EGU23-5212, updated on 25 Apr 2024
https://doi.org/10.5194/egusphere-egu23-5212
EGU General Assembly 2023

Event clustering for infrasound monitoring system

maurice charbit, alexis le pichon, benjamin poste, **françois roueff**, and julien vergoz

In the area of infrasonic monitoring system, an important task is to clustering the various cells containing a
signal of interest to a reduced event number. Indeed, to each time-frequency cell is attached a source direction of arrivals,
many of which arise from the same physical event spanning over a large time-frequency window. In PMCC (Progressive
Multi-Channel Cross-Correlation) the clustering is based on ad hoc rules and metrics provided by a large empirical expertise.

In this study we present a new approach based on a statistical model, associated to the multichannel maximum-
likelihood (MCML). More specifically, for each time-frequency cell, the MCML provides the estimations of the slowness
vector and of the signal-to-noise ratio (SNR), and the p-value computed from the generalized likelihood ratio (GLR).
These quantities are collected within a large window, typically 10 Hz by 1 hour. To reduce the computational time, only
the cells with p-values below a threshold are considered. The proposed mixture model (MM) is based on the following
4-dimensional vector: the time location, the log of the frequency location and the two components of the slowness vector.

Each cluster is modeled by a distribution, chosen in a flexible catalog that can still be improved. Today the catalog
consists of Gaussian distribution, uniform distribution and mixture of Gaussian and uniform. A few examples: short
time event is modeled by a 4D Gaussian, permanent event in a given frequency band, as microbarom or wind turbine, is modeled by
a 2D Gaussian for the DOA, a full time uniform distribution for the time location and a uniform distribution in the
known frequency band. For the DOA modeled by a Gaussian, the covariance is taken as the asymptotic covariance of
the MCML, using the estimated SNR in the corresponding cell. Moreover we introduce a specific cluster to trap falsely
detected signals, modeled by a full uniform distribution in the four dimensions.

The estimation of the parameters of the MM is performed by the so-called Expectation Maximization algorithm.
Then the maximum a posteriori provides the final clustering. We also present an estimation of the number of events
based on the Bayesian information criterion (BIC). Many real observations are considered to illustrate the method.

The main advantages of the proposed method are (i) taking into account the p-value for selecting the cells to cluster,
(ii) the flexibility of the model catalog, (iii) the statistical interpretation of the results.