About surrogates

July 16, 2019

For many test of significance, we need to build surrogates. For time-series, most of the time, we want to destroy the potential cross-correlation between time-series but to keep the autocorrelation within the time series to avoid artefacts. This can be done in the spectral domain. The simplest method is the so-called caterpillar procedure to these methods (Gouhier and Guichard, 2014), which only consist in shifting each time series by a different random lag. Among the more complicated methods used to build surrogates, we will present the Iterative Amplitude Adapted Fourier Transform method (IAAFT, Schreiber and Schmitz, 2000) and Ebisuzaki's method (Ebisuzaki, 1997).

IAAFT

We have access to the codes from Detto et al. (2013) and Kristian Haaga from the tstools R package `https://github.com/kahaaga/tstools`.

Some notes on the different algorithms I have seen:

- I am not sure why we should apply a Gaussian reordering: this is done in the AFFT in Schreiber and Schmitz (1996, 2000) but in both papers, the null hypothesis is that data were generated by a Gaussian linear process. For IAAFT, on the contrary, they remove this assumption which introduces some flatness (white noise) in the spectrum. Rank re-ordering based on data seems more appropriate[1].
- I am a bit at loss concerning the comparison between auto-correlations in the time series (Schreiber and Schmitz (2000), K. Haaga) and the use of cross-correlation between surrogate and observed time series by M. Detto.

[1] We can note that in K. Haaga's code, even though a Gaussian re-ordering is applied to the first iteration on surrogate, it seems quickly removed during the loop process [from what I've understood].

Algorithm 1 Bases for the IAAFT sequence with variations from M. Detto and K. Haaga

Compute surrogates
[surr]=IAAFT(x, maxiter, threshold)

function IAAFT(plou)
Input: x is the time series
maxiter is the maximum number of iterations for the aaft
threshold is the maximum difference between the surrogate and the time series to accept convergence
Output: surr is the surrogate

ModObs=amplitude(spectrum(x)) /* spectrum and spectrum^{-1} correspond to Fourier transform and inverse Fourier transform, respectively */
OrderObs=sorted indices of x
Shuffle x
Option
/*K. Haaga*/ 1 Gaussian re-ordering: surr=x[sort($\mathcal{N}$(0,2))]
/*M. Detto*/ 2 Random permutation: surr=x[permutation]
EndOption

/* Init */

Option
/*K. Haaga*/
1 acfObs=autocorrelation(x)
2 acfTmp=autocorrelation(surr)
3 diffOld=$\sqrt{\sum_l(\text{acfObs}_l - \text{acfTmp}_l)^2)}$=f(x,surr) /*l is the lag*/
4 diffNew=diffOld+2thresh /* could be anything for diffNew-diffOld>thresh */
/*M. Detto*/
5 corrObs=cross-correlation(x,surr)=g(x,surr)
6 corrTmp=corrObs
EndOption

iter=0

/* Start iteration */

Option
1 convergence=|diffOld-diffNew|>thresh.AND.iter<maxiter
2 convergence=(1-corrObs^2)>thresh.AND.corrTmp<1.AND.iter<maxiter
EndOption

while .NOT.convergence **do**
iter=iter+1
phaseTmp = phase(spectrum(surr))
surrNew = spectrum^{-1}(ModObs*exp(phaseTmp))
surrNew=surrNew[OrderObs]
Option
1 diffOld=diffNew; diffNew=f(surr,surrNew)
2 corrObs=g(x,surr); corrTmp=g(surr,surrNew)
EndOption
surr=surrNew
end while
return surr

end function

Ebisuzaki

Ebisuzaki (1997)'s method has been specifically designed to analyze pairs of time series. We can compare his algorithm to the code in (Ye et al., 2018) (version 0.7.1), which is also the one used by K. Haaga. It is basically a random phase test and does not involve iterations. In this case, a random vector of phases (of the same length n as the time series)[2] is added to the Fourier transformation of the original time series, which is then transformed back to a time-domain surrogate. In Ye et al. (2018), the vector of phases Φ is symmetric so that $\Phi = (0, \phi_1, ..., \phi_{n-1}, \phi_n, \phi_{n-1}, ...\phi_1)$.

The Fourier transform is also slightly different for the Nyquist frequency (n/2) (it differs by a $\sqrt{2}$coefficient).

References

Detto, M., G. Bohrer, J. Nietz, K. Maurer, C. Vogel, C. Gough, and P. Curtis, 2013. Multivariate conditional granger causality analysis for lagged response of soil respiration in a temperate forest. *Entropy* **15**:4266–4284.

Ebisuzaki, W., 1997. A method to estimate the statistical signifiance of a correlation when the data are serially correlated. *JOURNAL OF CLIMATE* **10**:7.

Gouhier, T. C. and F. Guichard, 2014. Synchrony: quantifying variability in space and time. *Methods in Ecology and Evolution* **5**:524–533.

Schreiber, T. and A. Schmitz, 1996. Improved Surrogate Data for Nonlinearity Tests. *Phys. Rev. Lett.* **77**:635–638.

Schreiber, T. and A. Schmitz, 2000. Surrogate time series. *Physica D: Nonlinear Phenomena* **142**:346–382.

Ye, H., A. Clark, E. Deyle, S. Munch, J. Cai, J. Cowles, Y. Daon, A. Edwards, O. Keyes, J. Stagge, M. Ushio, E. White, and G. Sugihara, 2018. rEDM: Applications of Empirical Dynamic Modeling from Time Series.

[2] With limit constraints for the 0 and N/2 frequencies