fractalRegression: An R package for multiscale regression and fractal analyses

Aaron D. Likens[1] & Travis J. Wiltshire[2]

[1] Department of Biomechanics, University of Nebraska at Omaha

[2] Department of Cognitive Science & Artificial Intelligence, Tilburg University

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line. Enter author note here.

The authors made the following contributions. Aaron D. Likens: Conceptualization, Software, Writing - Original Draft Preparation, Writing - Review & Editing; Travis J. Wiltshire: Writing - Original Draft Preparation, Writing - Review & Editing, Software.

Correspondence concerning this article should be addressed to Aaron D. Likens, Postal address. E-mail: alikens@unomaha.edu

Abstract

Time series data from scientific fields as diverse as astrophysics, economics, human movement science, and neuroscience all exhibit fractal properties. That is, these time series often exhibit self-similarity and long-range correlations. This `fractalRegression` package implements a number of univariate and bivariate time series tools appropriate for analyzing noisy data exhibiting these properties. These methods, especially the bivariate tools (Kristoufek, 2015a; Likens, Amazeen, West, & Gibbons, 2019) have yet to be implemented in an open source and complete package for the R Statistical Software environment. As both practitioners and developers of these methods, we expect these tools will be of interest to a wide audience of scientists who use R, especially those from fields such as the human movement, cognitive, and other behavioral sciences. The algorithms have been developed in C++ using the popular Rcpp (Eddelbuettel & Francois, 2011) and RcppArmadillo (Eddelbuettel & Sanderson, 2014) packages. The result is a collection of efficient functions that perform well even on long time series (e.g., $\geq$ 10,000 data points). In this work, we motivate introduce the package, each of the functions, and give examples of their use as well as issues to consider to correctly use these methods.

Keywords: long range correlation, fractal, multiscale, dynamics

Word count: X

fractalRegression: An R package for multiscale regression and fractal analyses

Introduction

Fractal analysis, in its many forms, has become an important framework in virtually every area of science, often serving as an indicator of system health (Goldberger et al., 2002), adaptability (Bak, Tang, & Wiesenfeld, 1987), control (Likens, Fine, Amazeen, & Amazeen, 2015), cognitive function (Euler, Wiltshire, Niermeyer, & Butner, 2016), and multi-scale interactions (Kelty-Stephen, 2017). In particular, various methods related to Detrended Fluctuation Analysis (DFA) (Peng et al., 1994) have rose to prominence due to their ease of understanding and broad applicability to stationary and nonstationary time series, alike.

The basic DFA algorithm has been implemented in numerous packages and software programs. However, advanced methods such as Multifractal Detrended Fluctuation Analysis (MFDFA) (Kantelhardt et al., 2002), Detrended Cross Correlation (DCCA) (Podobnik & Stanley, 2008; Zebende, 2011), and, in particular, fractal regression techniques such as Multiscale Regression Analysis (MRA) (Kristoufek, 2015a; Likens, Amazeen, West, & Gibbons, 2019) have not yet been implemented in a comprehensive CRAN Package for the R Statistical Software Environment. Thus, there is a clear need for a package that incorporates this functionality in order to advance theoretical research focused on understanding the time varying properties of natural phenomena and applied research that uses those insights in important areas such as healthcare (Cavanaugh, Kelty-Stephen, & Stergiou, 2017) and education (Snow, Likens, Allen, & McNamara, 2016).

Package Overview

Some foundational efforts in fractal analyses, which partially overlap with the functionality of this package, have been implemented elsewhere. For example, a number of univariate fractal and multifractal analyses have been implemented in the 'fracLab' library for MATLAB (Legrand & Véhel, 2003) and other toolboxes that are mainly targeted at multifractal analysis (Ihlen, 2012; Ihlen & Vereijken, 2010). In terms of open access packages, there are other packages that implement some, but not all of the same functions such as the `fathon` package (Bianchi, 2020) that has been implemented in Python as well as the R packages: `fractal` [defunct], `nonlinearTseries` (Garcia, 2020), and `MFDFA` (Laib, Golay, Telesca, & Kanevski, 2018). However, none of the above packages incorporate univariate monofractal and multifractal DFA with bivariate DCCA and MRA nor do they run on a C++ architecture. Our `fractalRegression` package is unique in this combination of analyses and efficiency. For instance, we are not aware of any other packages that feature MRA and Multiscale Lagged Regression (MLRA).

Methodological Details and Examples

In order to demonstrate the methods within the 'fractalRegression' package, we group this into univariate (DFA, MFDFA) and bivariate methods (DCCA, MRA, MRLA). For each method, we 1) highlight the key question(s) that can be answered with that method, 2) briefly describe the algorithm with sources for additional details, 3) describe some key consideration for appropriately applying the algorithm, and demonstrate the use of the functions on a 4) simulated and 5) empirical application of the function. An overview of the functions included in the package, the general objective of that function, and the output are shown below in Table 1. The additional details are included in the sections corresponding to those methods, in the package documentation, and in the original sources for the methods.

Table 1.

Overview of package functions, objectives, and output

Univariate Methods

Detrended Fluctuation Analysis. The key question that can be answered by Detrended Fluctuation Analysis (DFA) (Peng et al., 1994) is: *what is the magnitude and direction of long range correlation in a single time series?* While DFA has been described extensively elsewhere (Kantelhardt, Koscielny-Bunde, Rego, Havlin, & Bunde, 2001) and visualized nicely (Kelty-Stephen, Stirling, & Lipsitz, 2016), we provide a brief summary here. DFA entails splitting a time series into several small bins (e.g., 16). In each bin, the least squares regression is fit and subtracted within each window. Residuals are squared and averaged within each window. Then, the square root is taken of the average squared residual across all windows of a given size. This process repeats for larger window sizes, growing by, say a power of 2, up to N/4, where N is the length of the series. In a final step, the logarithm of those scaled root mean squared residuals (i.e., fluctuations) is regressed on the logarithm of window sizes. The slope of this line is termed α and it provides a measure of the long range correlation. α is commonly used an as estimator of the Hurst exponent (H), where $\alpha < 1 = H$, and for $\alpha > 1$, $H = 1 - \alpha$. Conventional interpretation of α is: $\alpha < 0.5$ is anti-correlated, $\alpha = 0.5$ is uncorrelated, white noise, $\alpha > 0.5$ is temporally correlated, $\alpha = 1$ is long-range correlated, 1/f-noise, pink noise, $\alpha > 1$ is non-stationary and unbounded, and $\alpha = 1.5$ is fractional brownian motion.

DFA Examples.

To demonstrate the use of `dfa()` we simulate three time series using the `fgn_sim()` function. This is a simple function based on the former `fARMA` R package. It only requires the number of observations `n`, and the Hurst exponent `H`. In particular, we simulate white noise, pink noise, and anti-correlated fractional Gaussian noise using the code below.

```
white.noise <- rnorm(5000)
pink.noise <- fgn_sim(n = 5000, H = 0.9)
anti.corr.noise <- fgn_sim(5000, H = 0.25)
```

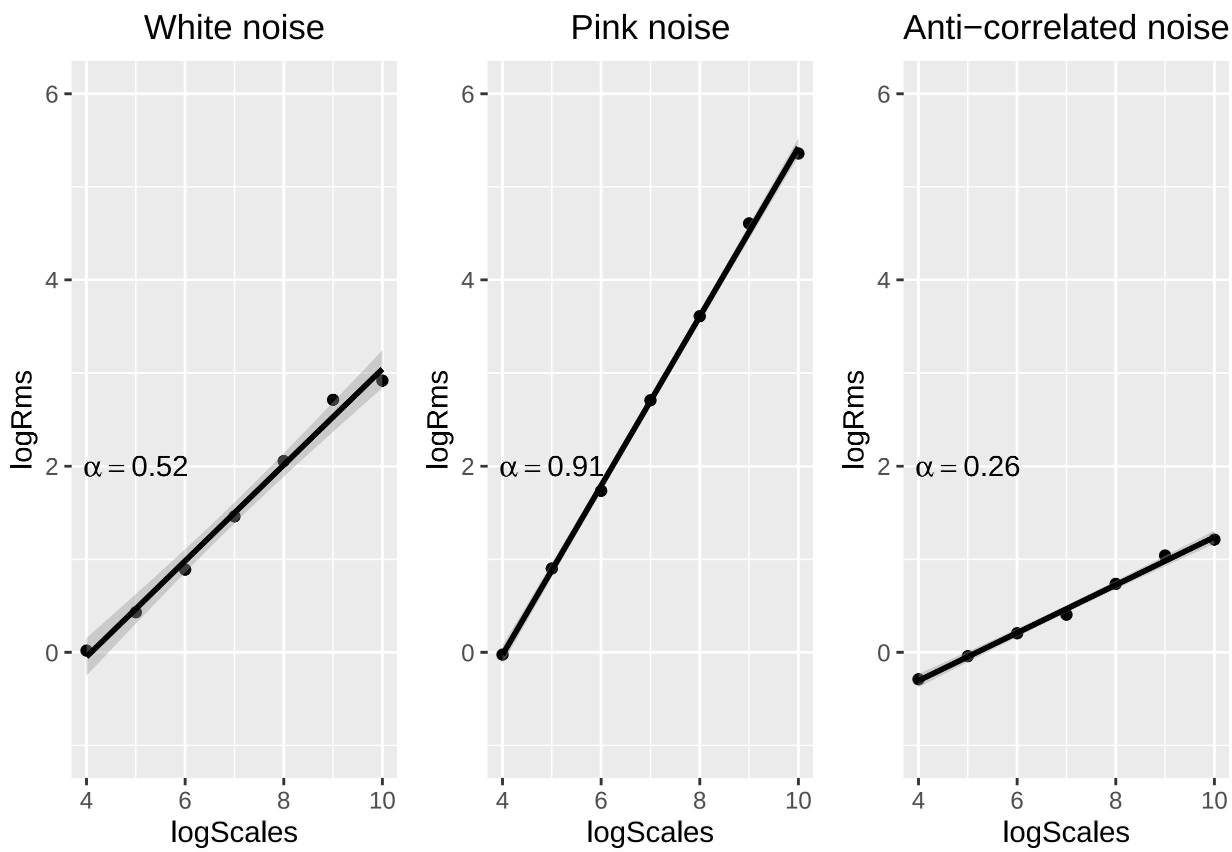
Then we can run DFA on these using the example code below. Note that this example uses a linear detrending with minimum scale (`sc_min`) of 16, a maximum scale (`sc_max`) of 1/4 the time series length, and logarithmically spaced scale factor (`scale_ratio`) of 2.

```
dfa.white <- dfa(x = white.noise, order = 1, verbose = 1,
                 sc_min = 16, sc_max = length(pink.noise)/4, scale_ratio = 2)
dfa.pink <- dfa(x = pink.noise, order = 1, verbose = 1,
                 sc_min = 16, sc_max = length(pink.noise)/4, scale_ratio = 2)
dfa.anti.corr <- dfa(x = anti.corr.noise, order = 1, verbose = 1,
                 sc_min = 16, sc_max = length(pink.noise)/4, scale_ratio = 2)
```

In terms of output from the above examples, for white noise, we observed that $\alpha = 0.52$, for pink noise we observed that $\alpha = 0.91$, and since we simulated anti-correlated noise with H $= 0.25$, we observed a close estimate of the $\alpha = 0.26$. In terms of the objects saved from the `dfa()` function, one commonly inspects the `logScales-logRms` plots. Given the estimates above, we see in Figure 1 that the slopes for white noise, pink noise, and anti-correlated noise conform to our expectations.

Figure 1

LogScale-LogFluctuation plots for white noise (left), pink noise (middle), and anti-correlated noise (right)



For an empirical example, we apply the `dfa()` function to DESCRIBE DATA HERE. Movement Data?

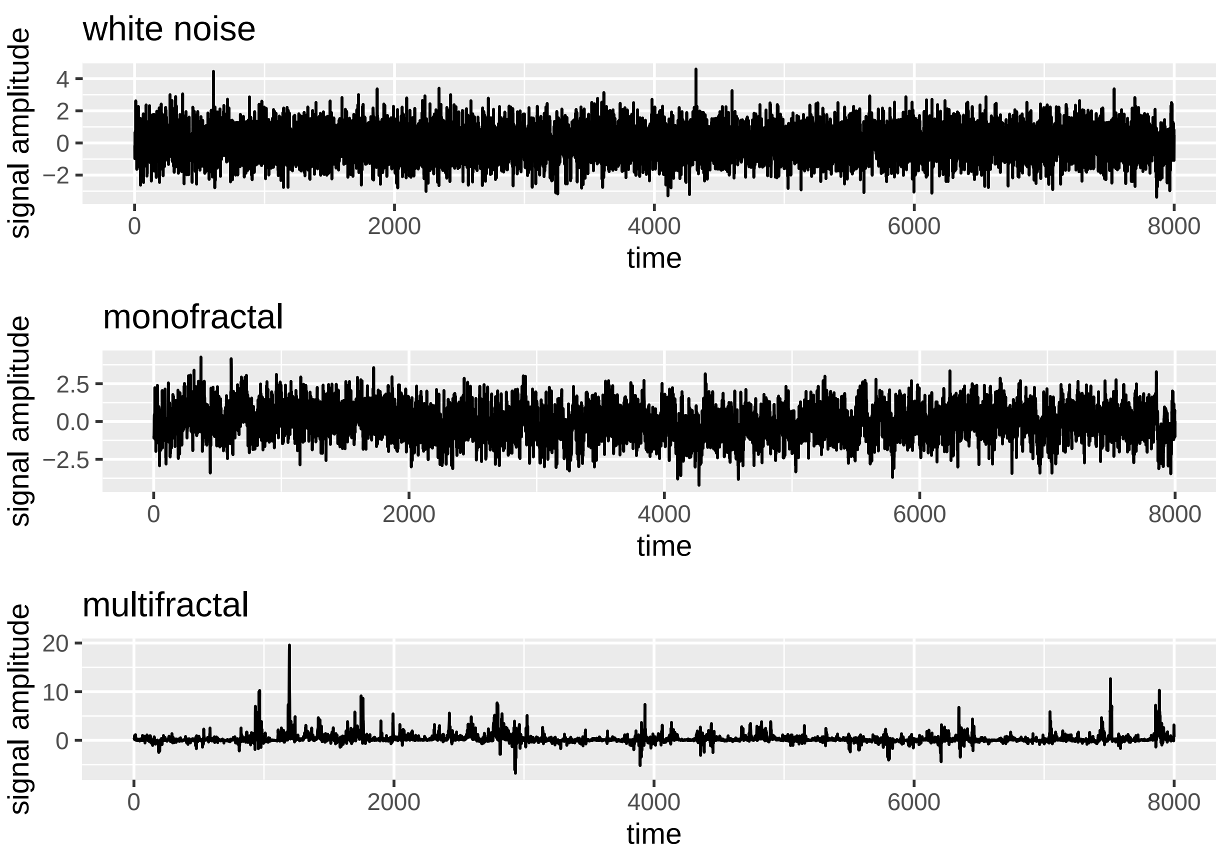
Figure 2

LogScale-LogFluctuation plots for empirical time series (ADD ME)

Multifractal Detrended Fluctuation Analysis. Multifractal Detrended Fluctuation Analysis (MFDFA; Kantelhardt et al. (2002)) is an extension of DFA by generalizing the fluctuation function to a range of exponents of the qth order. The key question that can be answered by MFDFA is: *how does the magnitude and direction of long range correlation change over time within a single time series?* Like DFA, MFDFA entails splitting a time series into several small bins (e.g., 16). In each bin, the least squares regression is fit and subtracted within each window. However, the residuals are raised to a range of exponents q and averaged within each window. So when $q = 2$, DFA is equal to

MFDFA. When $q > 2$, larger residual are emphasized and when $q < 2$, smaller residuals are emphasized. The rest of the DFA algorithm is performed for each window and windows size for all values of q. We refer the reader to the work of Kelty-Stephens and colleagues Kelty-Stephen, Stirling, and Lipsitz (2016) Figure 3 for a visualization of the algorithm and to Kantelhardt and colleagues Kantelhardt et al. (2002) for additional mathematical description.

MFDFA Examples. To demonstrate the use of `mfdfa()`, we work with data included in our package (`fractaldata`), that was originally provided by Ihlen (2012) . It includes a white noise time series, monofractal time series, and a multifractal time series. Investigation of the properties could also examine simulated multifractal Brownian motion and multifractal Gaussian noise.



Simulated data: mfbrownian motion from Ihlen matlab (Aaron might have R port) see mbm_mgn for R from aaron - Empirical data: EPICLE Movement Data?

DFA and MFDFA Considerations.

We recommend a few points of consideration here in using this function. One is to be sure to verify there are not cross-over points in the logScale-logFluctuation plots (Peng et al., 1994; Perakakis, Taylor, Martinez-Nieto, Revithi, & Vila, 2009). Cross-over points (or a visible change in the slope as a function of of scale) indicate that a mono-fractal characterization does not sufficiently characterize the data. If cross-over points are evident, we recommend proceeding to using the 'mfdfa()' to estimate the multi-fractal fluctuation dynamics across scales.

While it is common to use only linear detrending with DFA, it is important to inspect the trends in the data to determine if it would be more appropriate to use a higher order polynomial for detrending, and/or compare the DFA output for different polynomial orders (Kantelhardt, Koscielny-Bunde, Rego, Havlin, & Bunde, 2001).

General recommendations for choosing the min and max scale are an sc_min = 10 and sc_max = (N/4), where N is the number of observations. See Eke et al. (2002) (Eke, Herman, Kocsis, & Kozak, 2002) and Gulich and Zunino (2014) (Gulich & Zunino, 2014) for additional considerations.

Bivariate Methods

Detrended Cross-Correlation Analysis. Detrended Cross-Correlation Analysis (DCCA; Podobnik and Stanley (2008)) is a bivariate extension of the DFA algorithm generalizing it to a correlational case between two time series. The key questions that can be asked it are: a) *How does correlation between two time series change as a function of scale?* and b) *What is/are the dominant (time) scale(s) of coordination? (those that are beyond a threshold, or statistically significant given a criteria, or of a certain magnitude?* For DCCA, the DFA algorithm gets applied to both time series providing the scale-wise estimates for both. DESCRIBE DCCA ALGORITHM HERE.

DCCA Examples.

- - Simulated data: MC-ARFIMA
  - Empirical data: EPICLE Movement Data?

Multi-scale Regression Analysis. Multi-scale regression analysis (MRA) is an adaptation of DCCA that brings the analyses into a predictive, regression framework Kristoufek (2015b) . The key questions that can be answered by it are: a) *How does the influence of one time series on another time series change as a function of scale?* and b) *What is/are the dominant (time) scale(s) of influence of one time series on another time series?* DESCRIBE MRA ALGORITHM HERE.

MRA Examples.

- MRA
  - Simulated data:
  - Empirical data: FNIRS from Aaron?

Multi-scale Lagged Regression Analysis. Multi-scale lagged regression analysis is an extension of MRA that allows for examining the influence as a function of scale, but also of time lag. In paritucIar, the key questions that can be asked with MLRA are: a) *How does the influence of one time series on another time series change as a function of scale at different time lags?* and b) *Does the dominant time scale of influence change over successive time lags?* DESCRIBE MLRA ALGORITHM HERE.

MLRA Examples.

- MLRA
  - Key Question
  - Simulated data: Equation from Aaron from grant on MLRA
  - Empirical data: FNIRS from Aaron?

Surrogate Methods (and 'full' data analysis)

Methods are ranked in terms of increasing levels of rigor.

- Randomization - Estimates should be different.
- IAAFT - Estimates should be different.
- Model based surrogate (Simulated exponents) - See Likens 2019 paper with model of postural sway/control, taking an educated guess about the data generating process underlying the time series. Estimates should not be different. See Roume et al 2018 windowed detrended CCA
- Can we incorporate lags into MC-ARFIMA?

General Discussion

- General value of methods and the types of questions
- Practical consideration of univariate methods
- Practical consideration of bivariate methods
- Unique contribution of the methods

Acknowledgements

Author AL receives support from a National Institutes of Health Center grant (P20GM109090).

References

Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of the 1/f noise. *Physical Review Letters*, *59*(4), 381–384. https://doi.org/10.1103/PhysRevLett.59.381

Bianchi, S. (2020). Fathon: A python package for a fast computation of detrendend fluctuation analysis and related algorithms. *Journal of Open Source Software*, *5*(45), 1828.

Cavanaugh, J. T., Kelty-Stephen, D. G., & Stergiou, N. (2017). Multifractality, Interactivity, and the Adaptive Capacity of the Human Movement System: A Perspective for Advancing the Conceptual Basis of Neurologic Physical Therapy. Retrieved from https://www.ingentaconnect.com/content/wk/npt/2017/00000041/00000004/art00007

Eddelbuettel, D., & Francois, R. (2011). Rcpp: Seamless R and C++ Integration. *Journal of Statistical Software*, *40*(1), 1–18. https://doi.org/10.18637/jss.v040.i08

Eddelbuettel, D., & Sanderson, C. (2014). RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics & Data Analysis*, *71*, 1054–1063. https://doi.org/10.1016/j.csda.2013.02.005

Eke, A., Herman, P., Kocsis, L., & Kozak, L. R. (2002). Fractal characterization of complexity in temporal physiological signals. *Physiological Measurement*, *23*(1), R1–R38. https://doi.org/10.1088/0967-3334/23/1/201

Euler, M. J., Wiltshire, T. J., Niermeyer, M. A., & Butner, J. E. (2016). Working memory performance inversely predicts spontaneous delta and theta-band scaling relations. *Brain Research*, *1637*, 22–33. https://doi.org/10.1016/j.brainres.2016.02.008

Garcia, C. A. (2020). nonlinearTseries: Nonlinear Time Series Analysis. Retrieved from https://CRAN.R-project.org/package=nonlinearTseries

Goldberger, A. L., Amaral, L. A. N., Hausdorff, J. M., Ivanov, P. C., Peng, C.-K., & Stanley, H. E. (2002). Fractal dynamics in physiology: Alterations with disease and aging. *Proceedings of the National Academy of Sciences*, *99*(suppl 1), 2466–2472. https://doi.org/10.1073/pnas.012579499

Gulich, D., & Zunino, L. (2014). A criterion for the determination of optimal scaling ranges in DFA and MF-DFA. *Physica A: Statistical Mechanics and Its Applications*, *397*, 17–30. https://doi.org/10.1016/j.physa.2013.11.029

Ihlen, E. A. F. (2012). Introduction to Multifractal Detrended Fluctuation Analysis in Matlab. *Frontiers in Physiology*, *3*. https://doi.org/10.3389/fphys.2012.00141

Ihlen, E. A. F., & Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: Beyond 1/? fluctuation. *Journal of Experimental Psychology: General*, *139*(3), 436–463. https://doi.org/10.1037/a0019098

Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H. A., Havlin, S., & Bunde, A. (2001). Detecting long-range correlations with detrended fluctuation analysis. *Physica A: Statistical Mechanics and Its Applications*, *295*(3), 441–454. https://doi.org/10.1016/S0378-4371(01)00144-3

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and Its Applications*, *316*(1), 87–114. https://doi.org/10.1016/S0378-4371(02)01383-3

Kelty-Stephen, D. G. (2017). Threading a multifractal social psychology through within-organism coordination to within-group interactions: A tale of coordination in three acts. *Chaos, Solitons & Fractals*, *104*, 363–370. https://doi.org/10.1016/j.chaos.2017.08.037

Kelty-Stephen, D. G., Stirling, L. A., & Lipsitz, L. A. (2016). Multifractal temporal correlations in circle-tracing behaviors are associated with the executive function of rule-switching assessed by the Trail Making Test. *Psychological Assessment*, *28*(2), 171–180. https://doi.org/10.1037/pas0000177

Kristoufek, L. (2015a). Detrended fluctuation analysis as a regression framework: Estimating dependence at different scales. *Physical Review E*, *91*(2), 022802. https://doi.org/10.1103/PhysRevE.91.022802

Kristoufek, L. (2015b). Detrended fluctuation analysis as a regression framework: Estimating dependence at different scales. *Physical Review E*, *91*(2), 022802. https://doi.org/10.1103/PhysRevE.91.022802

Laib, M., Golay, J., Telesca, L., & Kanevski, M. (2018). Multifractal analysis of the time series of daily means of wind speed in complex regions. *Chaos, Solitons & Fractals*, *109*, 118–127. https://doi.org/10.1016/j.chaos.2018.02.024

Legrand, P., & Véhel, J. L. (2003). Signal and image processing with FracLab. *Thinking in Patterns: Fractals and Related Phenomena in Nature*, 321322.

Likens, A. D., Amazeen, P. G., West, S. G., & Gibbons, C. T. (2019). Statistical properties of Multiscale Regression Analysis: Simulation and application to human postural control. *Physica A: Statistical Mechanics and Its Applications*, *532*, 121580. https://doi.org/10.1016/j.physa.2019.121580

Likens, A. D., Fine, J. M., Amazeen, E. L., & Amazeen, P. G. (2015). Experimental control of scaling behavior: What is not fractal? *Experimental Brain Research*, *233*(10), 2813–2821. https://doi.org/10.1007/s00221-015-4351-4

Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, *49*(2), 1685–1689. https://doi.org/10.1103/PhysRevE.49.1685

Perakakis, P., Taylor, M., Martinez-Nieto, E., Revithi, I., & Vila, J. (2009). Breathing frequency bias in fractal analysis of heart rate variability. *Biological Psychology*, *82*(1), 82–88. https://doi.org/10.1016/j.biopsycho.2009.06.004

Podobnik, B., & Stanley, H. E. (2008). Detrended Cross-Correlation Analysis: A New Method for Analyzing Two Nonstationary Time Series. *Physical Review Letters*, *100*(8), 084102. https://doi.org/10.1103/PhysRevLett.100.084102

Snow, E. L., Likens, A. D., Allen, L. K., & McNamara, D. S. (2016). Taking Control: Stealth Assessment of Deterministic Behaviors Within a Game-Based System. *International Journal of Artificial Intelligence in Education*, *26*(4), 1011–1032. https://doi.org/10.1007/s40593-015-0085-5

Zebende, G. F. (2011). DCCA cross-correlation coefficient: Quantifying level of cross-correlation. *Physica A: Statistical Mechanics and Its Applications*, *390*(4), 614–618. https://doi.org/10.1016/j.physa.2010.10.022

Fun ction	Objective	Output
dfa	Estimate long-range correlation in a time series	Object containing the overall α estimate and, if desired the `logScales` and `logRMS`
mfdfa	Estimate the magnitude and range of long-range correlations in a time series	Object containing the $\log_2$ scales used for the analysis, the $\log_2$ fluctuation function for each scale and q, the various q-order exponents, Hq, Tau, h, and Dh
dcca	Estimates of sc ale-specific correlation between two time-series	Object containing the scales used for the analysis and the ρ 'rho' values for each scale
mra	Estimates the scale specific regression coefficients for a predictor time series on and outcome time series	Object containing the scales and scale specific β estimates, R^2, and t statistics

Fun ction	Objective	Output
mlra	Estimates the scale specific regression coefficients for a predictor time series on and outcome time series at p re-specified lags	Object with lag-specific β coefficients
fg n_sim	Simulate fractional Gaussian noise	Returns a vector of length `n` according to the specified `H` Hurst exponent
iaaft	Generate surrogate series using the iterative amplitude adjusted Fourier transform	Returns a vector of same length as input time series