ROBUST DETECTION OF PERIODICALLY BEHAVING BIOLOGICAL TIME SERIES

Miika Ahdesmäki[1], Harri Lähdesmäki[1], Ronald Pearson[2], Heikki Huttunen[1] and Olli Yli-Harja[1]
[1]Institute of Signal Processing, Tampere University of Technology, miika.ahdesmaki@tut.fi
[2]ProSanos Corporation, ronald.pearson@prosanos.com

Introduction

Periodicity detection in time series measurements is a usual application of signal processing in studying biological data. Periodically behaving biological events can be of interest because of many reasons, e.g. periodicity in gene expression time series could suggest cell cycle control over the gene expression and so on. What is usually the case, is that there are lots of measured targets but very few time points per target. This is particularly true for gene expression measurements where there can be thousands of genes measured but only at few time points.

We present a method to assess what time series among a set of time series are statistically significantly periodic. The method is based on a recently introduced robust, rank-based, non-parametric spectral estimator that is used to estimate the spectra of the biological time series. Based on the spectral estimate of a time series, we calculate a statistic, the g-statistic, that is the chosen spectral component of interest divided by the sum of the all the spectral components of the time series at hand. This statistic is then evaluated for all the time series and based on the estimated distribution of the g-statistic, we can find a p-value for each time series telling us whether or not a strong periodic component is present. Multiple test correction of the p-values is then necessary and we use the Benjamini-Hochberg false discovery rate (FDR) to choose a cut-off value for the p-values accepted as periodic. This results in a robust testing procedure which is insensitive to a heavy contamination of outliers, missing-values, short time series, nonlinear distortions, and is completely insensitive to any monotone non-linear distortion.

The Model and Results

Assume the model for a periodic time series as

$$y_n = \beta \cos(\omega n + \phi) + \epsilon_n, \tag{1}$$

where $\beta \geq 0$, $\omega \in (0, \pi)$, $n = 1, \ldots, N$, $\phi \in (-\pi, \pi]$, and ϵ_n is an i.i.d. noise sequence. To test for the periodicity, define the null hypothesis as $H_0 : \beta = 0$, i.e., time series consists of the noise sequence alone, $y_n = \epsilon_n$.

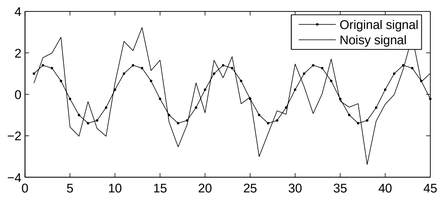

Figure 1: Three examples of time series, the first composed of a sine and additive standard Gaussian noise (topmost), the second with additive standard Gaussian and impulsive noise (middle), and the third with additive standard Gaussian noise and cubic distortion (bottommost).

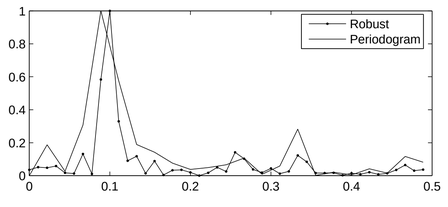

Figure 2: The spectral estimates for the time series in Figure 1, respectively, using both the standard periodogram and the proposed robust method.

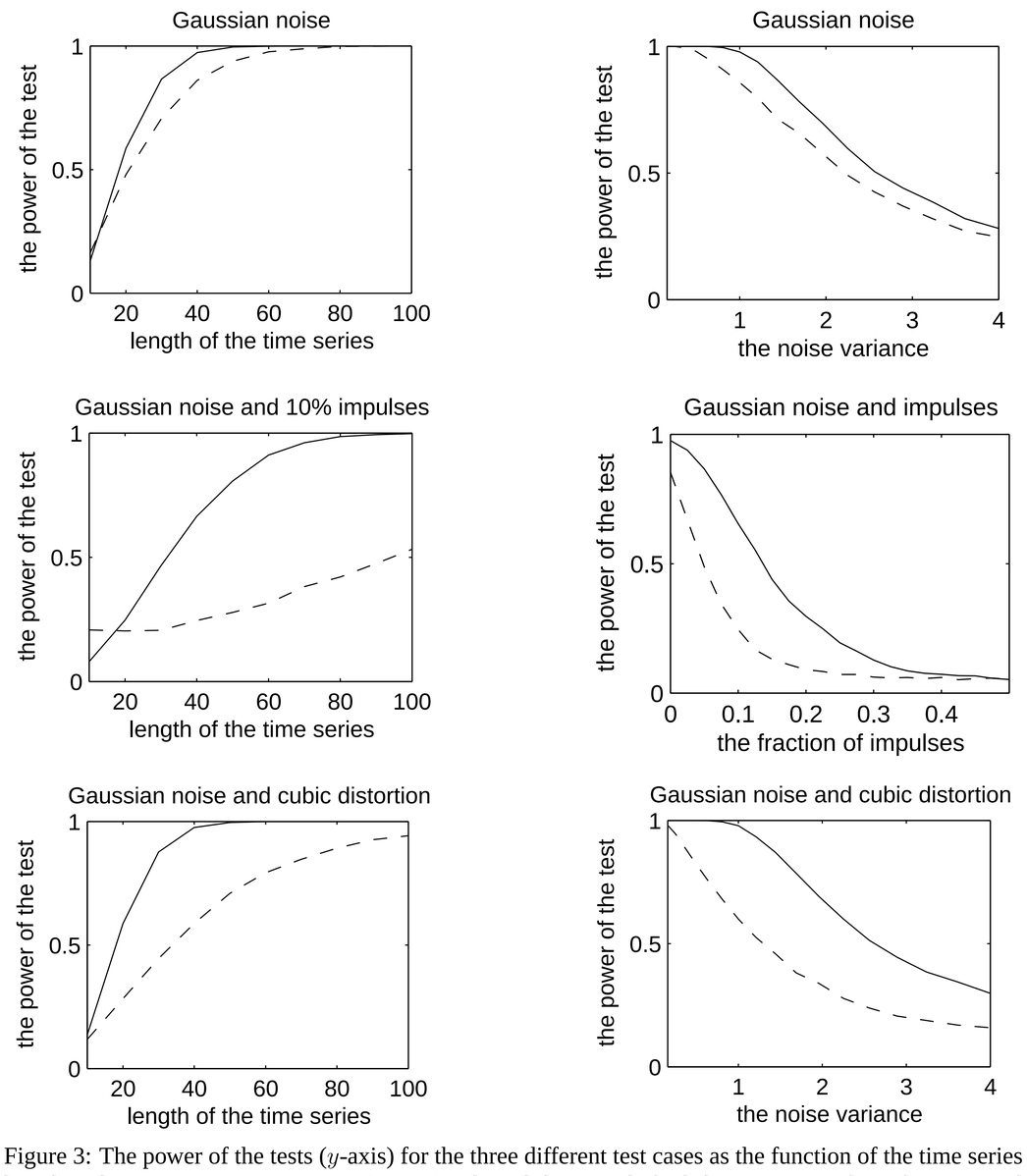

Figure 3: The power of the tests (y-axis) for the three different test cases as the function of the time series length and varying noise parameters (x-axis). The solid (resp. dashed) line corresponds to the proposed robust method (resp. Fisher's test). Three different types of non-idealities are considered, namely, pure standard Gaussian noise (the first row), standard Gaussian and impulsive noise (the second row), and standard Gaussian noise and x^3 distortion (the third row). The left (resp. right) column shows the results for different time series lengths (resp. different values of the noise parameters).

We first present some example time series created according to Eq. (1) that have a sinusoidal component and additive noise. The example time series can be seen in Figure 1, where different noise terms are used. The estimated spectra for the time series in Figure 1 are shown in Figure 2. As we can see, the robust spectrum estimator is clearly rejecting impulses and other artifacts better than the classical periodogram spectral estimator.

Next we compare the power of the presented method to the classical periodogram approach since there is a strong connection between these two methods. The power of the test, i.e., one minus the probability of the type II error (false negative), is estimated for three different test cases as well as for different time series lengths and for different noise parameters using 10000 Monte Carlo runs, see Figure 3. The significance level is set to $\alpha = 0.05$. In all the three cases, the case-specific noise assumptions are used for both the null hypothesis ($\beta = 0$) and the alternative hypothesis ($\beta > 0$). In this simulation, we use the signal model shown in Eq. (1) with $\beta = \sqrt{2}$ to represent a periodic signal (i.e., the alternative hypothesis). In the right column of Figure 3, the length of the time series is set to 40 and the power is shown as the function of varying noise parameters.

Summary

As the results show, the presented robust method performs well and clearly outperforms analyses based on the classical periodogram.