GLMMcosinor: Flexible cosinor modeling with a generalized linear mixed modeling framework to characterize rhythmic time series.

Authors:

Rex Parsons[1,†], Oliver Jayasinghe[2,†], Nicole White[1], Prasad Chunduri[2], Oliver Rawashdeh[2,*]

* Denotes corresponding author.

[†] Denotes equal contribution.

1. Australian Centre for Health Services Innovation and Centre for Healthcare Transformation, School of Public Health and Social Work, Faculty of Health, Queensland University of Technology, Kelvin Grove, Australia
2. The School of Biomedical Sciences, Faculty of Medicine, The University of Queensland, Brisbane, St Lucia, Australia

Key Points

- GLMMcosinor is an R Package for flexible cosinor modeling, a method used to estimate cyclic rhythm characteristics.
- It uses a generalized linear mixed modeling framework (glmmTMB) to allow the user to fit data with non-Gaussian response variables and hierarchical data structures.
- GLMMcosinor is a well documented and feature rich package allowing users to easily fit, interpret and visualize cosinor models.
- GLMMcosinor supports modeling of a wider range of experimental and observation datasets.

Abstract

The complexity, volume, and importance of time series data across various research domains highlight the necessity for tools that can efficiently analyze, visualize, and extract insights. Cosinor modeling is a widely used methodology to estimate or compare rhythmic characteristics in time series datasets. Time series are widely used in biomedical and clinical research studies, with a large amount of research focusing on circadian rhythms in physiology and their relationship to health outcomes. However, existing software for cosinor modeling fails to effectively equip researchers to analyze their data, often due to the hierarchical structure of the data (e.g., repeated measures over time) or non-Gaussian response variables being modeled. Here, we present GLMMcosinor, an R package for fitting the cosinor model to rhythmic time series, using a Generalized Linear Mixed Modeling framework (glmmTMB). This software extends cosinor modeling to non-Gaussian and hierarchical data due to the flexibility offered by glmmTMB. It offers multiple additional features unavailable in other cosinor modeling packages that use the linearized cosinor model or circacompare (which we previously developed), which uses nonlinear regression. GLMMcosinor includes several additional features to interpret, test, and visualize the produced models and can fit models with multiple cosinor components. A detailed description of the use of GLMMcosinor is available within the package's online documentation and vignettes. The GLMMcosinor R package is available from GitHub (https://github.com/ropensci/GLMMcosinor), CRAN, rOpenSci, and the R-universe. A shiny app is also available and can fit and visualize a model with GLMMcosinor without having to write R code (https://github.com/RWParsons/GLMMcosinor-shinyapp).

1. Introduction

The cyclic expression and occurrence of events with time have been closely studied at the fundamental level and are becoming a central component in preclinical and health research. Many physiological processes, including sleep, metabolism, and gene expression, demonstrate a circadian (~24-hour) rhythm. There are also many non-24 hour rhythms such as the menstrual cycle (infradian rhythm) (Itriyeva, 2022), and the pulsatile release of hormones (e.g., glucocorticoids (Kalafatakis, et al., 2021)) and neurotransmitters (e.g., dopamine) via ultradian rhythms (Blum, et al., 2014). A primary interest of scientists (especially chronobiologists) analyzing cyclic data is to estimate the characteristics of rhythms or differences between groups. Statistical models can be applied to rhythmic time series datasets to estimate different parameters of a rhythm, such as amplitude and cycle length (period). Differences in such parameters between multiple rhythms can also be estimated for rhythmic time series datasets belonging to two or more experimental groups. Quantifying and estimating the characteristics of a rhythm is distinctly different from detecting whether a rhythm exists, for which there are several other methods available (Hughes, et al., 2010; Hutchison, et al., 2015; Obodo, et al., 2023). Our study focuses on advancing the methods to characterize rhythms, not detecting them.

Cosinor models are a popular, regression-based method for analyzing rhythmic data. The general form of a cosinor model is given in Equation 1:

$$\mathrm{Y(t)} = \mathrm{M} + \mathrm{A}\cos\left(\frac{2\pi t}{\tau} - \phi\right) + \mathrm{e(t)} \qquad \text{Equation 1}$$

For a response variable (e.g. locomotor activity or gene expression levels) collected over time, $\mathrm{Y(t)}$, the cosinor model assumes that the response can be described in terms of the Midline Statistic Of Rhythm (MESOR), amplitude, and acrophase of a rhythmic dataset. The MESOR represents the rhythm-adjusted mean; amplitude estimates the difference between the MESOR and the rhythms peak or trough; and acrophase is the time at which a peak of the response variable occurs. Remaining variation (variability in the response variable not explained by the rhythmic characteristics) is assigned to a general residual term $\mathrm{e(t)}$.

The distribution of these residuals can be affected by underlying stochastic processes, as well as random error associated with measuring a particular physiological response.

For unbiased estimates of the rhythmic parameters, residuals should be normally distributed. However, for many biological rhythms, including relative gene expression (Xin, et al., 2021), this

is not the case and transformations are often required to satisfy cosinor model assumptions. Often the assumption of normally distributed residuals does not hold for biological processes as it assumes that the observed response variable can be any positive or negative number, despite only positive values being possible. Generalized linear models (GLMs) offer a more flexible approach allowing modeling of response variables from the exponential family (Dobson and Barnett, 2018). For example, the number of circulating tumor cells (Zhu, et al., 2021), is a form of count data and may be best estimated using a Poisson regression within a GLM framework (Cunningham, et al., 2013). Cosinor modeling using Gamma regression (via a GLM) of physiological measurements including heart rate, melatonin levels and locomotor activity – all of which cannot be negative – has been shown to increase the statistical power of the study and provide more appropriate estimates of circadian parameters (Doyle, et al., 2022).

For rhythms with normally distributed residuals, there are several software packages in R and Python that fit cosinor models and provide estimates for the parameters in Equation 1 (Hou, et al., 2021; Moškon, 2020; Parsons, et al., 2020). However, few packages accommodate cosinor models for non-Gaussian responses, and of those available, their functionality is limited to a small number of distributions (Velikajne and Moškon, 2022). Therefore, there is need for a flexible-approach cosinor modeling that can fit regression of a wider range of response variable distributions.

When modeling data with a hierarchical structure, for example collecting repeated measures from individuals over time, it is best to use a mixed-modeling framework to account for the within-individual variance. Relatively few R packages for estimating the cosinor model allow for mixed-models, with many studies inappropriately fitting the cosinor model without random effects to datasets with repeated measures. cosinoRmixedeffects (Hou, et al., 2021) and circacompare (Parsons, et al., 2020) allow for mixed models but neither allow for the functionality of a GLM to fit data with a non-Gaussian response variable.

The R package introduced here, GLMMcosinor, addresses this need by supporting modeling of rhythmic data while including the large number of possible distributions available within the glmmTMB package. This is important because non-normally distributed responses, that are assumed to be normally distributed, may lead to biased estimates and incorrect interpretations. Using an appropriate modeling approach for tailored data allows researchers to make more accurate and reliable conclusions.

GLMMcosinor fits a cosinor model using a generalized linear mixed modeling framework by first linearizing equation 1:

$$Y(t) = M + \beta x + \gamma z + e(t) \qquad \text{Equation 2}$$

Where $x = \cos\left(\frac{2\pi t}{\tau}\right)$, $z = \sin\left(\frac{2\pi t}{\tau}\right)$, $\beta = A\cos(\phi)$, $\gamma = A\sin(\phi)$.

The benefit of linearizing the cosinor equation is that it can be fit using the glmmTMB: an R package which allows the user to fit generalized linear mixed models. The estimates and standard errors for the parameters in Equation 1 can then be recovered using the delta method, the same approach used by the cosinor R package (Sachs, 2023) and explained in detail, in the context of cosinor modeling, by Mikulich and colleagues (Mikulich, et al., 2003). The delta method has been shown to estimate equivalent variances to using a nonlinear model (Equation 1), which can be more expensive to compute (Mikulich, et al., 2003). An additional benefit of using glmmTMB is that random effects can be easily specified as part of the linear predictor, for example, to account for correlation induced by repeated measures on the same individual.

By building GLMMcosinor over the glmmTMB package, it extends all the functionality of fitting the cosinor model to data with non-Gaussian response variables and hierarchical structure. GLMMcosinor also allows for the specification of models with multiple components, cosinor components grouping variables with 2 or more levels, and several methods to visualize and summarize the model to aid interpretation. For a detailed explanation of how to use the package with examples, see the package vignettes (https://docs.ropensci.org/GLMMcosinor/). The GLMMcosinor R package can also be used via an interactive webapp (https://github.com/RWParsons/GLMMcosinor-shinyapp).

2. GLMMcosinor R package

GLMMcosinor was built using the cosinor R package (Sachs, 2023) as a starting structure. GLMMcosinor therefore includes many similarities to cosinor in how it functions and how the user interacts with it. The main model fitting function, `cglmm()`, has three main arguments:

1. The model formula
2. The data
3. The family (defaults to “Gaussian” for a standard linear regression model)

The `cglmm()` function returns a model object which can be used for generating predictions (`predict()`), statistical inference from the model parameters (`summary()`), and visualizations (`autoplot()` and `polar_plot()`). For more details on how to use these functions, see the vignettes (https://docs.ropensci.org/GLMMcosinor).

The GLMMcosinor R package has been peer-reviewed by rOpenSci and is available on GitHub, CRAN, and the R-universe.

2.1 Model specification and interpretation

As the model uses the glmmTMB package, the formula is specified using the same (lme4-) style (Bates, et al., 2015). For an introduction to this style, see the lme4 documentation. In GLMMcosinor, we introduce the use of an `amp_acro()` call within the model formula to specify cosinor components. The `amp_acro()` requires the user to specify the variable representing time, the period of the cosinor component, and whether it is to be grouped by some other factor variable. The benefit of specifying `amp_acro()` is ease of use, as the function processes the inputs and creates the linear cosinor coefficients required for creating the linearized form of the model (transforming Equation 1 to Equation 2, for example). GLMMcosinor then returns the estimates for these linear coefficients as meaningful parameter estimates for amplitude and acrophase and MESOR.

By using the glmmTMB framework, GLMMcosinor also allows the user to fit zero-inflated and dispersion models by specifying these as separate, right-sided, formulas. Zero-inflated models are frequent in biological modeling when datasets contain a large number of zero values, which could represent true zeros (when there is no response), or false zeros (when there is failure to capture a response) (Gandia, et al., 2023). GLMMcosinor extends the capability of zero-inflated and dispersion models by allowing the use of `amp_acro()` within their formulas. The vignettes include a detailed explanation and examples.

The main `cglmm()` output returns the parameter estimates for each model specified (conditional, zero-inflated, and/or dispersion). The `summary()` function returns the parameter estimates, standard errors and 95% confidence intervals of the fitted model (Equation 2) but also applies the delta method to provide the same summary of recovered parameters (Equation 1), including the estimates for amplitude and acrophase.

It is often valuable to test for differences between groups or components within a model. For this purpose, the `test_cosinor_components()` and `test_cosinor_levels()` functions are used to assess differences between groups for amplitude or phase. In a multicomponent cosinor model, the `test_cosinor_components()` function can be used to estimate and test for differences between the different components in the model. For all cosinor models with grouped cosinor components, `test_cosinor_levels()` can be used to assess differences (for a single component) between levels of the grouping variable. These functions test for differences in acrophase and amplitude. As the MESOR is estimated in the model by the intercept term, differences between groups can be estimated by including a fixed effect for that grouping variable in the model formula and interpreting these effects in the model summary.

As GLMMcosinor fits a model using the glmmTMB package, all the same methods for model diagnostics can be used on a model fit using `cglmm()`. An example is included in the "Model visualizations" vignette using the DHARMa R package to produce a QQ residuals and residual versus predicted plots (Hartig, 2022).

2.3 Visualizations

GLMMcosinor includes two methods to visualize the model. The `polar_plot()` method visualizes the (one or more) cosinor components on a polar plot. On a polar plot, the angle from the pole to the outer boundary represents the phase angle of the cosinor component (the full circle representing the period), and the distance from the pole represents the amplitude. This allows, for example, easy visual comparison of only the cosinor portion of the model between two groups.

The `autoplot()` method returns a time series plot from the model. It can include the influence of covariates, including the grouping variables on the cosinor components, as well as superimposing the data used to fit the model or ribbon representing the confidence intervals. For more details, see the model visualizations vignette.

2.4 Example analysis

To demonstrate the importance of choosing an appropriate modeling approach when analyzing circadian rhythms, we used GLMMcosinor to fit a cosinor model to data that were originally modeled using circacompare, an R package that exclusively supports cosinor modeling of Gaussian response data (Parsons, et al., 2020), similar to most other cosinor modeling packages.

In figure 3A of Xin and colleagues' study, the circacompare was used to model relative gene expression over time (Xin, et al., 2021). In this figure, relative gene expression represents the ratio of Nr1d1 gene expression in the liver between two groups of mice: those fed during daytime (DRF - daytime restricted feeding), and nighttime (NRF - nighttime restricted feeding).

In this case, the problem with assuming a Gaussian response variable (standard linear regression modeling) is that it can estimate periods of the gene expression rhythm to be below zero, which is impossible. The model fit presented in this study exhibited this issue, with the NRF group having an (incorrectly) estimated mean gene expression level below zero between approximately zeitgeber time 16 and 24. We anticipated that it may be more appropriate for this model to be fit assuming a Gamma distribution for the gene expression, restricting the response variable to only ever be positive. To our knowledge, there is no other R or Python package available for fitting the cosinor model using a Gamma distribution and this problem would have occurred using any available software at the time of their publication.

We used the data presented in Figure 3A of Xin and colleagues' paper to models with GLMMcosinor: (1) with a Gaussian family (equivalent to the circacompare model), and (2) with a Gamma family. We also fit a third model which includes a random effect for the mouse, as this dataset included repeated measures and the within-mouse correlation is not otherwise accounted for by these first two models. Figure 1 visualizes these three models using `autoplot()`.

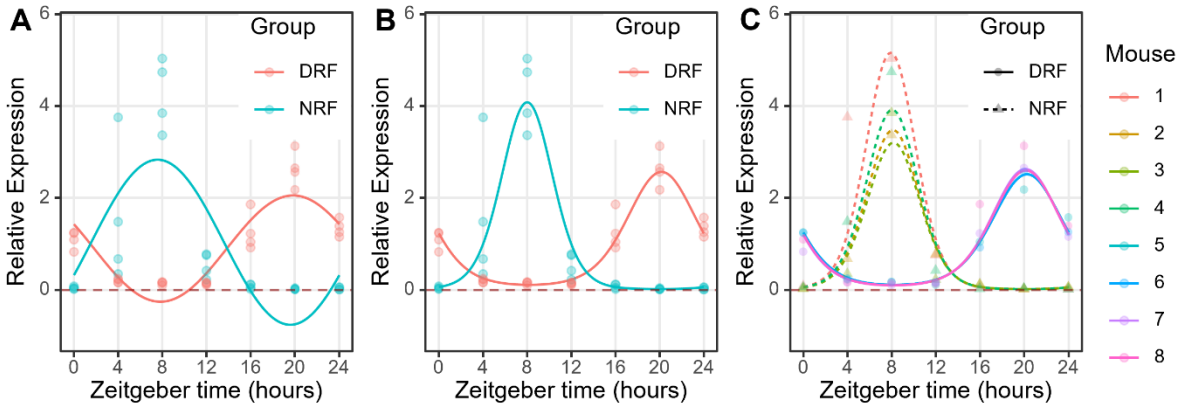

Figure 1: (A) Gaussian cosinor model of relative expression of Nr1d1 gene across. (B) Gamma cosinor model of relative expression of Nr1d1 gene expression. (C) Mixed, Gamma cosinor model of relative expression of Nr1d1 gene expression with random effects on each mouse.

Figure 1 suggests that the Gamma cosinor model fits the data better compared to the Gaussian model. It also prevents the possible misunderstanding that the gene expression levels for the NRF can be negative. Visually, the model fit is much better when using the Gamma model (Figure 1B and 1C) than when assuming the Gaussian response variable (Figure 1A). When fitting the model with a hierarchical structure, not only does it better account for the repeated-measures nature of the experimental design and dataset, but it can also allow the user to generate estimates of the rhythmic characteristics and generate predictions on a per-mouse basis. However, we note that the mixed model approach should be used regardless of this benefit, as we know that the data are hierarchical and specifying a model that ignores this within-mouse correlation would not be ideal.

3. Conclusion

GLMMcosinor is an R package designed to flexibly fit models to rhythmic data and to equip users with tools to effectively interpret and use their models. Existing cosinor regression R software lacks flexible support for modeling both non-Gaussian and hierarchical data. By using a generalized linear mixed modeling framework, GLMMcosinor supports modeling of a wider range of experimental and observation datasets. GLMMcosinor also includes features to generate visualizations, summaries, and predictions from fitted models, allowing researchers to efficiently interpret and use their cosinor models. More details on using the R package are available in the online documentation (https://docs.ropensci.org/GLMMcosinor/).

4. References

Bates, D.*, et al.* Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 2015;67(1):1 - 48.

Blum, I.D.*, et al.* A highly tunable dopaminergic oscillator generates ultradian rhythms of behavioral arousal. *eLife* 2014;3:e05105.

Cunningham, C.B.*, et al.* Competitive Ability in Male House Mice (Mus musculus): Genetic Influences. *Behavior Genetics* 2013;43(2):151-160.

Dobson, A.J. and Barnett, A.G. An Introduction to Generalized Linear Models. CRC press; 2018.

Doyle, M.M.*, et al.* Enhancing cosinor analysis of circadian phase markers using the gamma distribution. *Sleep Medicine* 2022;92:1-3.

Gandia, K.M.*, et al.* Understanding Circadian and Circannual Behavioral Cycles of Captive Giant Pandas (Ailuropoda melanoleuca) Can Help to Promote Good Welfare. *Animals* 2023;13(15):2401.

Hartig, F. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. *CRAN* 2022.

Hou, R., Tomalin, L.E. and Suárez-Fariñas, M. cosinoRmixedeffects: an R package for mixed-effects cosinor models. *BMC Bioinformatics* 2021;22(1):553.

Hughes, M.E., Hogenesch, J.B. and Kornacker, K. JTK_CYCLE: An Efficient Nonparametric Algorithm for Detecting Rhythmic Components in Genome-Scale Data Sets. *Journal of Biological Rhythms* 2010;25(5):372-380.

Hutchison, A.L.*, et al.* Improved Statistical Methods Enable Greater Sensitivity in Rhythm Detection for Genome-Wide Data. *PLOS Computational Biology* 2015;11(3):e1004094.

Itriyeva, K. The normal menstrual cycle. *Current Problems in Pediatric and Adolescent Health Care* 2022;52(5):101183.

Kalafatakis, K.*, et al.* Glucocorticoid ultradian rhythmicity differentially regulates mood and resting state networks in the human brain: A randomised controlled clinical trial. *Psychoneuroendocrinology* 2021;124:105096.

Mikulich, S.K.*, et al.* Comparing linear and nonlinear mixed model approaches to cosinor analysis. *Statistics in Medicine* 2003;22(20):3195-3211.

Moškon, M. CosinorPy: a python package for cosinor-based rhythmometry. *BMC Bioinformatics* 2020;21(1):485.

Obodo, D., Outland, E.H. and Hughey, J.J. LimoRhyde2: Genomic analysis of biological rhythms based on effect sizes. *PLOS ONE* 2023;18(12):e0292089.

Parsons, R.*, et al.* CircaCompare: a method to estimate and statistically support differences in mesor, amplitude and phase, between circadian rhythms. *Bioinformatics* 2020;36(4):1208-1212.

Sachs, M. cosinor: Tools for Estimating and Predicting the Cosinor Model. *CRAN* 2023.

Velikajne, N. and Moškon, M. RhythmCount: A Python package to analyse the rhythmicity in count data. *Journal of Computational Science* 2022;63:101758.

Xin, H.*, et al.* A multi-tissue multi-omics analysis reveals distinct kinetics in entrainment of diurnal transcriptomes by inverted feeding. *iScience* 2021;24(4):102335.

Zhu, X.*, et al.* In vivo flow cytometry reveals a circadian rhythm of circulating tumor cells. *Light: Science & Applications* 2021;10(1):110.