A Comparison of Segmentation Procedures and Analysis of the Evolution of Spectral Parameters

José B. Hernández Ch.
Universidad Central de Venezuela
Caracas, Venezuela

Joaquin Ortega S.
CIMAT, A.C.
Guanajuato, Gto, Mexico

ABSTRACT
In this work we consider the evolution of power spectra of waves during a period of one year. Soukissian and Samalekos (2005) have proposed a segmentation method for significant wave height based on determining periods of stability, increase and decrease using time-series techniques. The second segmentation method is based on the mean value over a moving window, and uses a fixed-width band to determine the change-points in the register. We compare both segmentation methods for several spectral characteristics and give a statistical analysis of duration and intensity of sea states in each case.

KEY WORDS: Spectral analysis; stationary periods; time series; segmentation procedure.

INTRODUCTION
In this work we consider the evolution of power spectra of waves during a period of one year with data from one recording station situated at Waimea Bay, Hawaii. Using the wave-height record we calculate the spectra every 15 minutes in order to capture the short term evolution of some wave characteristics that can be obtained from the spectra. WAFO was used for obtaining the spectra and the spectral characteristics.

Soukissian and Samalekos (2005) have proposed a segmentation method for significant wave height based on determining periods of stability, increase and decrease using time-series techniques. Their method is based on local linear regression and the initial and end points of the intervals are extreme points (local maxima and minima) of the time series. They use a cost function to determine the best configuration of intervals. We apply this method to some spectral characteristics and compare the results obtained with another segmentation method which will be described next.

The second segmentation method is based on calculating mean values over moving windows, and using a fixed-width band to determine change points in the wave-height data. Those intervals in which the values remain within a fixed-width interval around the mean are considered to be stationary, those in which the values go above (or below) will be considered increasing (or decreasing). In this way the stationary, increasing and decreasing intervals are determined. Both methods were implemented in MATLAB.

We will consider the following spectral characteristics: Significant wave height (\(H_m = 4\sqrt{\text{Var}(X)}\)), spectral moments of order zero (\(m_0 = \int_0^\infty \omega^2 \rho(\omega) d\omega\)), and two \(m_2 = \int_0^\infty \omega^2 \rho(\omega) d\omega\) and up-crossing peak periods (\(T_p = 2\pi \sqrt{m_0/m_2}\)).

After calculating the spectral characteristics the results were smoothed using a finite moving average filter of order 5, to get rid of the local noise, see Brockwell and Davis (1996) for details.

The time series we considered are from Station 10601 in Waimea Bay, Hawaii, with the following characteristics. Deployment latitude: 21°40.364' N, longitude: 158°06.949' W, water depth (m): 198.00. The time series has a sampling rate of 1.280 Hz.

The rest of the paper is organized as follows. In the next section we describe the Soukissian and Samalekos algorithm, and apply this method to the time series. Next we describe the band method and its application to the time series. In the following two sections we make an analysis of the results for both methods and give our conclusions.

SOUKISSIAN’S ALGORITHM
Consider a time series of significant wave height observations \(H_m = h_1, h_2, ..., h_n\) with \(n\) terms; the goal is to find a \(k\)-segmentation of \(H_m\) i.e. \(H_m = H_{m1}, H_{m2}, ..., H_{mk}\) with \(H_{mi}\) disjoint and non-overlapping intervals.

The first step in time series segmentation is to define a representation model that approximates the data in each segment. Once we find the representation model, the quality of the approximation is evaluated by a cost function to minimize the representation error. For this a linear regression model is used (Charbonnier, 2005), and the representation error is defined based on the sum of squares of distances between the actual values of the time series and the values of the representation model fitted.

The total cost of a \(k\)-segmentation is

\[
\text{COST} = \sum_{i=1}^{k} \text{cost}(i,k)
\]

where \(\text{cost}(i,k)\), \(1 \leq i \leq k\) is the cost of \(i\)-th segment of a \(k\)-segmentation.

The linear regression model is employed because \(H_m\) data exhibit