Delta f = F/N$, after which there is no new information in the data to inform new points in the periodogram, leading to autocorrelation in subsequent error terms and an overly smooth periodogram (section A.2 in Additional file 1). The DFT is now a standard tool for signal processing especially since it can be efficiently computed via the fast Fourier transform (FFT) [28], which is available in most statistical software.

The Lomb-Scargle periodogram: an extension of the DFT important for movement ecology

The condition that the sampling interval is constant, which is necessary to apply the DFT, is rarely met in animal tracking datasets. Tracking datasets are often scheduled to be collected on an even time grid but feature many missing records, or the sampling interval varies in duration due to delays in GPS signal acquisition, or the tracking devices may be purposely “duty cycled” to alternate between coarse and fine sampling rates. This renders the DFT inapplicable without first manipulating the data, which usually involves either thinning them until the sampling intervals are constant, or interpolating missing observations. The Lomb-Scargle periodogram (LSP) [20] extends the DFT to situations with missing data or variable sampling intervals, without ad hoc data manipulations. For this reason, the method has been progressively adopted in many fields where identifying periodicities in imperfect datasets is of interest, including astronomy, climate science, and biology [15–18, 22–24]. This article introduces the LSP as an effective non-parametric method to explore an animal tracking dataset for periodic patterns of space use, something that to our knowledge has not been done previously. Movement ecologists can also use the LSP to look for periodicity in activity level [15–18] but the present study does not emphasize this aspect (see however the section “African buffaloes” below).

A new algorithm to compute the Lomb-Scargle periodogram

First, we introduce and describe a new algorithm to compute the LSP. The motivation for this new algorithm is to be able to analyze long time series (>10⁶ records like [29]), or large number of moderately long times series (>50 individuals with >10⁴ records per individual like [30]), while maintaining computation times short enough that the protocol can be used at the data exploration stage. Importantly, the choice of an algorithm (our new approach or any of the pre-existing ones, see below) does not influence the inference; all LSP algorithms compute the same quantity, but their computation times vary. The full derivation

of the new algorithm is described in Additional file 1 and briefly outlined below. First, we define the sampling schedule function w on a regular time grid $\{t_j; j = 1 \dots N\}$:

$$w(t_j) = \begin{cases} 1 & \text{if } X(t_j) \text{ is recorded} \\ 0 & \text{if } X(t_j) \text{ is missing} \end{cases} \quad \text{with } t_{j+1} - t_j = \Delta t \quad (\forall j) \quad (2)$$

If the interval between two records was intended to be constant but some records are missing, $\{t_j\}$ simply codes for the intended sampling schedule. In the case of a duty cycle with varying sampling rate, $\{t_j\}$ codes for the finest sampling resolution. If the sampling interval was intended to be constant but eventually varied around that intended value (e.g., due to delays in GPS signal acquisition), our R package `ctmm` [31, 32] defaults to the median sampling interval for constructing w (see further details in section A.7 in Additional file 1). The user can however optionally refine the temporal resolution of $\{t_j\}$ through the `res.time` option. In this case, we use the Lagrange interpolation of the sinusoids [25] (see more detail below), causing an increase in computing time. We find `res.time=2` to be generally sufficient for most datasets affected by this issue of random variation in sampling interval, while `res.time=4` is almost exact within numerical precision.

Then we exploit the equivalence of the periodogram at frequency f to a least-square fit of the data to sinusoids of frequency f (Scargle 1982) [20].

$$X(t_j) \approx A(f)e^{2\pi if t_j} + A(f)^* e^{-2\pi if t_j} \quad (3)$$

where the amplitude $A(f)$ is the parameter of interest and the $*$ symbol denotes the complex conjugate. The squared error of this fit is, for each dimension k :

$$L^{(k)}(f) = \sum_j w(t_j) \cdot |X(t_j)|_k - (A(f)|_k e^{2\pi if t_j} + A(f)|_k^* e^{-2\pi if t_j})|^2 \quad (4)$$

To estimate the amplitude, we minimize the cost function $L(f) = \sum_k L^{(k)}(f)$. It is possible to perform this minimization analytically (Additional file 1; Eqs. A.9–A.11). Following Scargle (1982) [20], the LSP of the k^{th} dimension of X is then expressed as a function of the amplitude:

$$LSP^{(k)}(f) = \frac{1}{2} \left(\sum_j w(t_j) |X(t_j)|_k^2 - \min_{A(f)} L^{(k)}(f) \right) \quad (5)$$

yielding, after manipulations that are detailed in Additional file 1: Eqs. A13–A15, the following, new formula for the empirical Lomb-Scargle periodogram of a multidimensional time series:

$$\widehat{LSP}^{(k)}(f) = \frac{DFT\{w\}(0) \cdot |DFT\{wX\}(f)|_k^2 - \text{Re}(DFT\{w\}(2f) \cdot DFT\{wX\}(f)|_k^2)}{FT\{w\}(0)^2 - |DFT\{w\}(2f)|_k^2} \quad \widehat{LSP}(f) = \frac{1}{K} \sum_{k=1}^K \widehat{LSP}^{(k)}(f) \quad (6)$$

where $\text{Re}(\cdot)$ denotes the real part of a complex number and $\cdot|_k$ denotes the k^{th} dimension of a variable. In Eq. 6, we can see that if there are no missing data, the LSP reduces to the DFT-periodogram of X because then $DFT\{w\}(0) = N$ and $DFT\{w\}(f) = 0$ ($\forall f > 0$).

What makes this new formula for the LSP fast is that, contrary to the original implementation by Scargle (1982) [20], it is entirely based on the DFT, and can therefore be computed using the fast Fourier transform (FFT) [28]². Exploiting the proven speed of a known algorithm like the FFT is a standard method for developing more efficient computing techniques (e.g., [33]). For the case that interests us here, we assessed the gain in computing time using both algorithmic complexity considerations, and by directly measuring computing time. The computational complexity of an algorithm is conveyed with notation O , meaning “proportional to”. The complexity of the FFT algorithm is known to be $O(N \log N)$ [28], that is, the computing time of the FFT is proportional to $N \log N$, where N is the number of records in the time series. Being entirely based on the FFT, our new algorithm for the LSP is therefore also $O(N \log N)$. By contrast, Scargle’s original computation of the LSP requires running twice through the time series, yielding a $O(N^2)$ complexity, that is a computing speed proportional to N^2 . The difference between $O(N \log N)$ and $O(N^2)$ can be enormous in practice: our algorithm with $O(N \log N)$ complexity runs within seconds in situations where algorithms with $O(N^2)$ complexity need more than a day (Fig. 1). Previously, only Press & Rybicki managed to derive an $O(N \log N)$ routine for the LSP, also by using the FFT [25]. However, the Press & Rybicki algorithm is much more complex than ours; as a consequence, all R packages and freeware for biologists that have a LSP functionality, and of which we are aware [22, 34–37], are based on Scargle’s original algorithm, not on Press & Rybicki’s fast algorithm. Furthermore, although the Press & Rybicki algorithm is based on the FFT like our new algorithm, it requires additional calculations including finer gridding, Lagrange interpolation, and $O(N)$ square-root and trigonometric function evaluations [25], that our approach does not require as long as the time grid $\{t_j\}$ is even or can be approximated as such (see beginning of section for definition of $\{t_j\}$). These additional steps also render the Press & Rybicki algorithm only an approximation to the LSP, whereas our approach is exact as long as the time grid $\{t_j\}$ is even. When the interval durations are variable, i.e., $\{t_j\}$ is not even, both methods

are approximate. In the `ctmm` implementation of our fast algorithm, users can further control the amount of error through the frequency resolution option (`res.freq`).

In summary, because it is based on the FFT and is therefore of $O(N \log N)$ complexity, our new algorithm substantially increases the speed of the computation of the LSP compared to other R implementations, which are based on Scargle’s original $O(N^2)$ algorithm (Fig. 1). This makes LSP computation tractable even for long time series (like [29]), or for numerous time series (like [30]). Importantly, whatever the algorithm used (ours, Press & Rybicki’s, or Scargle’s), the biological inference and statistical power are not affected: the same quantity is estimated.

Lastly, our implementation in the R package `ctmm` [31, 32] also accommodates two important features that are commonplace in movement data but infrequent in other fields. The first of these features is multi-dimensionality; all other LSP implementations in R are 1-dimensional and therefore require post-hoc manipulations to combine results from latitude and longitude time series. In the `ctmm` implementation, periodicity in the latitude, longitude, and altitude can be investigated jointly or separately. The second feature is multiple individuals. Movement ecology datasets often comprise multiple individuals which are expected to behave in similar ways, i.e., exhibit the same periodic patterns. In such a situation, `ctmm` allows fitting multi-individual periodograms, in order to augment sample size and better separate the periodic patterns from the background noise. The principle is described in section A.3 of Additional file 1 and an example presented in the “Maned wolves” section below.

Interpreting periodograms

First, the overall shape of the periodogram of an animal’s track record is influenced by the temporal autocorrelation structure of the aperiodic background noise. This issue was previously largely ignored because in astronomy, which is where the LSP was developed and where most extensions of the LSP have been published, the background noise can be assumed to be independently and identically distributed (“white”) [27]. However, in movement ecology, the stochastic aperiodic component of the animals’ paths is expected to exhibit temporal autocorrelation in the location process, the velocity process, or both (“colored”) [26]. As a consequence, previously developed statistical tests of periodicity [22, 27] cannot be applied to

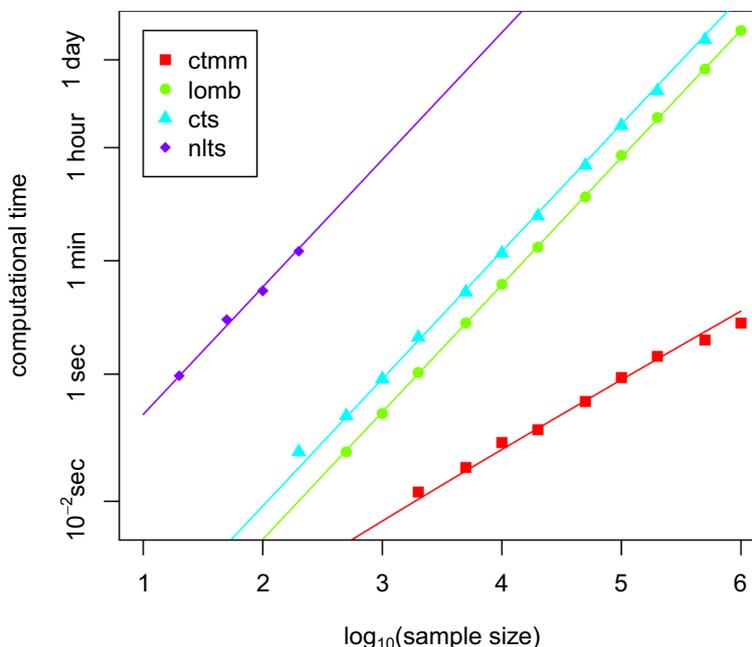


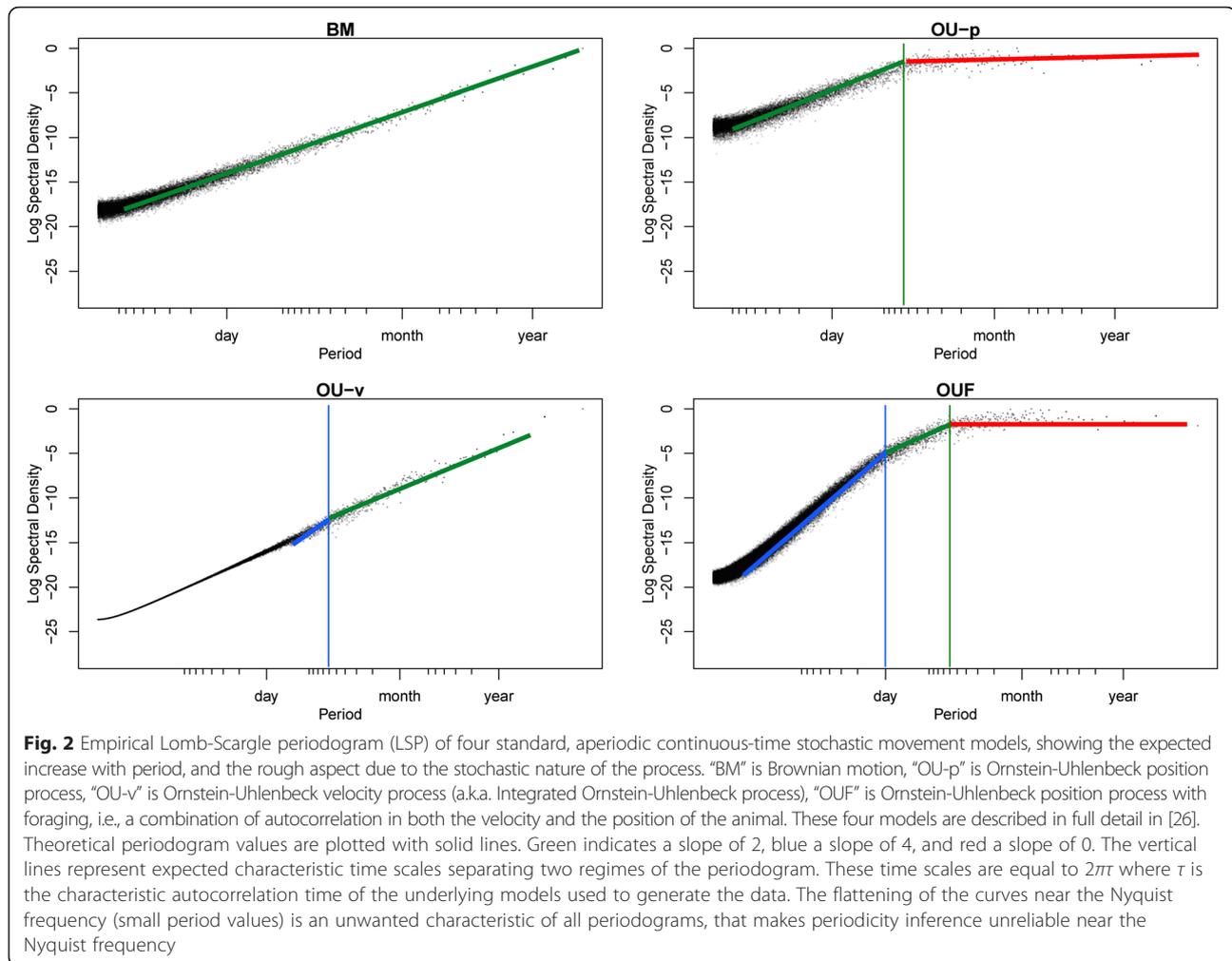
Fig. 1 Empirical computing times for the Lomb-Scargle periodogram (LSP) for 2D time series of varying length, using different R-based routines on a 2.5GHz i5 CPU. All time series featured 50 % randomly selected missing observations, so that the reported sample sizes are half the series' lengths. "ctmm" corresponds to our new, fast algorithm. "lomb" corresponds to the implementation in the **lomb** R-package [22], "cts" is from [34], and "nlts" is from [35]. The steeper slope of the "lomb", "cts", and "nlts" curves is due to these being based on a $O(N^2)$ algorithm, whereas "ctmm" is based on a $O(N \log N)$ algorithm (see main text) and therefore becomes increasingly faster than other R-based alternatives as the sample size increases. The different intercepts for "lomb", "cts", and "nlts" indicate a change in per-iteration efficiency across implementations of the same algorithm

movement data. Instead, the peaks in the periodograms must be compared to the autocorrelated background noise (see the section entitled "Artefactual periodicities and a new statistical test for periodic patterns of space use" below). A graphic representation of the LSP of the four most commonly used aperiodic continuous-time stochastic movement models is provided in Fig. 2. We do not recommend using the LSP as a diagnostic tool to distinguish different aperiodic stochastic models (see instead [26, 32, 38]). However, knowing what the periodogram of an aperiodic animal path should look like can be very useful to interpret empirical periodograms, especially given the rough aspect of most empirical periodograms.

The important features to look for in periodograms are peaks (Additional file 2, section C1). Peaks correspond to frequencies that resonate with the signal. These peaks vary in width and height (Figs. 3, 4 and 5; Additional file 2). Regarding the width of the peaks, as mentioned earlier, the default resolution of the periodogram is, in the frequency domain, $\Delta f = 1/(N \cdot \Delta t)$ where N is the number of recorded locations and Δt is the (median) sampling interval between two records. In the time domain, this yields a resolution that increases with the period: $\Delta T = T^2/(N \cdot \Delta t) \approx T$, meaning that, if the sampling schedule stays the same, the longer the period, the wider the peak. Peak width, therefore, is

typically an artifact of the periodogram's natural resolution, so that longer periods will often have wider peaks (Additional file 2, section C.4). Peak width thereby typically carries little biological information about periodicity in the mean of the movement process. Some of this artefactual variation in peak width can be removed using the `max = TRUE` option of the plot method for periodogram objects in `ctmm`, and by increasing the frequency resolution (the number of points) with the `res.freq` option when computing the periodogram. With the `max=TRUE` option, only local maxima of the periodogram are plotted. This post-hoc coarsening of the frequency resolution removes in particular the artefactual oscillations caused by the autocorrelation in the error term of the periodogram with a period of $1/D$ on the frequency scale ($\sim T^2/D$ on the period scale), where D is the overall study duration. These oscillations, when not discarded using the `max` option, are typically visible for large periods (Figs. 3, 4 and 5).

By contrast, the height of the peaks, i.e., the difference between the periodogram value at the peak and around the peak, conveys information about the signal-to-noise ratio [22, 27]. The statistical significance of the periodic pattern can be inferred by comparing the peak height to the baseline periodogram of the background noise [27] (see more details under "Artefactual periodicities and a new statistical test for periodic patterns of space use" below).



Third, periodogram interpretation needs to take harmonic series into account (section C.2 in Additional file 2). A harmonic is defined as a component frequency of the signal that is an integer multiple of the fundamental frequency (the frequency with the largest signal-to-noise ratio). Harmonic series occur if the periodic signal is not perfectly sinusoidal, i.e., if the waveform is different from a sinusoid, or if the repeated pattern is different from an ellipse. In movement ecology, periodic patterns of space use are therefore expected to always generate a series of harmonics. We performed a simulation study to illustrate the occurrence of harmonics in a few simplified situations (section C.2 in Additional file 2). These simulations illustrate that users should not over-interpret peaks in the periodogram that occur for periods that are a fraction of the fundamental period (the highest peak in the periodogram). Even in very simplified cases, predicting which harmonics would be activated was far from straightforward, and depended on the sampling schedule, the velocity, and the shape of the repeated pattern. In real-life applications, harmonic series are unlikely to bring interpretable information

about the shape of the repeated pattern. The difficulty further lies in identifying whether the signal is multi-periodic (e.g., both daily and hourly periods) or mono-periodic with harmonics. If peaks occur for periods that are not integer multiples of each other, the signal is multi-periodic. If the peak for a short period (e.g., one hour) is higher than the peak for a long period (e.g., one day), the signal is also multi-periodic. In most other situations, the signal is likely monoperiodic with harmonics.

Artefactual periodicities and a new statistical test for periodic patterns of space use

The issue at hand in this section is the identification of peaks in the periodogram that are not related to an underlying periodic behavior, but are instead caused by the sampling schedule. This can for example be the case if the tracking device is duty cycled to alternate between fine and coarse sampling rates, in effect creating a periodic sampling schedule, or if there is autocorrelation in the way missing data occur. In such situations, periodicity in the sampling schedule can be transferred into periodicity

in the animal tracking record, even if the true, underlying path is not periodic (section C.3 in Additional file 2). In most previous applications of the LSP, this issue was largely ignored: researchers assumed missing observations or sampling interval variations occurred without autocorrelation [22]. In movement ecology, the issue cannot be ignored anymore (e.g., [26]), requiring innovative treatments.

In order to jointly deal with the issues of 1) temporally autocorrelated background noise; 2) possible confusion of random peaks with periodicity-induced peaks; and 3) artefactual periodicities created by the sampling schedule, we devised the following two-pronged approach based on a null model simulation routine and a visual diagnostic.

Visual diagnostic

The objective here is to visually compare the periodogram of the location time series, $X(t)$, and the periodogram of the sampling schedule, $w(t)$. If the periodicity observed in the data is artefactual, then the same peak should occur in the periodogram of the sampling schedule and in the periodogram of the locations. After rescaling the two periodograms so that they both have a maximum value of zero (automatically performed by `ctmm`), the periodogram of the recorded locations is plotted alongside the periodogram of w . If a peak occurs in both periodograms for the same period, then it is likely that the detected periodicity is artefactual. However, a true periodicity could plausibly be superimposed on an artefactual one, so that the occurrence of a peak of same periodicity in the two periodograms does not prove the absence of a periodic pattern, but only indicates the risk of false positive due to issues in the sampling schedule. This visual diagnostic approach has the advantage of being fully non-parametric. It can thus be employed on any time series irrespective of which type of underlying movement process generated the data or how irregular the sampling schedule is. It can therefore be used without restriction. The visual diagnostic approach is available in package `ctmm` through the `diagnostic=TRUE` option of the `plot` method for periodogram objects.

Null model approach

Here, we first use an information-theoretic approach to select a preferred aperiodic continuous-time stochastic model and fit it to the data. The theory for these models is developed in e.g., [38] and recommendations for model fitting and model selection with `ctmm` are detailed in [32]. The preferred aperiodic continuous-time stochastic model acts as the null model, representing the hypothesis “there are no periodic patterns of space use in the data”. It is the reference frame against which peaks in the periodogram are going to be compared. We use this model to generate simulated datasets with the same sampling schedule as the

real data. Any irregularity in the sampling schedule or pattern in the way missing data occur is thus carried over in the simulations. Finally, we compute the proportion of the simulated periodograms in which the value at the period of interest is larger than the value from the real data. This proportion is the P -value of the periodicity test (the sensitivity of the test depends on the number of simulations). An analogous approach was proposed for the case where the background noise is independently and identically distributed (“white”) by [27] and is implemented by default in the `lomb` package [22]. Here, we thus generalize this original statistical test of periodicity to the case where the background noise is temporally autocorrelated (“colored”), which will typically be the case with movement data. An R script implementing this approach (assuming a preferred null model has already been selected) is provided in Additional file 3.

African buffaloes

This case study is aimed at 1) illustrating the issue of artefactual periodicities created by the sampling schedule (see also section C.3 in Additional file 2 for a simulated example), and 2) highlighting that periodicities in activity levels need not be translated into periodic patterns of space use. Data came from two African buffalo cows named Cilla and Pepper in the Kruger National Park of South Africa, with hourly GPS position records [39]. Cilla had no missing observations and displayed aperiodic space use, whereas Pepper had many missing observations and a periodogram that exhibited a clear daily periodicity. Based on the visual diagnostic, Calabrese et al. [32] suspected that the periodicity in Pepper’s periodogram was artefactual and caused by a collar malfunction. To prove that point, we discarded data points from Cilla in order to mimic Pepper’s sampling schedule. To do so, we first shifted the sampling schedules so that the time series for the two individuals started at the same time, and then discarded all fixes from Cilla’s tracking record that were collected more than 10 min from a fix from Pepper. This yielded a subsample of Cilla’s locations affected by the same missing data problem as Pepper. We predicted that the periodogram from Cilla’s resampled tracking record would exhibit the same (artefactual) peak as Pepper’s. Sample sizes were 3528 fixes for Cilla, of which 2153 were discarded in the manipulation, and 1726 fixes for Pepper. Lastly, we tested Pepper’s periodicity using the above-described null-model approach. If artefactual, the periodicity should not be significant in this test (P -value >0.05).

We also used Cilla’s (complete) data to illustrate the difference between periodic patterns of space use and periodicity in activity level. The former corresponds to the existence of a characteristic revisitation time between two passages of an animal in a given place,

and is our focus in this paper. The latter corresponds to rest/activity cycles, and has been the focus of most other studies applying periodograms to movement data [4, 14–18]. In this case, African buffaloes characteristically show reduced activity during the mid-day heat, so we predicted a daily periodicity in Cilla's activity level whether or not her space use was periodic. Activity was measured by movement speed, which we quantified as the length of each recorded step divided by the duration of the corresponding time interval.

Waved albatrosses

We analyzed 90-min resolution GPS tracking data from waved albatrosses during their 2008 breeding season [40, 41]. We selected data from four adult birds breeding on Española island in the Galapagos archipelago, that were monitored for more than 45 days (range 47–97 days) and undertook more than one fishing trip during that time. During incubation, the male and female of a pair alternate nest duties and fishing trips, so that the nest is always attended [40]. Most fishing trips are towards nutrient-rich up-welling zones off the coast of Peru about 1200 km from the breeding island. Waved albatross' diet include a large proportion of squids [42] which are mainly available to them at night [6]. Although waved albatrosses exhibit some morphological adaptations such as larger eyes and larger distance between them than related species, we suppose that they still have poor night vision like most other albatrosses [43, 44]. Hence, we formulate the hypothesis that foraging behavior should be constrained by moon light, and therefore exhibit a periodicity of c. 29.5 days.

Maned wolves

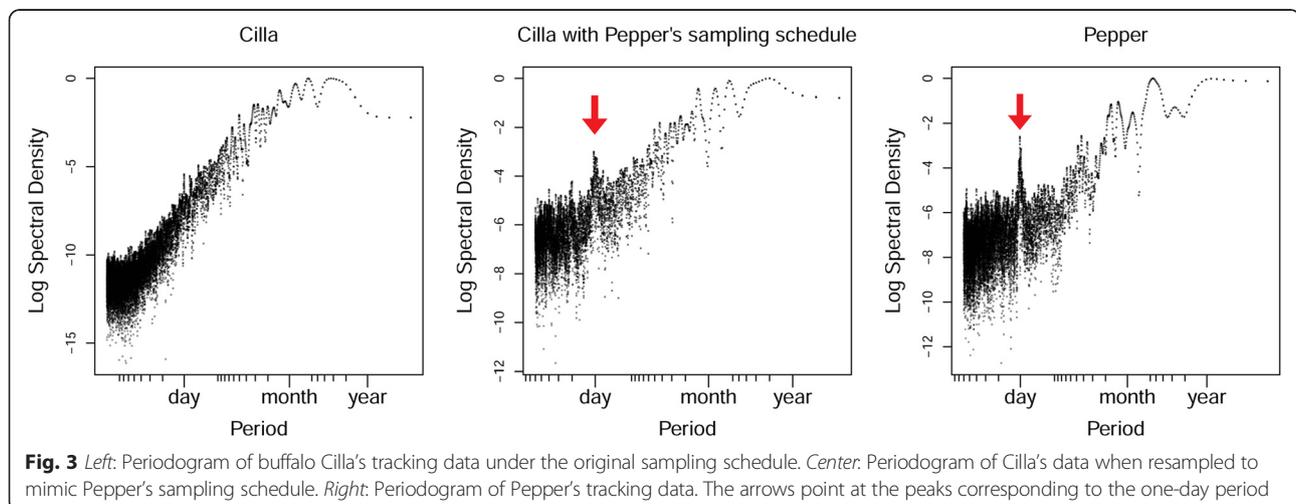
Maned wolf, the largest South American canid, is exceptional among similarly-sized canids for its solitary

foraging behavior and omnivorous diet, which raises questions regarding its daily energy budget [45] and social system [46]. The occurrence of a periodic pattern of space use corresponding to “routine behaviors” is potentially of foremost importance for energy budget and sociality, but remains poorly understood. While there is anecdotal evidence that maned wolves repeatedly use the same paths when patrolling their home range in search of food in Brazil [47], routes are diversified in Bolivia [48]. Interestingly, the behavior is documented in other, more intensively-studied canid species, but even in these species the extent to which it is expressed is variable and poorly known [49, 50]. We used tracking data from eight wolves collected in or near the Serra da Canastra National Park, Brazil, a grassland ecosystem, and used the LSP methodology to test for a daily periodicity in those eight individuals. Maned wolves were captured using live-traps baited with cooked chicken and sardines, sedated (with direct injection of tiletamine-zolazepan), and equipped with VHF/GPS-Collars (Lotek Wireless Inc. GPS 3300S and Iridium Track 1D, and Sir-track Limited Pinnacle Lite G5C 275D). The devices were set to record one location every 1 to 4 h (depending on the animal).

Results

African buffaloes

The LSP of Cilla's locations was characteristic of an aperiodic random walk (no peak), whereas the LSP of Pepper's locations suggested a daily cycle (Fig. 3). When applying Pepper's pattern of missing observations to Cilla's location time series, we retrieved, as predicted, a signal of daily periodicity (Fig. 3). This illustrated that the periodic pattern was created by autocorrelation in the way observations were missing, not the animals' behavior. The P -value of 0.29 from 150 simulations in the null model test applied to Pepper's locations confirmed



that the observed periodic pattern in Pepper's tracking record was artefactual.

When applying the LSP to Cilla's speed record (vs. location record previously), we found, as predicted, a significant daily periodicity (P -value <0.01 from 150 simulations; Additional file 2: Figure C5). Cilla's *activity* was thus periodic even though her *space use* was not (Fig. 3). This illustrates how periodic patterns of space use are decoupled from periodicity in activity levels in this species.

Waved albatrosses

All four albatross individuals exhibited a significant periodicity (all P -values <0.005 from 1000 simulations each) with estimated period ranging from 22 to 35 days. In the individual that undertook the longest trips, reaching waters 1800 km from its breeding colony with an estimated periodicity of 29 days, the association with moon phase was very pronounced (Fig. 4). During three monitored cycles, the bird left the colony shortly before the new moon and started its return journey shortly after the full moon, presumably so that both it and its mate could forage during periods with some moon light.

Maned wolves

The one-day periodicity was statistically significant for all eight individuals (all P -values <0.05 from 1000 simulations each; Fig. 5, right panel), even if the periodic pattern of space use was not directly evident in the raw tracking data (Fig. 5, left panel). In other words, those eight wolves showed a significant tendency to patrol their home range along daily-repeated routes, but this pattern was obscured by the important background noise (the aperiodic stochastic components of the paths).

Discussion

In this study, we introduced an analytical protocol based on the Lomb-Scargle periodogram (LSP) to facilitate exploration of animal tracking datasets for periodic patterns, even if those datasets feature missing observations or variation in the duration of sampling intervals. We described a new, fast algorithm to compute the LSP. The gain in speed makes it tractable to quickly analyze large datasets ($>10^6$ records) that would otherwise require days of computing time with other R packages. We also gave a detailed primer about periodicity analysis, focusing on the specificities of movement data. In particular, we described a methodology to identify artefactual periodicities created by the sampling schedule, an issue that is commonplace in animal tracking data but much rarer in other uses of the LSP. We developed a statistical test of periodicity that applies when the background noise is temporally autocorrelated, which is typically the case in movement data, and that is robust to irregularities in the sampling schedule. We also accommodated multidimensionality (2D, 3D) and the possibility that several animals exhibit the same type of periodic patterns, both features being commonplace in movement datasets. We highlighted that novel biological insights can be gained from analyzing periodicity in the *locations* of the animals, not (only) in their *activity* levels, as is more commonly done. Applying our LSP-based analytical protocol to location data from three species revealed that an ungulate exhibited periodicity in its movement speed but not in its locations, that a central place forager tracked moon phase, and that the paths of a range-resident canid included a daily patrolling component that was initially masked by the stochasticity of the paths.

On the use of periodograms in movement ecology

As outlined above, the study of periodic behaviors can be based on the animals' locations or on their activity

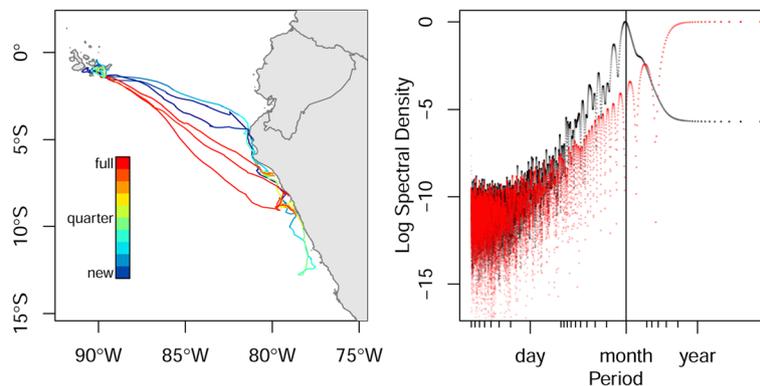
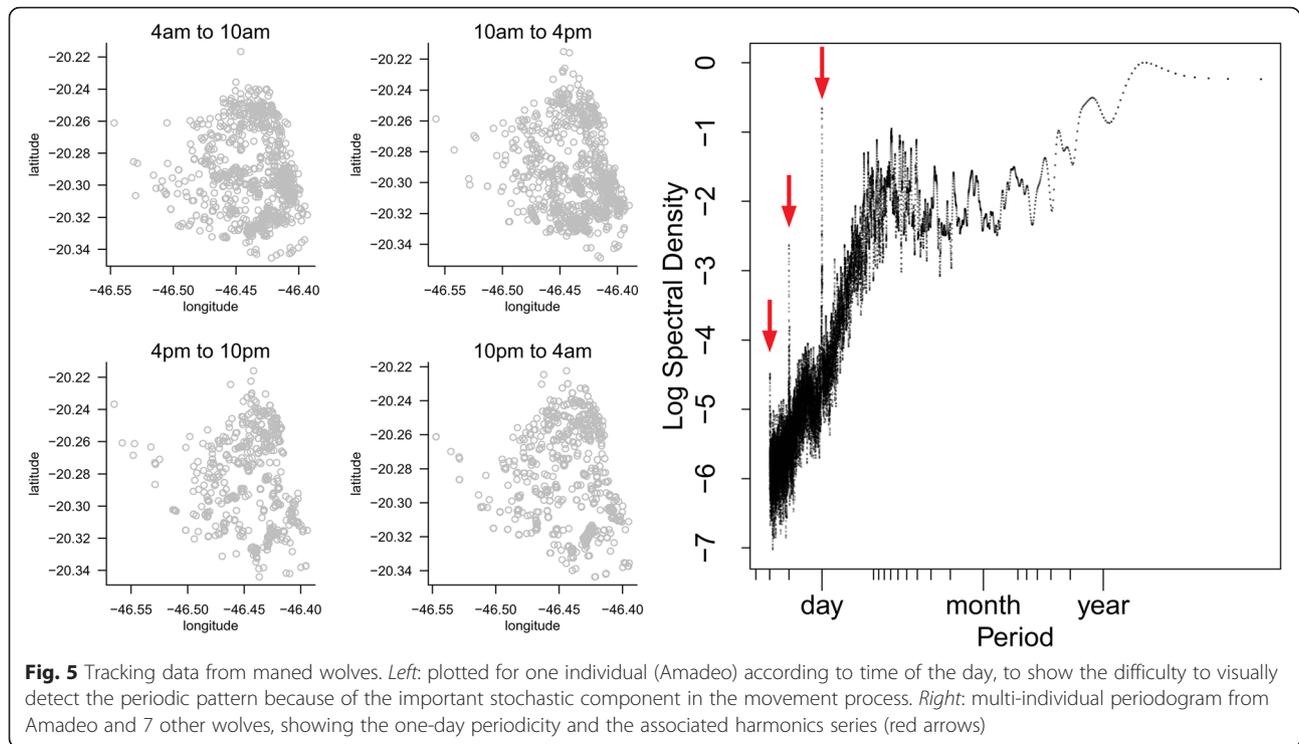


Fig. 4 Tracking data from one waved albatross. *Left*: plotted with color scale indicating moon phase. Blue colors are close to the new moon and red colors to the full moon. *Right*: periodograms. The black periodogram is from the tracking data and the red periodogram is from the sampling schedule. The presence of a peak in the black periodogram for a period of one lunar cycle (vertical line), and its absence in the red periodogram, indicate that the periodic pattern is not caused by the sampling schedule



levels, or both. In all cases, the LSP represents the best available non-parametric tool to detect periodic patterns in imperfect data [22, 24]. Periodicity in activity level and periodic patterns of space use represent two distinct phenomena, as the case of African buffalo Cilla illustrates. Cilla showed no sign of periodicity in her locations, i.e., no periodic pattern of space use (Fig. 3), but a highly significant one-day periodicity in her movement speed, i.e., statistically significant rest/activity cycles (Additional file 2: Figure C5). Rest/activity cycles are expected to occur in almost all species [4, 14–19], but see [5]. By contrast, periodic patterns of space use are expected to be much less widespread, but likely to be of fundamental biological interest when and where they do occur. Periodic patterns of space use are indeed only expected under specific conditions, such as when breeding and foraging habitats are distinct [51], or when forage quality and predation risk are positively correlated [13], or when being predictable allows avoiding direct contacts with neighbors and therefore reduces the potential for territorial conflicts [10]. The few previous studies that focused on periodicity in animals' locations first defined an area of interest and then transformed the tracking record into a series of Booleans representing presence, entry, or exit from that area of interest [13, 52]. This procedure restricts the inference to a specific area of interest that the researcher a priori designates as likely to be revisited periodically. The LSP allows researchers to

avoid making such a priori decisions, and is thus a more broadly applicable data exploration tool.

Semivariogram or periodogram?

Temporal autocorrelation, of which periodic patterns are a particular form, is typically investigated using semivariograms [26, 53]. Semivariograms could therefore theoretically be used for inference about periodicity (Additional file 2, section C.4). However, the empirical semivariogram often suffers strong autocorrelation in its error term, and is not by construct negative definite [53], which can lead to flawed inference about periodicity when the data are noisy and finely sampled. By contrast, as long as the minimum frequency resolution F/N is not exceeded, the empirical periodogram is largely without autocorrelation in its error term, and is by construct always positive [21]. As a side note, the absence of autocorrelation in the error term is also the cause for the rough appearance of most empirical periodograms. In addition, the periodogram is constructed in the frequency domain, which makes visual diagnosis much more straightforward, especially in the presence of multiple periodicities in the data. When the data are relatively coarsely sampled relative to the period of the periodic pattern of space use, the semivariogram can nevertheless yield a useful confirmation of a pattern that might be hard to detect in the periodogram because of the width of the peak (Additional file 2, section C.4). Lastly, neither periodograms nor semivariograms accommodate regime shifts

or non-stationary dynamics, which have been a topic of interest recently [4, 14, 19]. In other words, current versions of empirical periodograms and semivariograms average over non-stationarity and allow users to visualize the time-averaged processes.

Conclusions

In conclusion, we recommend the Lomb-Scargle periodogram (a fast implementation of which is available in the R-package `ctmm`) as a non-parametric, exploratory tool to explore animal tracking datasets for periodic patterns of space use and other periodic patterns such as periodicity in activity level. The often used Discrete Fourier Transform (DFT) requires constant time intervals between subsequent locations records, a condition that is rarely met in animal tracking datasets and therefore forces users to perform ad hoc data manipulations before analysis. The LSP is strictly equivalent to the DFT-periodogram if the time intervals are constant, and does not require ad hoc data manipulation if the time intervals are not constant. We confirm however that both approaches are challenged when sampling intervals vary in a strongly temporally autocorrelated way. Using any periodogram blindly in this case can lead to the detection of artefactual periodicities. Diagnostic tools presented in this study can be used to identify artefactual periodicities created by the sampling schedule, so that further analyses can be adapted accordingly. As animal tracking datasets become longer and finer-resolution, and more individuals can be tracked for the same budget, we envision increased interest in exploratory non-parametric methods like the periodogram, which pave the way for case-tailored parametric tests and movement models.

Endnotes

¹As a reminder, the frequency of a signal is the inverse of its period.

²In the `ctmm` implementation, we use the `fftw` R package, which is a wrapper around the FFTW C-library [54], and is faster and more robust than R's built in FFT routines. By default, we inflate the frequency resolution until the grid size is a power of two, to avoid slowing down the FFT. This excess frequency resolution can be removed with the `max = TRUE` option of our plot method for periodogram objects.

Additional files

Additional file 1: The Lomb-Scargle periodogram in movement ecology. (PDF 229 kb)

Additional file 2: Simulation study. (PDF 581 kb)

Additional file 3: R code for a statistical test of periodicity in animal tracking data. (R 1 kb)

Abbreviations

DFT, discrete Fourier transform; FFT, fast Fourier transform; GPS, global positioning system; LSP, Lomb-Scargle periodogram; OU, Ornstein-Uhlenbeck process

Funding

This work was supported by the US NSF Advances in Biological Informatics program (ABI-1458748 to JMC). GP was supported by a Smithsonian Institution CGPS grant, and CHF was supported by a Smithsonian Institution postdoctoral fellowship. We thank Paul Cross for providing the African buffalo data. The buffalo data collection was supported by the US NSF and NIH Ecology of Infectious Disease program (DEB-0090323 to W.M. Getz).

Authors' contributions

GP, CF, and JC conceived of the study. CF conducted the theoretical developments and coded the LSP function in the R package. RdP collected the data. GP performed the statistical analyses and drafted the manuscript. All authors contributed to and approved the final manuscript.

Competing interest

The authors declare that they have no competing interests.

Ethics approval and consent to participate

The maned wolf research was conducted under government permit ICMBio/SISBIO 11124.

Author details

¹Smithsonian Conservation Biology Institute, National Zoological Park, Front Royal, VA 22630, USA. ²Department of Biology, University of Maryland, College Park, MD 20742, USA. ³National Research Center for Carnivore Conservation (CENAP/ICMBio), Atibaia, Sao Paulo, Brazil.

Received: 6 February 2016 Accepted: 18 June 2016

Published online: 01 August 2016

References

- Boitani L, Barrasso P, Grímódi I. Ranging behaviour of the red fox in the Gran Paradiso National Park (Italy). *Ital J Zool.* 1984;51:275–84.
- Berthold P, Gwinner E, Sonnenschein E, editors. *Avian Migration*. Berlin: Springer; 2003.
- Panda S, Hogenesch JB, Kay SA. Circadian rhythms from flies to human. *Nature.* 2002;417:329–35.
- Polansky L, Wittemyer G, Cross PC, Tambling CJ, Getz WM. From moonlight to movement and synchronized randomness: Fourier and wavelet analyses of animal location time series data. *Ecology.* 2010;91:1506–18.
- van Oort BEH, Tyler NJC, Gerkema MP, Folkow L, Blix AS, Stokkan K-A. Circadian organization in reindeer. *Nature.* 2005;438:1095–6.
- Cruz SM, Hooten M, Huyvaert KP, Proaño CB, Anderson DJ, Afanasyev V, Wikelski M. At-sea behavior varies with lunar phase in a nocturnal pelagic seabird, the swallow-tailed gull. *PLoS One.* 2013;8:e56889.
- Bar-David S, Bar-David I, Cross PC, Ryan SJ, Knechtel CU, Getz WM. Methods for assessing movement path recursion with application to African buffalo in South Africa. *Ecology.* 2009;90:2467–79.
- Gautestad AO. Memory matters: influence from a cognitive map on animal space use. *J Theor Biol.* 2011;287:26–36.
- Péron G, Lebreton J-D, Crochet P-A. Breeding dispersal in black-headed gull: The value of familiarity in a contrasted environment. *J Anim Ecol.* 2010;79:317–26.
- Laidre KL, Born EW, Gurarie E, Wiig Ø, Dietz R, Stern H. Females roam while males patrol: divergence in breeding season movements of pack-ice polar bears (*Ursus maritimus*). *Proc R Soc Biol Sci.* 2013;280:20122371.
- Fleming CH, Fagan WF, Mueller T, Olson KA, Leimgruber P, Calabrese JM. Rigorous home-range estimation with movement data: A new autocorrelated kernel-density estimator. *Ecology.* 2015;96:1182–8.
- Kranstauber B, Kays R, Lapoint SD, Wikelski M, Safi K. A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *J Anim Ecol.* 2012;81:738–46.
- Riotte-Lambert L, Benhamou S, Chamaillé-Jammes S. Periodicity analysis of movement recursions. *J Theor Biol.* 2013;317:238–43.
- Wittemyer G, Polansky L, Douglas-Hamilton I, Getz WM. Disentangling the effects of forage, social rank, and risk on movement autocorrelation of

- elephants using Fourier and wavelet analyses. *Proc Natl Acad Sci U S A*. 2008;105:19108–13.
15. Jolivet A, Chauvaud L, Thébault J, Robson AA, Dumas P, Amos G, Lorrain A. Circadian behaviour of *Tectus (Trochus) niloticus* in the southwest Pacific inferred from accelerometry. *Mov Ecol*. 2015;3:26.
 16. Ehlinger GS, Tankersley RA. Endogenous rhythms and entrainment cues of larval activity in the horseshoe crab *Limulus polyphemus*. *J Exp Mar Bio Ecol*. 2006;337:205–14.
 17. Campos MC, Costa JL, Quintella BR, Costa MJ, Almeida PR. Activity and movement patterns of the Lusitanian toadfish inferred from pressure-sensitive data-loggers in the Mira estuary (Portugal). *Fish Manag Ecol*. 2008; 15:449–58.
 18. Helm B, Visser ME. Heritable circadian period length in a wild bird population. *Proc R Soc B Biol Sci*. 2010;277:3335–42.
 19. Heurich M, Hilger A, Küchenhoff H, Andrén H, Bufka L, Krofel M, Mattisson J, Odde J, Persson J, Rauset GR, Schmidt K, Linnell JDC. Activity patterns of Eurasian lynx are modulated by light regime and individual traits over a wide latitudinal range. *PLoS One*. 2014;9:e114143. Public Library of Science.
 20. Scargle JD. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys J*. 1982;263:835–53.
 21. Smith SW. Chapter 8: The Discrete Fourier Transform. *The Scientist and Engineer's Guide to Digital Signal Processing*. Secondth ed. San Diego, CA: California Technical Publishing; 1999. p. 141–68.
 22. Ruf T. The Lomb-Scargle periodogram in biological rhythm research: Analysis of incomplete and unequally spaced time-series. *Biol Rhythm Res*. 1999;30:178–201.
 23. Glynn EF, Chen J, Mushegian AR. Detecting periodic patterns in unevenly spaced gene expression time series using Lomb-Scargle periodograms. *Bioinformatics*. 2006;22:310–6.
 24. Schulz M, Statterger K. Spectrum: spectral analysis of unevenly spaced paleoclimatic time series. *Comput Geosci*. 1997;23:929–45.
 25. Press WH, Rybicki GB. Fast algorithm for spectral analysis of unevenly sampled data. *Astrophys J*. 1989;338:277.
 26. Fleming CH, Calabrese JM, Mueller T, Olson KA, Leimgruber P, Fagan WF. From fine-scale foraging to home ranges: a semivariance approach to identifying movement modes across spatiotemporal scales. *Am Nat*. 2014;183:e154–67.
 27. Nemeč AFL, Nemeč JM. A test of significance for periods derived using phase-dispersion-minimization techniques. *Astron J*. 1985;90:2317–20.
 28. Rockmore DN. The FFT: an algorithm the whole family can use. *Comput Sci Eng*. 2000;2:60–4.
 29. Strandburg-Peshkin A, Farine DR, Couzin ID, Crofoot MC. Shared decision-making drives collective movement in wild baboons. *Science*. 2015;348:1358–61.
 30. Singh NJ, Börger L, Dettki H, Bunnefeld N, Ericsson G. From migration to nomadism: movement variability in a northern ungulate across its latitudinal range. *Ecol Appl*. 2012;22:2007–20.
 31. Fleming CH, Calabrese JM. ctm: Continuous-Time Movement Modeling. R package version 0.3.2. <http://cran.r-project.org/package=ctm>. 2016. Accessed 23 Jun 2016.
 32. Calabrese JM, Fleming CH, Gurarie E. ctm: an r package for analyzing animal relocation data as a continuous-time stochastic process. *Methods Ecol Evol*. 2016. In press.
 33. Kleinberg J, Tardos É. *Network flow*. Algorithm Des. Boston, USA: Pearson/Addison-Wesley; 2006. p. 337–450.
 34. Wang Z. cts: an R package for continuous time autoregressive models via Kalman filter. *J Stat Softw*. 2013;53:1–19.
 35. Bjornstad ON. Package "nlts" <https://cran.r-project.org/web/packages/nlts/>. 2013. Accessed 23 Jun 2016.
 36. Spoelstra K. ChronoShop software for activity cycle analysis Version 1.04. <http://www.lichtopnatuur.org/files/ChronoShop.exe>. Wageningen: Netherland Institute of Ecology; 2015. Accessed 23 Jun 2016.
 37. Refinetti R, Lissen GC, Halberg F. Procedures for numerical analysis of circadian rhythms. *Biol Rhythm Res*. 2007;38:275–325. <http://www.circadian.org/software.html>. Accessed 23 June 2016.
 38. Fleming CH, Calabrese JM, Mueller T, Olson KA, Leimgruber P, Fagan WF. Non-Markovian maximum likelihood estimation of autocorrelated movement processes. *Methods Ecol Evol*. 2014;5:462–72.
 39. Getz WM, Fortmann-Roe S, Cross PC, Lyons AJ, Ryan SJ, Wilmers CC. LoCoH: nonparametric kernel methods for constructing home ranges and utilization distributions. *PLoS One*. 2007;2:e207.
 40. Dodge S, Bohrer G, Weinzierl R, Davidson SC, Kays R, Douglas D, Cruz S, Han J, Brandes D, Wikelski M. The environmental-data automated track annotation (Env-DATA) system: linking animal tracks with environmental data. *Mov Ecol*. 2013;1:3. BioMed Central Ltd.
 41. Cruz S, Proaño CB, Anderson D, Huyvaert K, Wikelski M. Data from: The Environmental-Data Automated Track Annotation (Env-DATA) System: Linking animal tracks with environmental data. 2013. doi:10.5441/001/1.3hp3s250.
 42. Harris MP. The biology of the waved albatross *Diomedea irrorata* of Hood Island, Galapagos. *Ibis*. 1973;115:483–510.
 43. Harrison CCS, Hida TTS, Seki MMP. Hawaiian seabird feeding ecology. *Wildl Monogr*. 1983;85:3–71.
 44. Phalan B, Phillips RA, Silk JRD, Afanasyev V, Fukuda A, Fox J, Catry P, Higuchi H, Croxall JP. Foraging behaviour of four albatross species by night and day. *Mar Ecol Prog Ser*. 2007;340:271–86.
 45. Carbone C, Mace GM, Roberts SC, Macdonald DW. Energetic constraints on the diet of terrestrial carnivores. *Nature*. 1999;402:286–8. Macmillan Magazines Ltd.
 46. Macdonald DW, Creel S, Mills MGL. *Society*. In: Macdonald DW, Sillero-Zubiri C, editors. *Biology and Conservation of Wild Canids*. Oxford, UK: Oxford University Press; 2004. p. 85–106.
 47. Motta Junior JC, Martins K, Levey DJ, Silva WR, Galetti M. The frugivorous diet of the maned wolf in Brazil. In: Levey DJ, Silva WR, Galetti M, editors. *Seed dispersal and frugivory: Ecology, Evolution and Conservation*. Wallingford: CAB publishing; 2002. p. 291–303.
 48. Emmons LH, Chávez V, Del Aguila LF, Angulo S, Muir M. Ranging patterns. The Maned Wolves of Noel Kempff Mercado National Park Smithsonian Contributions to Zoology. Washington D.C: Smithsonian Institution Press; 2012. p. 25–35.
 49. Cavallini P. Variation in the social system of the red fox. *Ethol Ecol Evol*. 1996;8:323–42. Taylor & Francis Group.
 50. Doncaster C, Macdonald D. Drifting territoriality in the red fox *Vulpes vulpes*. *J Anim Ecol*. 1991;60:423–39.
 51. Weimerskirch H. Are seabirds foraging for unpredictable resources? *Deep Sea Res Part II Top Stud Oceanogr*. 2007;54:211–23.
 52. Li Z, Han J, Ding B, Kays R. Mining periodic behaviors of object movements for animal and biological sustainability studies. *Data Min Knowl Discov*. 2011;24:355–86.
 53. Cressie NAC. *Statistics for spatial data*. New York: Wiley; 1993.
 54. Frigo M, Johnson SG. The design and implementation of FFTW3. *Proc IEEE*. 2005;93: 216–31.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

