

# On wavelet analyses with the `mvwt` package

July 29, 2020

To study synchrony, Keitt (2008) presents the wavelet modulus ratio, i.e.  $\rho(s, t) = \frac{\Lambda_{s,t}(|\sum_k w_k(s, t)|)}{\Lambda_{s,t}(\sum_k |w_k(s, t)|)}$  where  $\Lambda_{s,t}(y) = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(\frac{t-\tau}{s})^2} (y) d\tau$ ,  $s$  is the scale (“frequency”) and  $t$  is the location (“date”) for which the wavelet  $w_k$  is computed for species  $k$ . The package `mvwt` is meant to compute this index.

The package works as follows:

1. the function `mvwt` takes a table of dates and time series and computes the corresponding Morlet wavelets for each scale and location required. By default, the number of locations  $L$  is the number of dates, and is equal to the number of scales  $S$ .
2. the function `wmr.boot` takes the wavelets and compute the wavelet modulus ratio (`wmr`) of the observed time-series. It then computes  $R$  surrogates to provide the quantiles corresponding to the observed value at each scale and location.
  - (a) it first computes the observed `wmr` (called  $z$ )
  - (b) for each scale, it shifts the phase of the observed wavelets for the number of surrogates required. This can be compared to surrogates produced by shifting in the time-domain. In total, there is therefore  $S \times R$  bootstrapped wavelets.
  - (c) for each scale, it compares the value of the observed `wmr` at location  $t$  to the distribution of the index obtained with the surrogates. There is therefore  $S \times L$  comparisons to a distribution made of  $R$  samples. This is performed with the `ecdf` function (Fig. 2) which returns  $Pr(X \leq x)$ . The table made of  $S \times L$   $Pr(X \leq x)$  is called `z.boot`.
3. the function `image.mvwt` takes an object from `wmr.boot` (or `wmr`, which contains only the observed values of the index,  $z$ ). It first plots the values in  $z$  as an image ; then uses `z.boot`. We describe first the classical way it does this, and will then explain the modifications we made.
  - (a) it first computes  $1 - |1 - 2 * Pr(X \leq x)|$ . This formulation corresponds to the usual two-sided p-value, that is  $p_{val} = 2 \min(Pr(X \leq x), Pr(X \geq x))$ .
  - (b) it computes the adjusted p-value  $p_{val,adj}$  with a Benjamini-Yekutieli false discovery rate.
  - (c) it plots the contour for  $p_{val}$  and  $p_{val,adj}$  at the chosen  $\alpha$  level (0.05 by default).

## What we changed

There were several adjustments made to the `image.mvwt` function. The first ones were only cosmetic. Panels were created to be able to plot a `wmr` image as well as other plots on the same figure, add a colorbar and increase the

size of certain labels. We also changed the Benjamini-Yekutieli fdr to Benjamini-Hochberg following the suggestion of Referee 2 in Revision 1 (who actually only suggested that we changed Bonferroni to something else).

We also differentiated between low values and high values of the observed wmr. To do so:

1. we first compute  $p_{val}$  and  $p_{val,adj}$
2. for each z.boot value, we check if  $Pr(X \leq x) < Pr(X > x)$  (x has a low value and might therefore indicate compensation) or  $Pr(X \leq x) > Pr(X > x)$  (x has a high value and might therefore indicate synchrony).
3. we draw a contour at the alpha level with the following code: red for each  $p_{val}$  (or  $p_{val,adj}$ ) where  $Pr(X \leq x) > Pr(X > x)$  and blue for each  $p_{val}$  (or  $p_{val,adj}$ ) where  $Pr(X \leq x) < Pr(X > x)$ .

## What is the issue?

There were several issues (theoretical and not due to the package itself) noted thanks to the suggestions of Referee 1 in Revision 2.

- normalizing the time-series led to a change in observed values (see for instance the wavelet/wader wavelet plots in which the synchronizing effect of 2006 is much stronger with the scaled time-series). Normalised wavelet modulus ratios are not, however, comparable to the Gross index and the notion of compensation they examine loses its ecosystem level meaning (biomass replacement)
- increasing the number of surrogates from 100 (Fig. 1) to 1000, most of the significant zones disappeared. As the number of surrogates increase, we could assume that the distribution of possible values increases in width, and observed values which were really close to the limits of the distribution (most of them differ by  $10^{-2}$ – $10^{-3}$  from the 5% limits of the null distribution) are finally included in the distribution. This, however, is not clear when empirical cumulative distributions are plotted (Fig. 2).

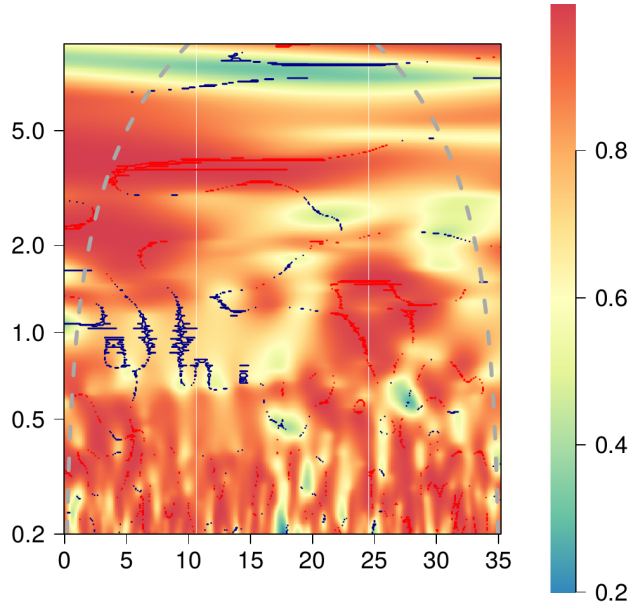


Figure 1: Reference wavelet analysis for the Cormoran/Heron/Egret case with 100 surrogates, BH correction and a fine grid (423×423)

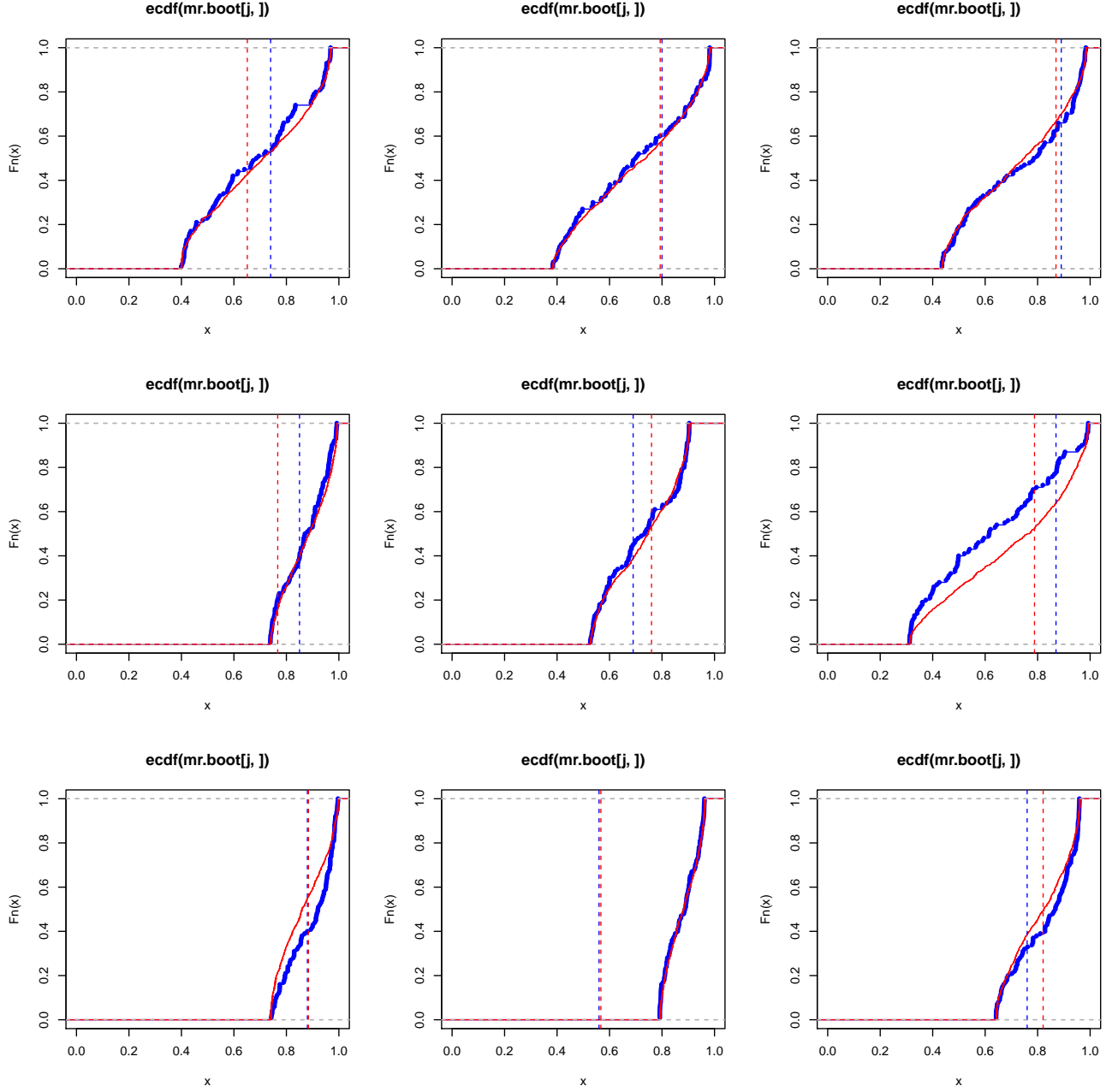


Figure 2: Empirical cumulative distribution of  $\rho(s, t)$  for several randomly chosen  $s$  and  $t$  with 100 (blue) and 1000 (red) surrogates. Observed values are indicated by vertical dashed lines.

- the p-value BH-adjustment could lead from 12% p-values  $< 0.1$  to 2% p-values  $< 0.1$  in the Cormoran/Heron/Egret case with 100 surrogates and from 11% to 0.2% with 1000 surrogates. When using a fine scale, there is indeed 178929 (423 scales  $\times$  423 locations) comparisons. It appears that only the p-values equal to 0 when not adjusted are considered significant with a correction (Fig. 3)

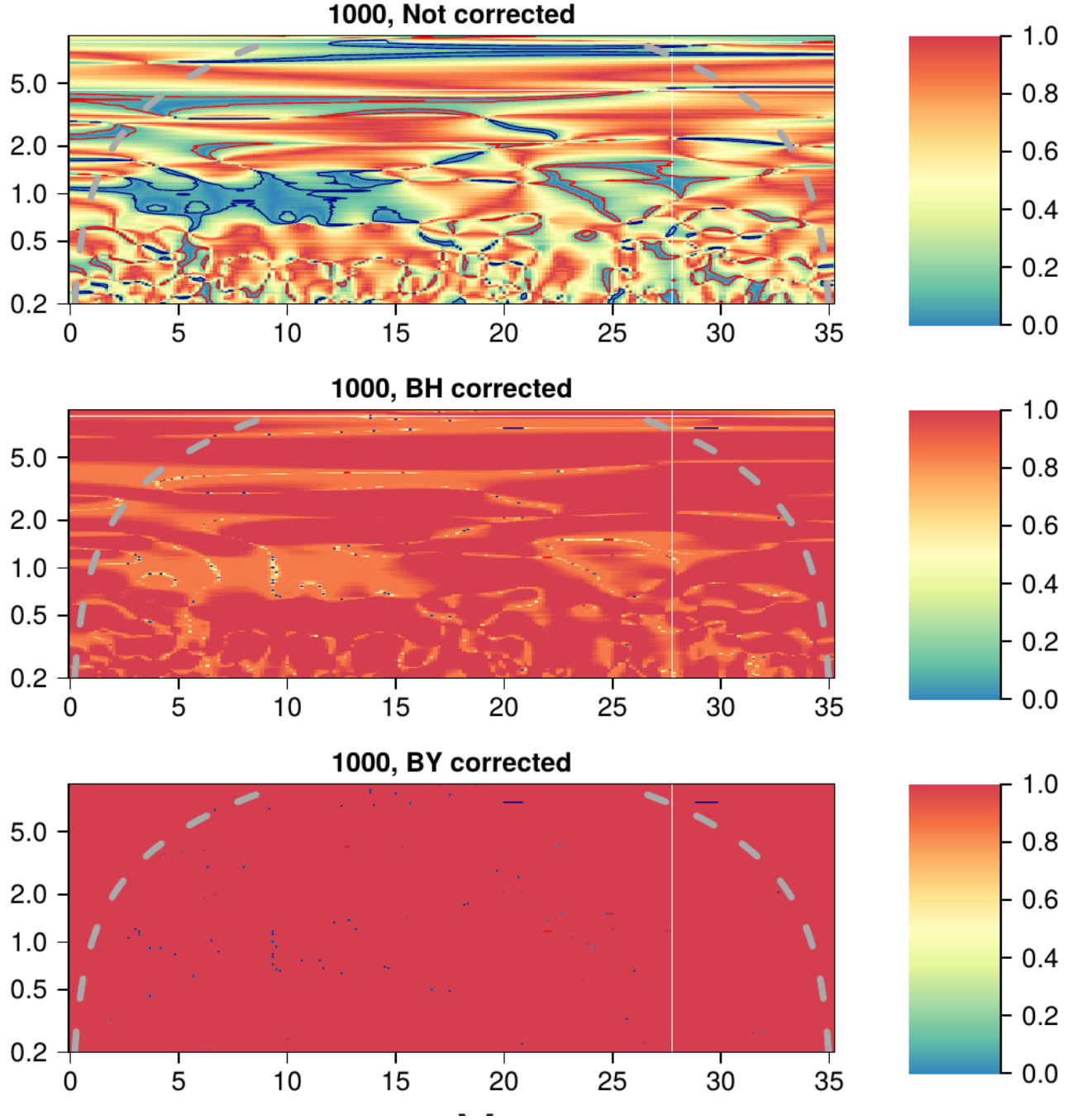


Figure 3: P-values images for 1000 surrogates without fdr correction, with the BH correction and with BY correction

### What we also checked

- The contours do correspond to small p-values; see Fig. 3.
- Instead of phase-randomization of the wavelets, we built new time series (that is, in the time domain) with the IAAFT method and then built the p-values using each image pixel's surrogate distribution, each surrogate  $\rho_j(s, t)$  being obtained through wavelet transformation and wavelet modulus ratio computation of the IAAFT-obtained time series. This could lead to even less significant zones (Cazelles et al. 2014, Journal of the Royal

Society Interface). In Fig. 4, the number of significant pixels has not changed markedly, but all significant  $\rho(s, t)$  are now in one block instead of being spread out.

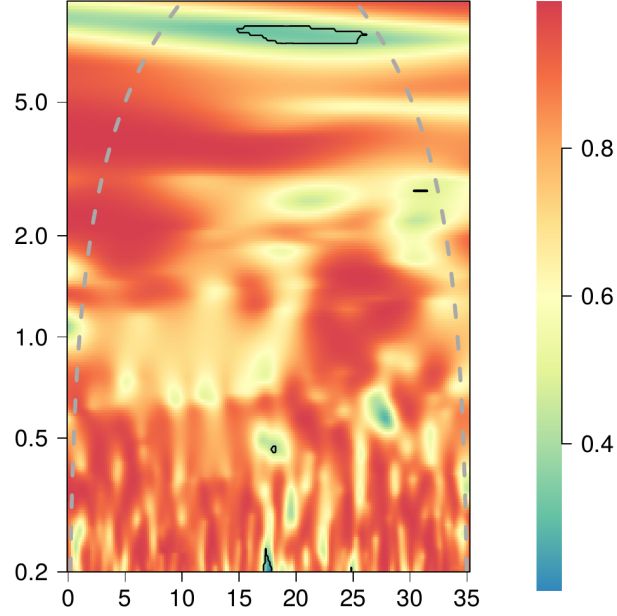


Figure 4: Wavelet analysis with 100 iaftf surrogates (in the time-domain) and a BH-correction.

- Finally, we changed the number of scales (from 423 to 100) and locations (from 423 to 212). This reduced the noise in significance contours (as well as the size of the images), see Fig. 5.

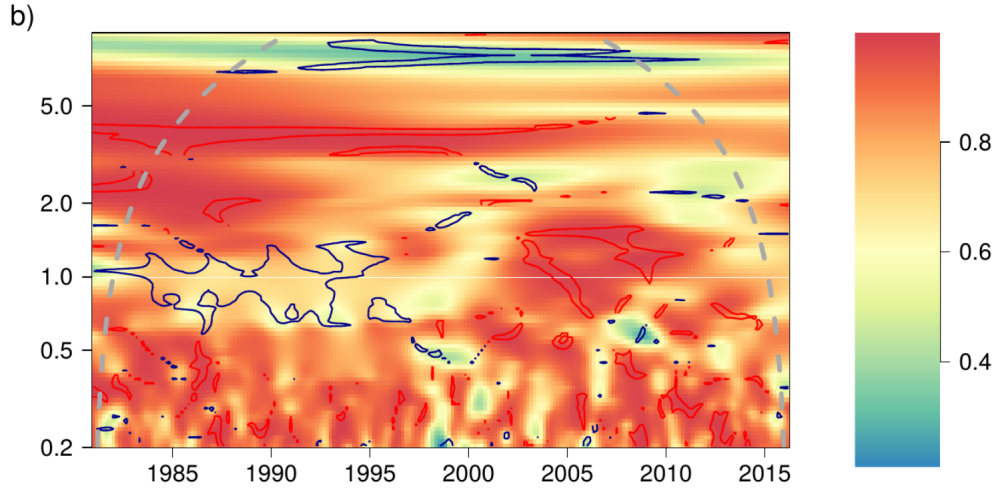


Figure 5: Wavelet analysis on a coarser grid, with 1000 surrogates and no correction