Resampling Methods, Lomb-Scargle Analysis, and Empirical Orthogonal Functions: A Combined Approach to Gappy Data in the Maritime Continent

Christopher Dupuis and Courtney Schumacher

Calculating EOFs with Gappy Data

Empirical Orthogonal Functions are eigenvector solutions of a covariance matrix. The covariance matrix elements can be defined as an evaluation of two time series' cross-correlation function:

$$C = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots \\ \sigma_{21} & \sigma_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

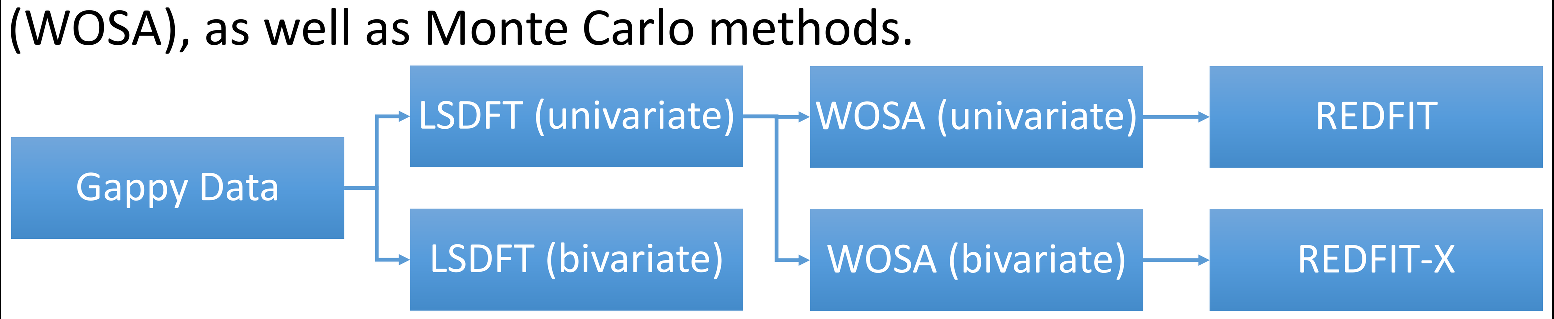
Prior research (in revision) shows it is possible to deduce EOF results from gappy geospatial time series by applying complex-valued Lomb-Scargle Discrete Fourier Transforms (LSDFTs) to each time series, then convolving each pair of spectra over a bandwidth.

Existing Techniques and Possible Workflows

The basic LSDFT workflow can be enhanced with a number of stackable techniques, satisfying the Cross-Correlation Theorem (bivariate workflow, right integral) or, preferably, the Wiener-Kinchin Theorem (univariate workflow, left integral):

$$\sigma_{xy} = \int_{-\infty}^{\infty} \tilde{X}(\omega)\tilde{Y}^*(\omega)\, d\omega = \int_{-\infty}^{\infty} \left|\widetilde{XY}(\omega)\right|^2 d\omega$$

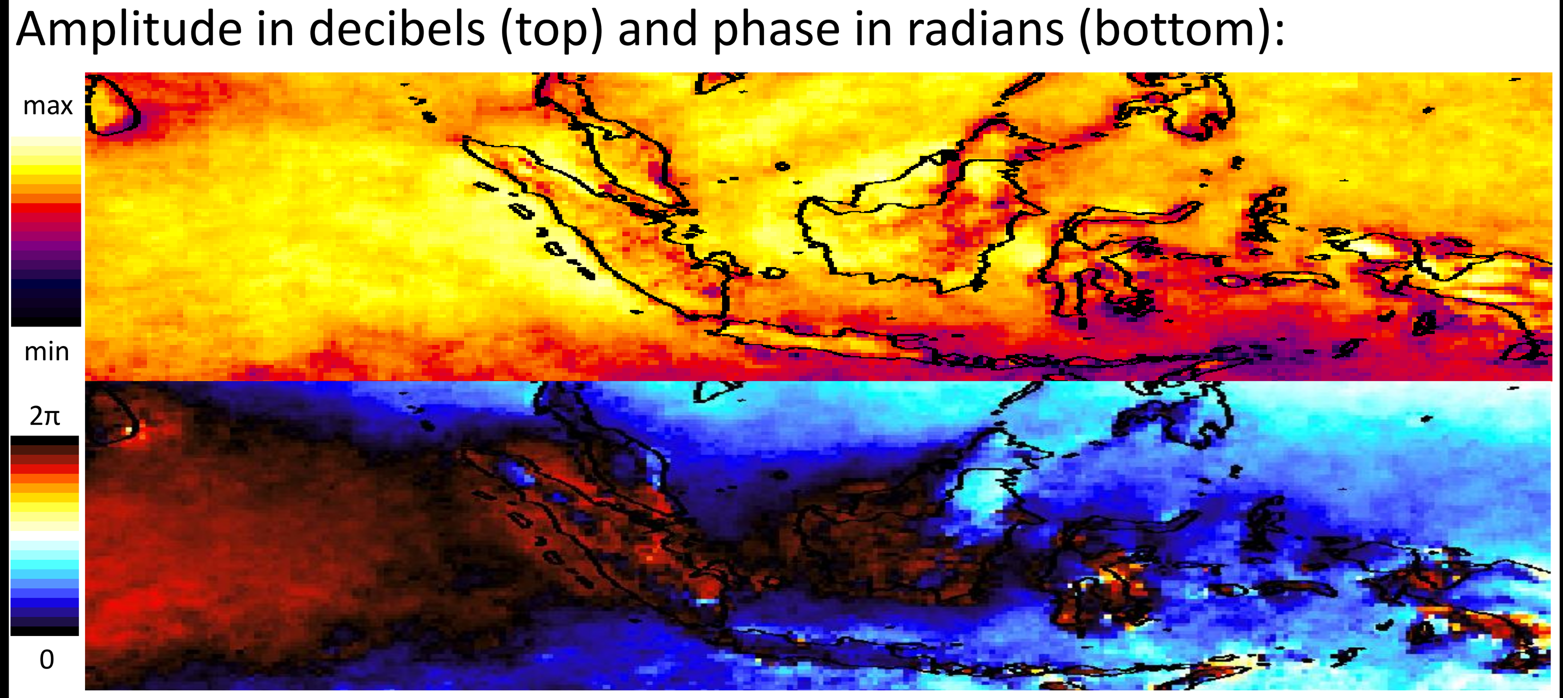
The basic LSDFT method can be enhanced with Welch's method (WOSA), as well as Monte Carlo methods.



Bivariate methods are time-consuming, scaling with the square of the expense for univariate methods, making them extremely expensive and therefore impractical for large data sets.

TRMM 3B42 MJO EOF (Raw WOSA)

Amplitude in decibels (top) and phase in radians (bottom):



The MJO mode appears as the second EOF in the 0-60 day band, accounting for about 75% of the non-diurnal variance. We note the relatively smooth eastward propagation shown in the phase map, as well as landmass-influenced structure in both the strength and timing of MJO rainfall effects.

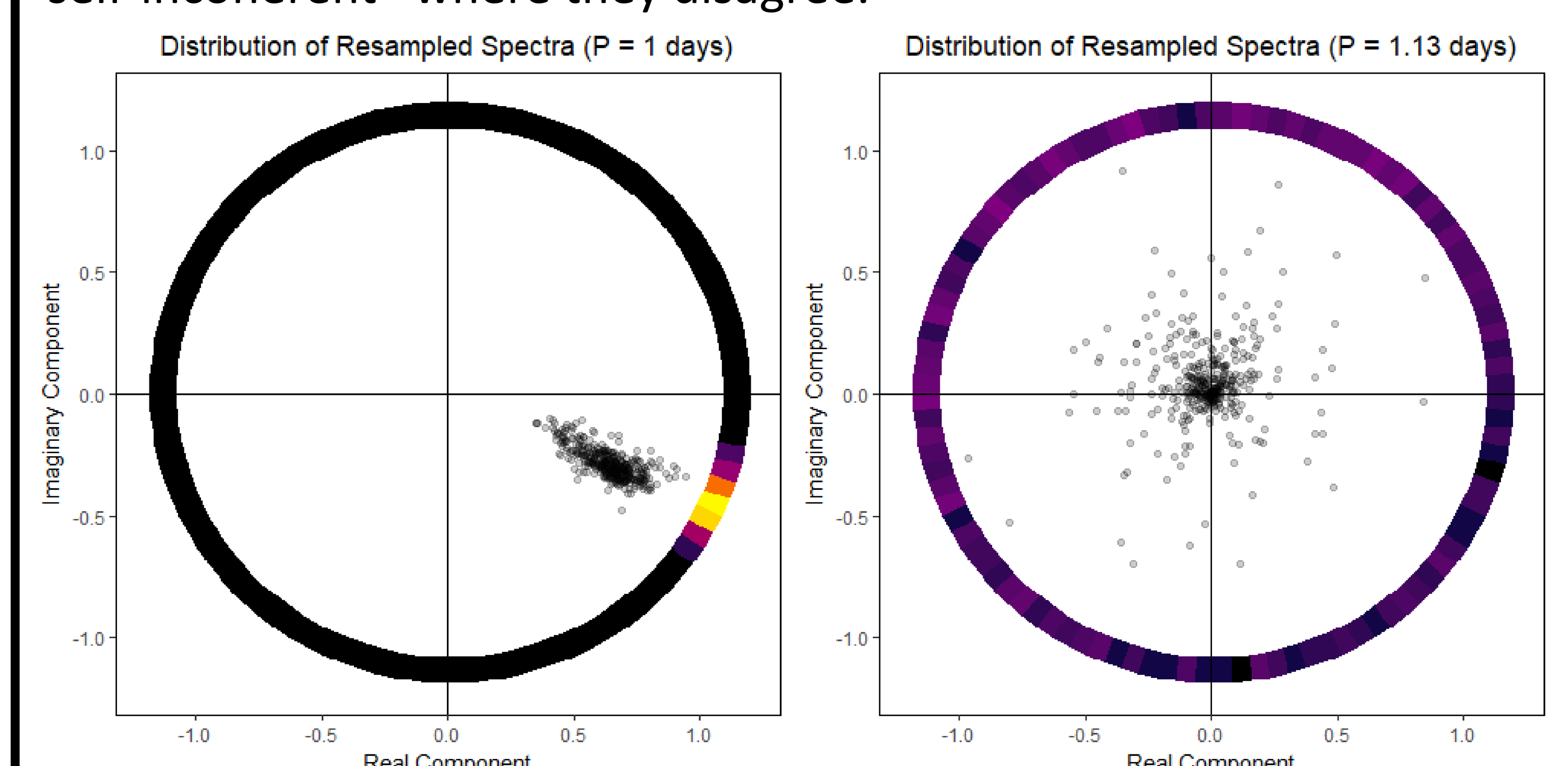
Resampling with Phase Preservation

Resampling methods like REDFIT and REDFIT-X provide a way to reduce the noise of spectral amplitude, but say nothing about phase, potentially even damaging the phase information.

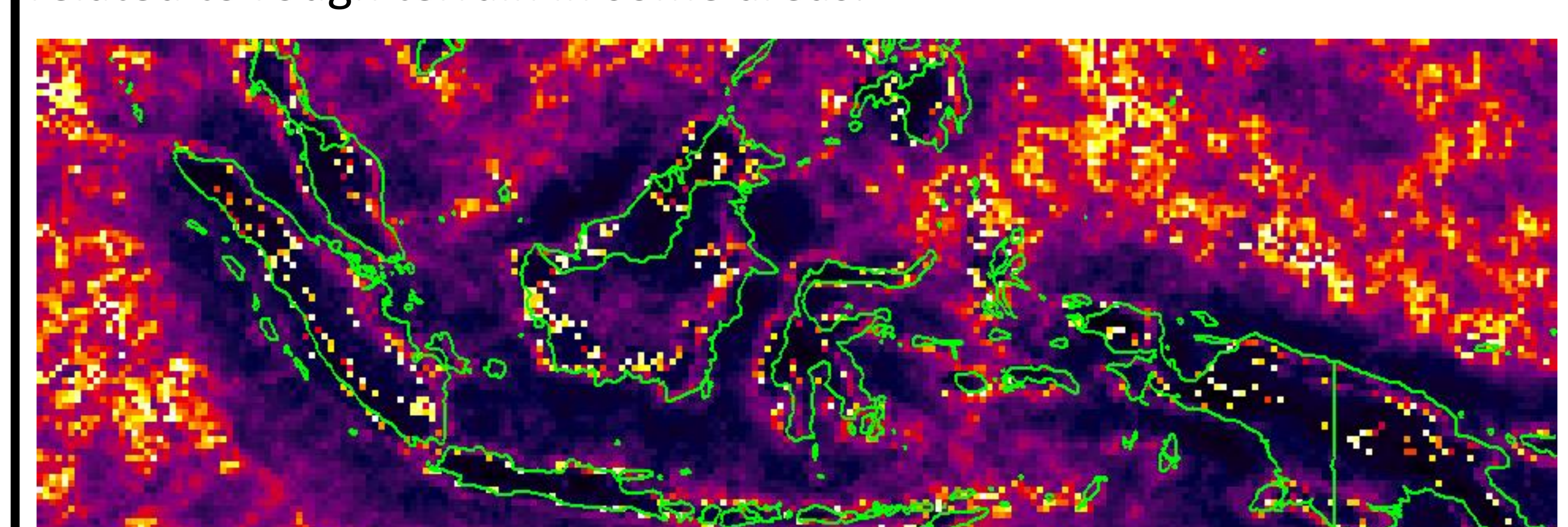
Therefore, we must consider phase-preserving methods. A simple way is to assign uniformly distributed values to each sample, and throw out any that fall below a threshold (akin to rolling a die). This is similar to the Delete-*d* Jackknife, except that *d* is binomially distributed rather than a constant.

Self-Coherence and Self-Incoherence

The spectra of time series resamples ostensibly result in a similar phase if there is significant phase information, and white noise (a circular uniform distribution) if not. We can call spectra "self-coherent" at frequencies where resamples agree on a phase, and "self-incoherent" where they disagree.



A heat map of the diurnal cycle circular standard deviation over the Maritime Continent is shown below. Self-incoherence in areas of high diurnal cycle amplitude appear to have spatial structure, possibly related to rough terrain in some areas.

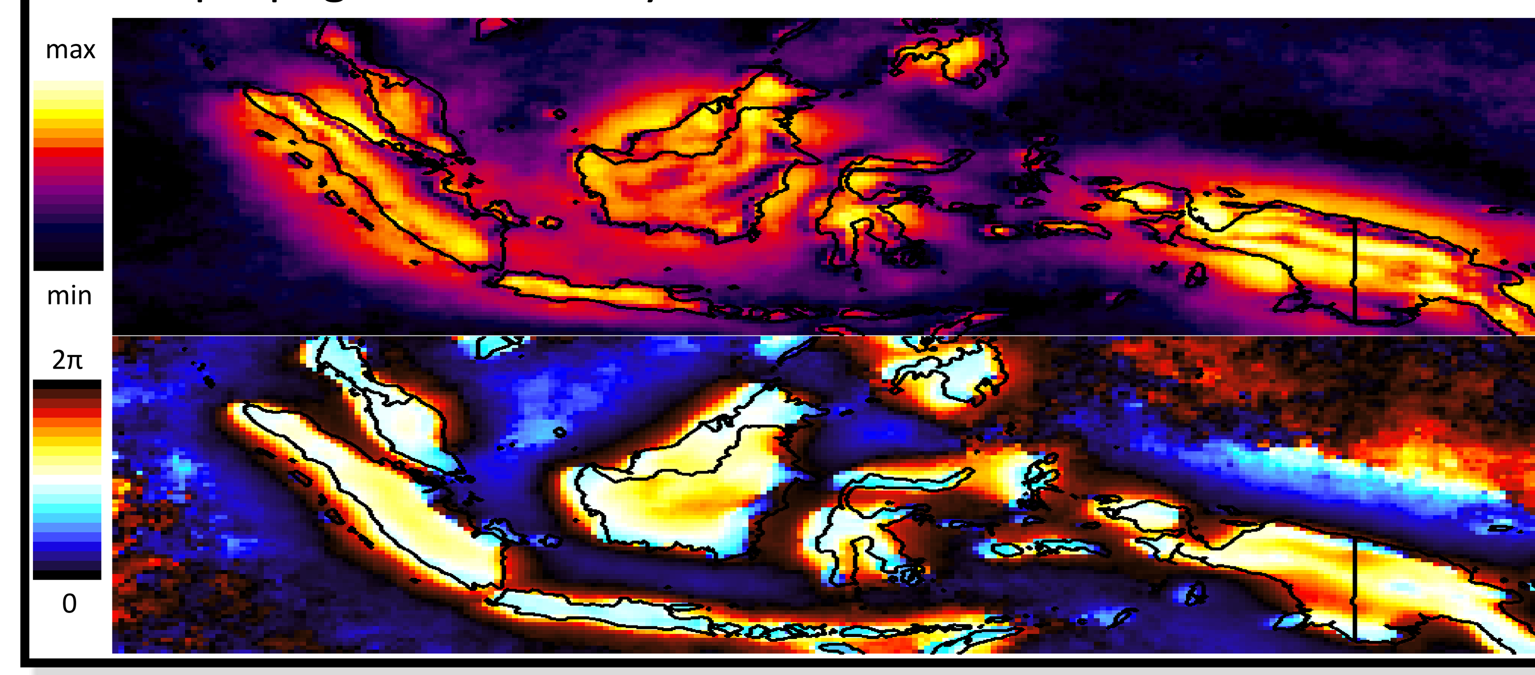


Further Reading:

Scargle (1982); North (1984); Schulz and Stattegger (1997); Schulz and Mudelsee (2002); Ólafsdóttir et al. (2016); Dupuis and Schumacher (in revision)

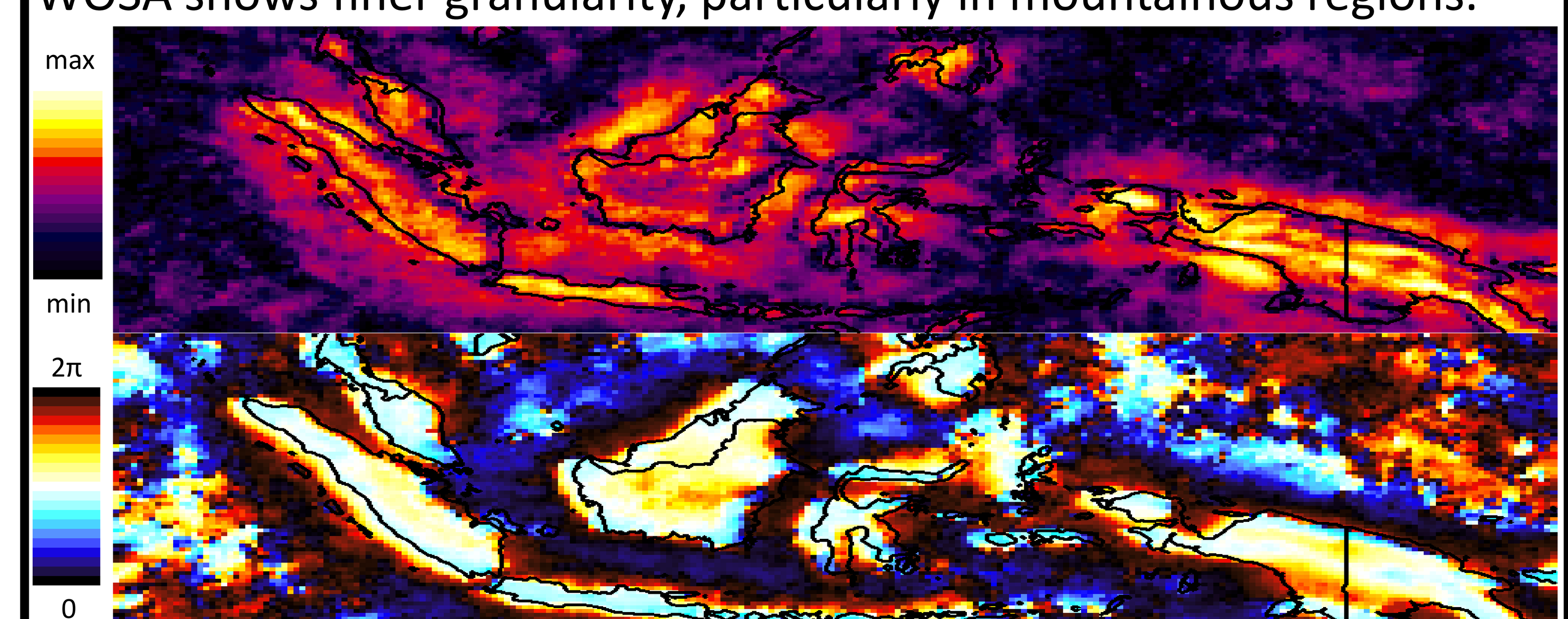
TRMM 3B42 Diurnal Cycle (Raw WOSA)

Amplitude in decibels (top) and phase in radians (bottom); the diurnal cycle is much stronger over land, and shows land breeze and sea breeze propagations in many areas.



TRMM 3B42 Diurnal Cycle (Delete ~10% Resampled WOSA)

Amplitude in decibels (top) and phase in radians (bottom); resampled WOSA shows finer granularity, particularly in mountainous regions.



Conclusions

Phase-preserving resampling methods are able to discern hidden features of geospatial time series. However, this method has mixed success in enhancing the spectra, and generally performs worse with EOF analysis, at least with the crude workflow presented. Circular averaging of resampled spectra as well as weighting the original spectra by circular standard deviation have not shown consistent success so far.

Questions Going Forward

- What is the best way to parameterize and weight self-coherence in time series resamples?
- What determines the spatial distribution of self-incoherence?
- Why does weighting by self-coherence succeed in reducing the noise floor in some regions, and fail in others?