NOTE

Detecting lunar cycles in marine ecology: periodic regression versus categorical ANOVA

Adrian M. H. deBruyn[1],*, Jessica J. Meeuwig[1,2]

[1]Department of Biology, [2]Project Seahorse, McGill University, 1205 Dr. Penfield Avenue, Montreal, Quebec H3A 1B1, Canada

ABSTRACT: Lunar cycles are commonly observed in the movement, feeding and reproduction of marine fishes and invertebrates. The statistical techniques employed to examine these cycles are unstandardized, complex, and typically lacking in statistical power. Here we suggest a simple, sensitive and robust alternative for the detection of cyclical patterns: periodic regression. We use Monte Carlo simulation to demonstrate that periodic regression is more powerful and less sensitive to missing data than categorical ANOVA (the most commonly employed technique in the literature). Finally, we use real seahorse bycatch data to show that periodic regression is superior to categorical ANOVA for the detection and description of more complex cycles. We encourage researchers to use periodic regression in the analysis of lunar cycles and other cyclical patterns of known period.

KEY WORDS: Lunar cycles · Statistical techniques · Periodic regression

The lunar cycle provides a strong, predictable set of environmental cues for marine species. Environmental cycles (e.g., tidal water movement, moonlight) entrain endogenous reproductive cycles, synchronizing gamete release within a population and ensuring that movement, feeding and reproduction occur under favourable conditions (Taylor 1984, Omori 1995). Lunar-synchronized spawning, for example, is commonly documented for species of shallow waters with large tidal fluctuations (Korringa 1947, Taylor 1984). Reef fishes often mass in spawning aggregations on a specific lunar and seasonal cycle (Johannes 1981, Robertson et al. 1990). In addition, lunar cycles have been detected in spawning and settlement of intertidal (Taylor 1984) and pelagic-spawning fish and invertebrates (Crabtree 1995, Robertson et al. 1999). Lunar

cycles in fish behaviour have long been recognized and exploited by artisinal and commercial fisheries (Johannes 1981, Parrish 1999). The implications of lunar cycles for the design of sampling programs are discussed by Gaudreau & Boisclair (2000).

A bewildering array of statistical techniques has been employed to test for lunar cycles. Most common is a categorical treatment, in which the lunar cycle is subdivided into phases. Examples include multiway (Taylor et al. 1998), split-plot (Courtney et al. 1996) or repeated measures ANOVA (Rooker & Dennis 1991), Kruskal-Wallis nonparametric ANOVA (von Herbing & Hunte 1991), chi-square contingency (Nakai et al. 1990), and correspondence analysis (Laroche et al. 1997). Most authors define 4 lunar phases, but this ranges from 2 (Grau et al. 1981) to 25 (Coblentz 1995). Examples of continuous treatments of lunar cycles are less common, but include Kolmogorov-Smirnov goodness-of-fit (Rudloe 1985), Spearman's rank correlation (Oxenford et al. 1995), time-series analysis (Robertson et al. 1990), generalized additive models (Bigelow et al. 1999), and the Rayleigh test for uniform angular distribution (Robertson et al. 1990). Finally, there are a large number of papers that forego statistical analysis and use simple graphical representations to demonstrate lunar periodicity (Dufour et al. 1996).

Here we suggest a simple, sensitive, and robust alternative to the above tests for the detection of cyclical patterns of known period: periodic regression (Batschelet 1981). In periodic regression (also called linear-circular regression), the independent variable is an angular representation of time. As with other forms of regression analysis, this approach has clear advantages over categorical tests when the independent variable is continuous. Regression techniques are less sensitive to missing data, and if the remaining data are not evenly spaced, a categorical treatment will probably be unbalanced, with the associated loss of

---
*E-mail: adebruyn@bio1.lan.mcgill.ca

statistical power. Regression also does not require the arbitrary subdivision of a continuous variable into a set of categories, an approach which discards information and reduces statistical power. Furthermore, the definition and allocation of such categories (phases) is partially subjective and likely to influence statistical outcomes. Periodic regression is simple, flexible, and can be performed on any commonly used statistics software.

The period of a lunar cycle is fixed *a priori* at 29.53 d. Parameters to be estimated in periodic regression are the amplitude of the rhythm (the range of oscillation of *Y*) and the phase angle (the point on the circular time scale at which *Y* is maximum). The procedure is simple: the lunar month is divided into 360° (or 2π radians) to give each day an angular equivalent, θ. The transformed data are then analyzed by simple linear regression, using the model:

$$Y = b_0 + b_1 \cdot \sin\theta + b_2 \cdot \cos\theta$$

where *Y* is the dependent variable, b_0 is the mean level of *Y*, and b_1 and b_2 are model coefficients which together define the phase shift and amplitude (Batschelet 1981). A cosine term is sufficient to describe a phase shift near 0° or 180°, and a sine term is sufficient to describe a phase shift near 90° or 270°. In these cases, the coefficient is simply the amplitude of the oscillation. Other cycles require both terms, but the phase shift and amplitude may be readily calculated from b_1 and b_2 (Batschelet 1981). A semilunar cycle (2 peaks per lunar month) may be analyzed using the terms $\sin 2\theta$ and $\cos 2\theta$; if the 2 peaks are substantially unequal in amplitude, the latter model will also require either $\sin\theta$ or $\cos\theta$ (e.g. Davis & West 1993).

We used a Monte Carlo simulation to compare the abilities of ANOVA and periodic regression to detect a lunar cycle. A simulated 29 d cycle was generated with the function:

$$Y = A \cdot \cos\theta + s \cdot rand$$

where *A* was the desired amplitude of the cycle, *s* the desired standard deviation of the cycle, and *rand* a number randomly selected from a normal distribution with mean 0 and standard deviation 1. Including only the cosine term fixed the phase shift of the resulting cycle at 0°. Amplitude was held constant at 6, while *s* was varied from 1 to 9, giving a series of cycles with signal:noise ratios ranging from 6 to 2/3. The 9 cycles were then analyzed (1) by simple linear regression of *Y* on $\cos\theta$, and (2) by 1-way ANOVA of lunar phase (4 levels). Lunar phase categories were defined to maximize the power of ANOVA while using

as many data as possible. Each phase comprised 7 d, centered on the peak day of the lunar phase. To compare the 2 techniques across a range of sampling intensities, analyses were performed on complete data (n = 29), on only even days (n = 14), and on only every fourth day (n = 7). Each simulation was repeated 250 times. The simulated cycle is shown in Fig. 1. We estimated the power of regression and ANOVA to detect the simulated cycle as the proportion of the 250 simulations in which the technique returned a p-value less than the standard alpha level of 0.05. We did not compare the techniques by the coefficients of determination (r^2), as r^2 is sensitive to the model degrees of freedom and these differ between regression and ANOVA.

With 29 data per cycle, ANOVA displayed good statistical power (>0.8) at signal:noise ratios as low as 1.2. At lower sampling densities, the power of this technique declined precipitously (Fig. 2). Power of ANOVA exceeded 0.8 at only the highest signal:noise ratio with 14 data, and never exceeded 0.1 with 7 data. Periodic regression displayed excellent power to detect the cycle at signal:noise ratios as low as 0.86 (complete data). Moreover, it was relatively insensitive to more widely spaced data, retaining good power at signal:noise ratios of 1.2 with 14 data and 3 with only 7 data (Fig. 2).

Next, we compared regression and ANOVA by analyzing daily catch per unit effort (CPUE) for seahorses *Hippocampus* spp. caught as bycatch in a Vietnam trawl fishery (A. Vincent unpubl. data). We analyzed data collected from June to December 1996 (n = 201 nights), a period that includes both the high CPUE season (June–October) and the low CPUE season (November–December). The categorical ANOVA model
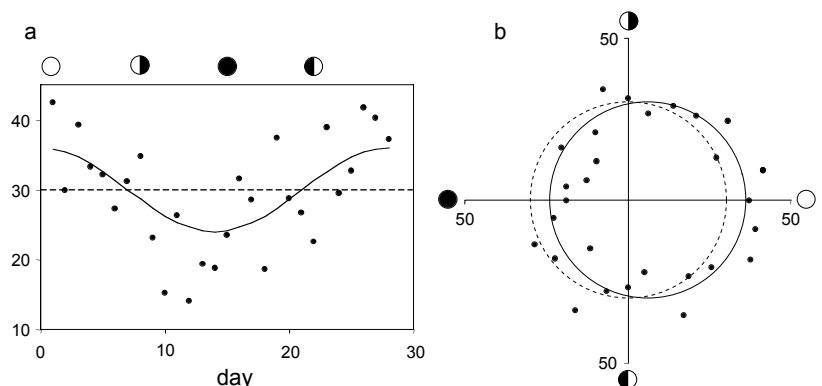

Fig. 1. Simulated lunar cycle (solid line) and an example of simulated data with standard deviation equal to the amplitude of the cycle (points). The mean elevation of the *Y* variable is shown as a dashed line. Simulated data are plotted across a lunar cycle, (a) on linear axes, and (b) on polar axes, where day is expressed as the angle θ and *Y* is distance from the origin. The shift between circles in (b) is a circular representation of the difference between a constant value (dashed) and a cosine function (solid)
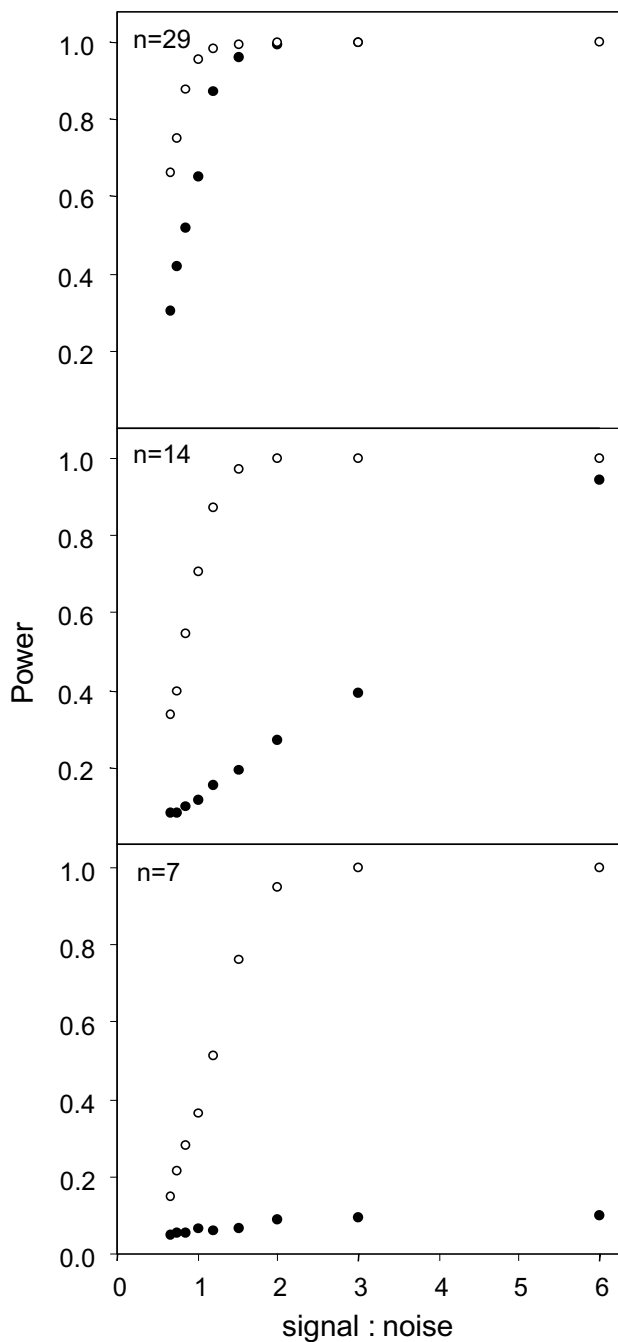
Fig. 2. Statistical power of categorical ANOVA (●) and periodic regression (○) to detect a simulated lunar cycle. Results are shown for analyses of complete data (upper panel), every second day (middle panel), and every fourth day (lower panel)

included lunar phase (4 equal-size levels, defined as above), season (2 levels), and the interaction term. Regression was performed as an analysis of covariance (ANCOVA), with the continuous variable θ representing day of the lunar month, a categorical variable for season, and the interaction term.

Categorical ANOVA revealed a significant main effect of season ($F = 14.7$; $p = 0.00017$), a significant main effect of lunar phase ($F = 2.75$; $p = 0.044$), and an interaction ($F = 2.85$; $p = 0.039$). One-way ANOVA of each season, followed by Tukey multiple comparisons, indicated an effect of lunar phase in the high season ($F = 4.07$; $p = 0.00834$), with lower CPUE around the new moon than either the first quarter ($p = 0.045$) or full moon ($p = 0.004$). ANOVA could detect no main effect of lunar phase in the low season ($F = 2.06$; $p = 0.12$).

Regression analysis (ANCOVA) also revealed a main effect of season ($F = 15.6$; $p = 0.00011$), but this approach detected a *semilunar* cycle ($r^2 = 0.19$) described by $\sin 2\theta$ ($F = 17.1$; $p = 0.00005$) and $\cos \theta$ ($F = 3.95$; $p = 0.048$):

$$\log \mathrm{CPUE} \ = \ 0.13 \cdot \sin 2\theta - 0.14 \cdot \cos \theta$$

This indicates a periodicity in CPUE with 2 unequal peaks per lunar cycle. Regression analysis also indicated a marginally significant interaction between $\cos \theta$ and season ($F = 3.23$; $p = 0.07$), indicating that the inequality of the 2 peaks may differ between seasons. We therefore performed a simple multiple regression analysis for each season separately. The high season exhibited a cycle similar to that described above ($\sin 2\theta$, $\cos \theta$; $r^2 = 0.17$; Fig. 3), but no cycle was detected in the low season ($p = 0.09$).

Periodic regression is a more powerful and robust technique than categorical ANOVA for the detection of cyclical patterns of known period. Periodic regression was able to detect a weaker simulated lunar cycle and was less sensitive to missing data than ANOVA. When applied to real data, periodic regression readily detected and described a semilunar cycle in seahorse bycatch, where ANOVA was able to detect only a single peak. Furthermore, periodic regression provided an equation describing the cycle. Such an equation may be used to estimate the value of the *Y* variable on any day in the cycle, whereas ANOVA only provides estimates of the mean value of *Y* for each phase. Incorporating a categorical effect of season into the periodic regression analysis was simple, and a continuous seasonal cycle could have been added to the model as a second set of periodic terms (spherical regression).

We must stress that as with any regression analysis, it is critically important not to overfit a periodic model. A number of more complex terms can be added to periodic regression to more closely describe the shape (e.g. skewness and kurtosis) of observed cycles (Batschelet 1981). Additional terms can always improve regression model parameters, but this improvement is frequently spurious. Overly complex models sacrifice ecological realism and defy interpretation. We recommend restrict-
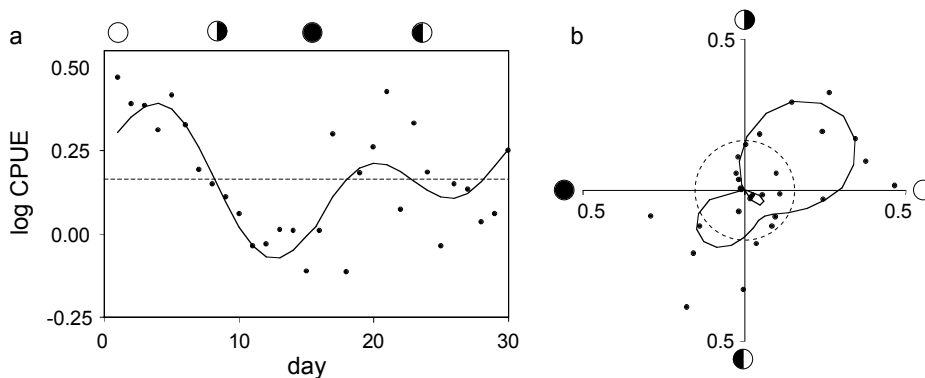
Fig. 3. Daily catch per unit effort (CPUE) of seahorse bycatch in a Vietnam trawl fishery during the high-CPUE season, shown on (a) linear and (b) polar axes. Points are averages across 5 mo in 1996, aggregated by lunar day, solid line is the fitted periodic model ($\log CPUE = 0.16 + 0.13 \cdot \sin 2\theta + 0.14 \cdot \cos\theta$), and dashed line is the mean value of CPUE

ing the model to those terms employed here, unless data (or theory) strongly support the inclusion of more complex terms.

We encourage researchers to use periodic regression in the analysis of lunar cycles. Our findings based on simulated data are consistent with previous work, in which ANOVA failed to detect an 'obvious cyclic trend' in reproduction of Australian penaeid prawns, while periodic regression was able to explain 93% of the variation among samples (Courtney et al. 1996). Furthermore, our analysis of seahorse data demonstrates how regression provides more information than ANOVA about the nature of the lunar cycle.

LITERATURE CITED

Batschelet E (1981) Circular statistics in biology. Academic Press, London

Bigelow KA, Boggs CH, He X (1999) Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. Fish Oceanogr 8:178–198

Coblentz BE (1995) Reproductive biology of the dwarf herring (*Jenkinsia lamprotaenia*) in the Virgin Islands. Bull Mar Sci 56:602–608

Courtney AJ, Die DJ, McGilvray JG (1996) Lunar periodicity in catch rate and reproductive condition of adult eastern king prawns, *Penaeus plebejus*, in coastal waters of south-eastern Queensland, Australia. Mar Freshw Res 47:67–76

Crabtree RE (1995) Relationship between lunar phase and spawning activity of Tarpon, *Megalops atlanticus*, with notes on the distribution of larvae. Bull Mar Sci 56:895–899

Davis TLO, West GJ (1993) Maturation, reproductive seasonality, fecundity, and spawning frequency in *Lutjanus vittus* (Quoy and Gaimard) from the North West Shelf of Australia. Fish Bull 91:224–236

Dufour V, Riclet E, Lo-Yat A (1996) Colonization of reef fishes at Moorea Island, French Polynesia: temporal and spatial variation of the larval flux. Mar Freshw Res 47:413–422

Gaudreau N, Boisclair D (2000) Influence of moon phase on acoustic estimates of the abundance of fish performing daily horizontal migration in a small oligotrophic lake. Can J Fish Aquat Sci 57:581–590

Grau EG, Dickhoff WW, Nishioka RS, Bern HA, Folmar LC (1981) Lunar phasing of the thyroxine surge preparatory to seaward migration of salmonid fish. Science 211:607–609

Johannes RE (1981) Words of the lagoon: fishing and marine lore in the Palau District of Micronesia. UCLA Press, Los Angeles

Korringa P (1947) Relations between the moon and the periodicity in the breeding of marine animals. Ecol Monogr 17:349–381

Laroche J, Baran E, Rasoanandrasana NB (1997) Temporal patterns in a fish assemblage of a semiarid mangrove zone in Madagascar. J Fish Biol 51:3–20

Nakai K, Yanagisawa Y, Sato T, Nimura Y, Gashagaza MM (1990) Lunar synchronization of spawning in cichlid fishes of the tribe Lamprologini in Lake Tanganyika. J Fish Biol 37:589–598

Omori K (1995) The adaptive significance of a lunar or semi-lunar reproductive cycle in marine animals. Ecol Model 82:41–49

Oxenford HA, Mahon R, Hunte W (1995) Distribution and relative abundance of flyingfish (Exocoetidae) in the eastern Caribbean. III. Juveniles. Mar Ecol Prog Ser 117:39–47

Parrish JK (1999) Using behavior and ecology to exploit schooling fishes. Environ Biol Fishes 55:157–181

Robertson DR, Petersen CW, Brawn JD (1990) Lunar reproductive cycles of benthic-brooding reef-fishes: reflections of larval biology or adult biology? Ecol Monogr 60:311–329

Robertson DR, Swearer SE, Kaufmann K, Brothers EB (1999) Settlement vs. environmental dynamics in a pelagic-spawning reef fish at Caribbean Panama. Ecol Monogr 69:195–218

Rooker JR, Dennis, GD (1991) Diel, lunar and seasonal changes in a mangrove fish assemblage off southwestern Puerto Rico. Bull Mar Sci 49:684–698

Rudloe A (1985) Variation in the expression of lunar and tidal behavioral rhythms in the horseshoe crab, *Limulus polyphemus*. Bull Mar Sci 36:388–395

Taylor MH (1984) Lunar synchronization of fish reproduction. Trans Am Fish Soc 113:484–493

Taylor RG, Grier HJ, Whittington JA (1998) Spawning rhythms of common snook in Florida. J Fish Biol 53:502–520

von Herbing IH, Hunte W (1991) Spawning and recruitment of the bluehead wrasse *Thalassoma bifasciatum* in Barbados, West Indies. Mar Ecol Prog Ser 72:49–58