A scan statistics approach to picking seismic wave arrival times using arbitrary-sized sliding windows

Aristides Zenonos

Supervisor: Prof. Nicholas Rawlinson

a.zenonos@abdn.ac.uk

UNIVERSITY OF ABERDEEN

Introduction and Aims

•Seismic wave arrival times are often used in seismic tomography to constrain the structure of the Earth's interior.
•The objective of scan statistics is to detect whether a subset of data points (for example in a time series) deviates from the population in a way that is statistically significant.
•This is done with the use of arbitrary-sized sliding windows, which progressively analyses the dataset to locate regions with high statistical significance.
•It is observed that seismic wave arrivals are not easily seen in a raw seismogram, due to the presence of coda noise and the oscillatory nature of the seismogram.

Methodology

•In this work it is shown that by computing the spectrogram and integrating over teleseismic frequencies, the scan statistics event detection approach can be used reliably on seismic wave arrivals.

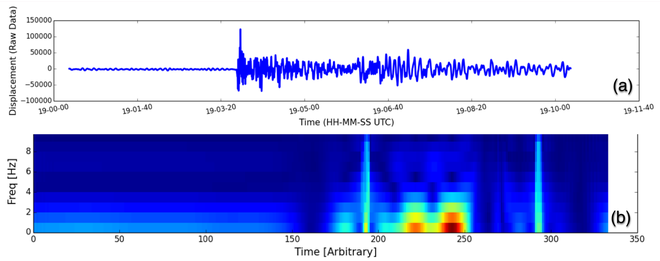

Fig. 1: The first figure (a) shows the raw data as downloaded from FSDN. AK135 model is used to estimate the p-wave arrival. Scan statistics is applied on an interval of 7.5s before and after the estimated arrival. In the second figure (b) the spectrogram is computed for teleseismic frequencies up to 10Hz.

•In order to minimise the influence of the noise in a seismic wave arrival, we convert the time series into rank-space i.e. rank each point in the time series from 1 to the number of total points *N* with 1 being the lowest and *N* the greatest as shown in Fig. 2.

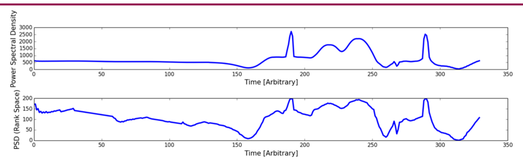

Fig. 2: Power Spectral Density (PSD) is converted into rank-space data where minimum value gets the value of 1 and the maximum the value of N.

•The approach for detecting the events, is based on examining the sum of the values inside a sliding window.

•First, a probability distribution is generated using a Monte Carlo approach in which we repeatedly select w (where w is the window length) unique integers in the range [1,N] (where N is the length of the time series) and compute their sum.
•The total number of times each sum occurs in the simulation divided by the total number of Monte Carlo samples yields the probability of each sum occurring.
•A probability distribution depends only on the number of values being added together and the length of the time series, so a new probability distribution needs to be computed for each distinct choice of w and N. as shown in Fig 3.

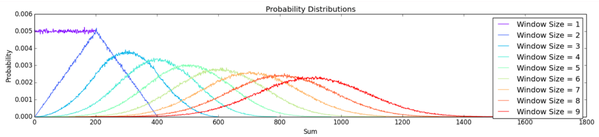

Fig. 3: Numerically generated probability distributions for N=200, and w from 1 to 9.

•During implementation, as the sliding window moves across the time-series, the p-value is obtained by comparing the sum of values in the sliding window with the probability distribution; statistically significant events will have a low p-value.
•By following this method and plotting the probability surface (contour plot of probabilities for all starting point and window size pairs) shown in Fig. 4 some interesting results are revealed.

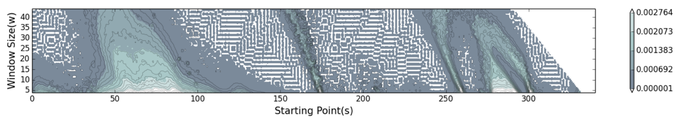

Fig. 4: Contour plot of probabilities for all starting point and window size pairs. Areas with low probability indicate possible event regions.

•The probability surface plot indicates clearly the event regions and they are suitable for further investigation using an optimisation method.
•Powell's optimisation method is used on the event regions to identify minima, which correspond to the occurrence of an event.

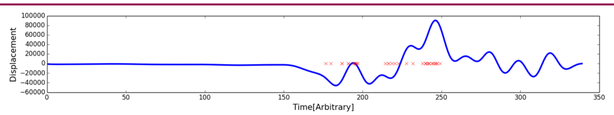

Fig. 5: Minima are found with the use of Powell's optimisation algorithm.

•In combination with this method, random restarts are also applied.
•Random restarts alleviate the dependence of the resulting minima on the original search vector of the local optimisation method.

•An agglomerative clustering algorithm is applied after random restarts to detect similar instances of the same solution and to find the best solution for the event region.

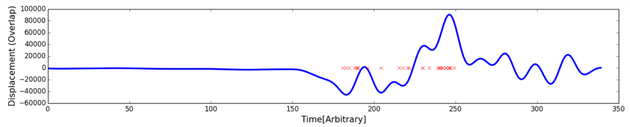

Fig. 6: Random restarts and an agglomerative clustering algorithm was applied.

Results

•Using the aforementioned procedure P-wave arrival times were tested and reliably found with a low rate of false positives.

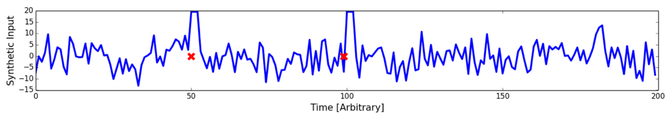

Fig. 7: Results for synthetic data with peaks at 50 and 100. The scan statistics parameters were: N=200 and w=5.

•P-wave arrival from passive data. The point of inflection prior the first peak was obtained.

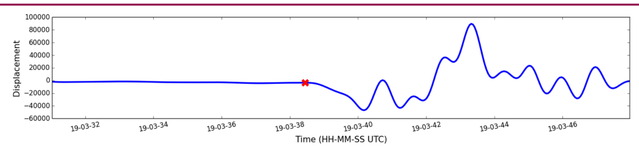

Fig. 8: Results for the P-phase arrival with parameters: Date: 2016-02-01, Time: 19:00:46, network: 'II', station: 'MSVF' and channel: 'BHZ'.

Conclusions

•A method for detecting seismic wave arrivals using rank-space based scan statistics is presented.
•This method does not require to model the noise and it has no restriction of fixed-size sliding window.
•Most scan statistics approaches in literature require either or both of these two restrictions.
•The method was tested with both synthetic and real data. It was observed that synthetic data were more reliable than the real ones.
•Minimisation techniques were used to reduce the brute force complexity including random restarts, overlap parameter of the clustering algorithm and the acceptable error which provides a tractable probability distribution.