

ASSOCIATION OF ARRAY PROCESSING AND STATISTICAL MODELLING FOR SEISMIC EVENT MONITORING

Paul Bui Quang, Pierre Gaillard, Yoann Cano

CEA, DAM, DIF
F-91297 Arpajon, France

ABSTRACT

We associate an array processing method, called progressive multi-channel correlation (PMCC), and statistical modelling, to detect and classify seismic events. PMCC detects any coherent wavefront crossing an array of seismometers, including the wavefronts not generated by actual seismic events. We use machine learning techniques to classify the PMCC detections between "events" and "noise". These techniques are based on the statistical modelling of features extracted from the seismic signal. The features we select combine features computed directly from the raw signal and features retrieved by the PMCC detector. We apply our method on a real data set from the Songino seismic station, in Mongolia. We compare the performance of four classifiers: Gaussian naive Bayes classifier, logistic regression, Gaussian mixture models (GMM), and hidden Markov models (HMM). In our case study, the GMM and the HMM yield the highest performance.

Index Terms— progressive multi-channel correlation, classification, Gaussian mixture models, hidden Markov models, seismic monitoring

1. INTRODUCTION

A seismic event releases energy that propagates in the Earth structure in the form of seismic waves [8]. These seismic waves can be body waves, traveling through the interior of Earth, or surface waves. Body waves, which carry most of the seismic energy, are of two types: primary waves (P-waves) and secondary waves (S-waves). P-wave are compressional wave polarized in the propagation direction, while S-wave are shear waves polarized orthogonally to the propagation direction. The P-wave speed is approximately 60% higher than the S-wave speed in the crust and in the mantle. Time pickings of P-wave and S-wave arrivals from an earthquake in Mongolia are plotted in Figure 1.

Seismometers monitor the pulse of the Earth. They record seismic events as well as seismic background noise. Seismic events can be earthquakes produced by plate tectonics or volcanic activity, or they can be related to human activity (quarry blast, nuclear explosion, etc.). Seismic events create particular patterns in the signal recorded by the seismometers. De-

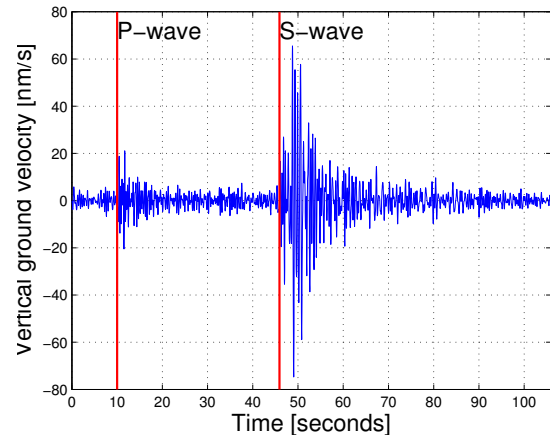


Fig. 1: Earthquake with P-wave and S-wave arrivals recorded in Mongolia and observed at 307km distance.

tecting and classifying these patterns is a difficult task, which involves dedicated processing tools as well as an important amount of human analysis.

In this paper, we use an array processing method called progressive multi-channel correlation (PMCC) as a detector. PMCC detects coherent wavefronts crossing a seismometer array and estimate the propagation parameters, i.e., the propagation direction and velocity. However, PMCC triggers detections even when the wavefront is not generated by an actual seismic event, e.g., road traffic, avalanche, cracking ice on a lake, etc. Hence, it is necessary to discriminate the detections of events and the detections not related to an event, referred to as "noise". For that purpose, we transform the seismic signal using an adapted feature representation, then we train statistical models to classify the features between "event" and "noise". We compare the performance of four machine learning classification techniques: Gaussian naive Bayes classifier, logistic regression, Gaussian mixture models, hidden Markov models.

The outline of the paper is as follows. In Section 2, we describe the PMCC algorithm. In Section 3, we present the feature representation and the statistical approach that we use to classify detections as events or noise. We apply our method in Section 4 on a real dataset from the Songino seismic station

in Mongolia.

2. ARRAY PROCESSING FOR DETECTION

Progressive multi-channel correlation (PMCC) is an array processing method designed to detect a coherent wavefront crossing an array of sensors and to estimate its propagation parameter, i.e., its propagation direction and speed. PMCC is used for seismic [4] and infrasound [3] event monitoring. A comprehensive presentation of the method can be found in [3] in the infrasound context.

The PMCC algorithm processes multi-channel signals to compute *pixels*, which are aggregated in *families*. A pixel is computed using the signal of sensors selected from the array and filtered in a given elementary time–frequency domain. The correlations between the filtered signals are computed and time delays are derived from them. A time delay Δt_{ij} is defined as the difference between the arrival times of a wavefront at two different sensors i and j . The consistency, defined as

$$C = \sum_{i < j < k} |\Delta t_{ij} + \Delta t_{jk} + \Delta t_{ki}|^2, \quad (1)$$

is then computed. This quantity is close to 0 when a coherent wavefront is crossing the array (and greater than 0 when the sensors only record incoherent noise), hence it is taken as a detection criterion. The consistency is sequentially computed using a growing number of sensors starting from an initial triplet. The sensors are progressively added as long as the consistency remains low. At the end of this process, if the final consistency remains close to 0 and the number of involved sensors is sufficiently large, the propagation parameters (azimuth and speed) are computed and stored with the time–frequency coordinates in the pixel. Starting from a triplet of sensors with small aperture allows a high detection rate, while the progressive approach leads to a better estimation of the propagation parameters [4]. Multiple pixels corresponding to multiple overlapping time–frequency windows are independently computed.

The next step of the detection process consists in aggregating pixels with similar attributes into detection families. The metric between pixels which allows family aggregation (i.e., pixel clustering) is a weighted Euclidean distance based on the time, the frequency, and the stored propagation parameters. If the number of pixels aggregated in a family is large enough, a detection is triggered. The propagation parameters of the detected wavefront are computed by averaging the azimuth and speed attributes of the family pixels.

Pixel and family computations are respectively described in Algorithms 1 and 2.

Algorithm 1 Pixel computation.

```

input multi-channel signals filtered in time–frequency
window centered at  $(\tau, f)$ 
select triplet of sensors  $\mathcal{S}$ 
compute time delays  $\Delta t_{ij}$  for all sensors  $i, j$  in  $\mathcal{S}$ 
compute consistency  $C$  from time delays  $\Delta t_{ij}$  (see Eq. (1))
while  $C < \varepsilon$  and  $|\mathcal{S}| < S_{\max}$  do
   $\Delta t'_{ij} \leftarrow \Delta t_{ij}$  for all sensors  $i, j$  in  $\mathcal{S}$ 
   $C' \leftarrow C$ 
  add new sensor to sub-array  $\mathcal{S}$ 
  compute time delays  $\Delta t_{ij}$  for all sensors  $i, j$  in  $\mathcal{S}$ 
  compute consistency  $C$  from time delays  $\Delta t_{ij}$  (see
  Eq. (1))
end while
if  $C < \varepsilon$  and  $|\mathcal{S}| \geq S_{\min}$  then
  compute azimuth  $\theta$  and speed  $v$  from time delays  $\Delta t_{ij}$ 
  return pixel  $(\tau, f, \theta, v, C)$ 
end if
if  $C \geq \varepsilon$  and  $|\mathcal{S}| - 1 \geq S_{\min}$  then
  compute azimuth  $\theta'$  and speed  $v'$  from time delays  $\Delta t'_{ij}$ 
  return pixel  $(\tau, f, \theta', v', C')$ 
end if

```

3. STATISTICAL CLASSIFICATION

3.1. Feature extraction

The raw data are seismic multi-channel signal segments that have triggered PMCC detections. To enhance relevant information with respect to our classification problem, we extract feature vectors from this raw data [2]. We select features computed from the seismic signal (standard features), and features retrieved by the PMCC detector (PMCC features).

For the standard features, the signal from each channel of the array is filtered in three overlapping frequency bands: low frequency band [0.8, 3], middle frequency band [1.5, 6], high

Algorithm 2 Family computation.

```

input multi-channel signals
for all time–frequency domains do
  filter signals in the time–frequency domain
  run Algorithm 1 on filtered signals
end for
aggregate pixels with similar attributes in families
for all families do
  if (number of pixels in family)  $\geq P_{\max}$  then
    trigger detection
    compute family azimuth and speed by averaging over
    pixels
  return families
end if
end for

```

frequency band [3, 9] (in Hz). For each channel, three standard features are computed after filtering in each frequency band: kurtosis, relative spectral energy, envelope variation (see Table 1). The feature values are then averaged over the channels, yielding nine standard features.

Regarding the PMCC features, we select four outputs from the PMCC detector: estimated wavefront speed, consistency, number of pixels of the detection, ratio of the maximal and minimal frequencies of the pixels (see Table 2).

For each processed signal segments, we take the logarithm of each standard and PMCC features and stack them to get a 13-dimensional feature vector.

kurtosis	$\frac{\frac{1}{T} \sum_{t' \in \{t \pm T/2\}} (S_{t'}^i - \bar{S}_t^i)^4}{\left[\frac{1}{T} \sum_{t' \in \{t \pm T/2\}} (S_{t'}^i - \bar{S}_t^i)^2 \right]^2}$ <p>where $\bar{S}_t^i = \frac{1}{T} \sum_{t' \in \{t \pm T/2\}} S_{t'}^i$</p>
relative spectral energy	$\frac{\sum_{f_{i,k} \in [f_i^{\min}, f_i^{\max}]} c_t(f_{i,k}) ^2}{\sum_i \sum_{f_{i,k} \in [f_i^{\min}, f_i^{\max}]} c_t(f_{i,k}) ^2}$
envelope variation	$\sum_{t' \in \{t \pm T/2\}} S_{t'}^{i,\text{env}} - S_{t'-1}^{i,\text{env}} $

Table 1: Standard features computed over the time window centered at t with length T (for one channel). $i = 1, 2, 3$ denote low, middle and high frequency bands respectively. $\{S_t^i\}_t$ is the seismic signal filtered in the frequency band $[f_i^{\min}, f_i^{\max}]$ and $\{S_t^{i,\text{env}}\}_t$ is its envelope. $c_t(f)$ is the Fourier coefficient at frequency f computed over the time window $\{t \pm T/2\}$.

number of pixels	$ \mathcal{F}_t $
consistency	$\frac{1}{ \mathcal{F}_t } \sum_{j=1}^{ \mathcal{F}_t } C_{j,t}$
speed	$\frac{1}{ \mathcal{F}_t } \sum_{j=1}^{ \mathcal{F}_t } v_{j,t}$
frequency ratio	$\frac{\max_j f_{j,t}}{\min_j f_{j,t}} - 1$

Table 2: PMCC features computed over the time window centered at t with length T . $\mathcal{F} = \{p_j\}_j$ denote a PMCC family and $p_j = (\tau_j, f_j, v_j, \theta_j, C_j)$ denote a pixel (i.e., a vector containing the time–frequency coordinates, the estimated azimuth and speed, and the consistency).

The statistical classifiers we consider in the sequel are trained using samples of feature vectors.

3.2. Classification and model training

We consider two classes: the "event" class or positive class, and the "noise" class or negative class. A distribution density of feature vectors (likelihood) is associated with each class. This density is denoted f_1 for the event class and f_0 for the noise class. We associate a feature vector Y (extracted from a signal segment) with the event class if $\frac{f_1(Y)}{f_0(Y)} \geq s$, or with the noise class if $\frac{f_1(Y)}{f_0(Y)} < s$, where $s \in \mathbb{R}^+$ is a decision threshold [9].

This classification based on a likelihood ratio has a Bayesian interpretation. Indeed, let $c \in \{0, 1\}$ be the sought class and let $p(\cdot|Y)$ be the conditional probability of c with respect to Y . The maximum a posteriori (MAP) estimator of c is defined as

$$\hat{c} = \operatorname{argmax}_{c \in \{0,1\}} \{p(c|Y)\}.$$

According to Bayes' rule,

$$p(c|Y) \propto f_c(Y) \pi_c,$$

where π_c is the prior probability of c . Thus, estimating c by the likelihood ratio is equivalent to estimating c by the MAP when the decision threshold is $s = \frac{\pi_1}{\pi_0} = \frac{\pi_1}{1 - \pi_1}$. For example, when $\pi_0 = \pi_1 = 1/2$, which corresponds to a noninformative uniform prior on the classes, then $s = 1$.

Since the densities f_0 and f_1 are unknown, they must be learned from training data (feature vectors) for which we know the class. Here, we consider four well-known machine learning methods:

- Gaussian naive Bayes classification [2, Chap. 8],
- logistic regression [2, Chap. 4],
- Gaussian mixture models (GMM) [2, Chap. 9],
- hidden Markov models (HMM) [7].

In these methods, the unknown densities f_0 and f_1 are supposed to belong to a parametrized set of densities. Learning f_0 (resp. f_1) consists in estimating the parameter such that it best fits the training data belonging to the noise (resp. event) class. This is done by maximum likelihood estimation. The optimization is straightforward for the naive Bayes classifier and the logistic regression, but it requires an expectation–maximization algorithm for the GMM [2, Chap. 9] and the HMM [7].

Among the four methods listed above, the HMM is different because it does not model simple feature vectors, but feature vector *sequences* (i.e., multidimensional time series)¹.

¹In this article we consider HMMs with multidimensional Gaussian observations and not symbolic observations [7].

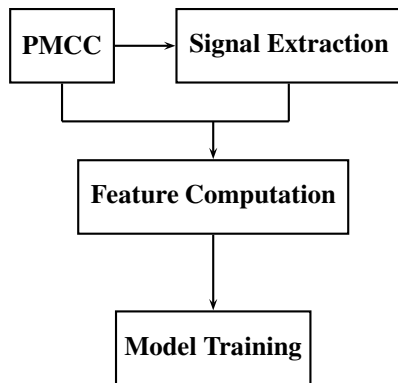


Fig. 2: Processing flowchart.

Hence, seismic signal segments are transformed into simple feature vectors when used for naive Bayes, GMM, and logistic regression, whereas they are transformed into feature vector sequences for HMMs, using a computation window sliding over the signal [1, 6].

The overall processing, including the PMCC method and the statistical model training, is summarized in Figure 2.

4. CASE STUDY

As a case study, we consider the Songino seismic station in Mongolia, which belongs to the International Monitoring System of the Comprehensive Nuclear-Test-Ban Treaty Organization². To build the training data and to assess the classification performance, we use as ground truth a seismic bulletin provided by analysts from the Research Center of Astrophysics and Geophysics (RCAG) of the Mongolian Academy of Sciences. The bulletin contains the P and S-wave arrival time pickings.

Signal segments corresponding to PMCC detections from August 25th to August 31st 2008 (seven full days) are extracted and transformed into feature vectors to constitute the training data. If the detection matches a seismic wave arrival from the bulletin, it is associated with the class "event". Otherwise, it is associated with the class "noise". Training data are made of 8816 PMCC detections, among which 7721 correspond to noise and only 1095 to events. The class population size is thus highly uneven (seven times more noise than events), which can lead to poor classification performance [5]. The noise class is therefore undersampled to get the same size than the event class. For the training data, the extraction of

signal segments starts 3s before the arrival time picking and ends 10s after, so that the extracted segment are 13s long.

To test the four classifiers, we use data from September 1st to September 2nd 2008 (two full days), which were not used for training. Test data are made of 3916 PMCC detections (3664 for noise, 252 for events). These detections are classified as event or noise, and the estimated class is compared with the actual class thanks to the seismic bulletin. For the test, the arrival time pickings are supposed to be unknown. Therefore, the extraction of the signal segments is done with regard to the PMCC detection, and not to the actual arrival from the bulletin. The extraction starts 3s before the detection starting time (i.e., $\min_j \tau_j$ with the notations from Table 2) and ends 10s after it.

For HMMs, the feature vectors are computed over a 2.5s time window sliding with a 0.5s step. Thus, for HMM training, a 13s-long signal segment is transformed into a sequence of 21 feature vectors.

ROC curves are displayed in Figure 3 using classification results on the training data. The GMM and the HMM yield the best performance regardless of the decision threshold. In particular, in an operational monitoring perspective, it is more important not to miss any event than to classify noise as event, i.e., true positive is preferable than true negative. The naive Bayes classifier, the logistic regression, the GMM, and the HMM reach a true positive rate above 99% for a true negative rate of 59%, 36%, 61%, and 64% respectively. Thus, the GMM and the HMM get the best performance in this regard: they attain a high classification performance for the event class at the price of a moderate loss of performance for the noise class.

The confusion matrices on the test data are displayed in Table 3. The highest classification rate for the noise class is obtained by the logistic regression, but at the price of a low classification rate for the event class. The GMM and the HMM exhibit good performance for both classes. The HMM has the highest classification rate for the event class. In the results in Table 3, the decision threshold is set to $s = 1$, which in Bayesian terms corresponds to the absence of *a priori* knowledge about the class.

The GMM and the HMM yield the best performance because these models are flexible, i.e., they are able to accurately describe various patterns of data. However, recall that the training of the GMM or the HMM involves an EM algorithm, so that the computation cost is much higher than for the naive Bayes classifier or the logistic regression.

Lastly, note that the HMM models feature vector sequences, whereas the other methods model single feature vectors. In the HMM, one PMCC detection corresponds to a feature vector sequence (computed with a short time window sliding over the signal). In the naive Bayes classification, the logistic regression, and the GMM, one PMCC detection corresponds to one feature vector (computed with a fixed large time window). Thus, the feature data processed by

²<http://www.ctbto.org/verification-regime/building-theinternational-monitoring-system/the-future-role-of-theinternational-monitoring-system/>

the HMM is less "smoothed" than for the other classifiers, and it contains more information (but also possibly irrelevant information with respect to the classification problem).

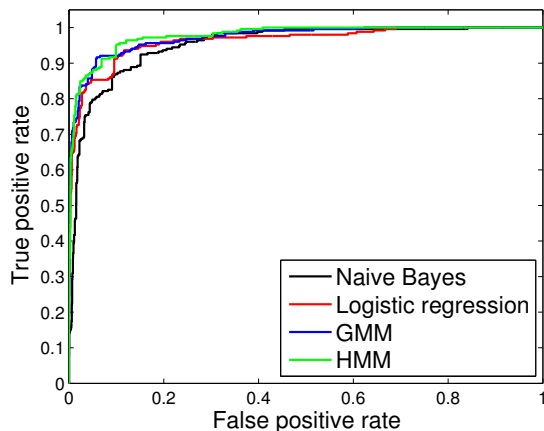


Fig. 3: ROC curves (training data).

	event	noise		event	noise
event	79	21	event	72	28
noise	5	95	noise	17	98
(a) Naive Bayes classifier			(b) Logistic regression		
	event	noise		event	noise
event	84	16	event	87	13
noise	28	97	noise	5	95
(c) GMM			(d) HMM		

Table 3: Confusion matrices (percentage, test data). The actual class is in line and the estimated class is in column.

5. CONCLUSION

PMCC is an array processing method, which detects coherent wavefronts crossing a seismometer array. PMCC also triggers a detection for wavefronts not generated by actual seismic events, hence the need to distinguish detections triggered by actual events from detections not related to an event.

We use machine learning techniques to perform this classification task. These techniques are based on the statistical modelling of features extracted from the raw seismic signal. The features we select are computed directly from the signal (standard features) or retrieved by the PMCC detector (PMCC features).

We compare four classifiers: Gaussian naive Bayes classifier, logistic regression, GMM, and HMM. This comparison is performed with a real data set from the Songino seismic station, in Mongolia. The GMM and the HMM yield the best classification performance. However the likelihood maximization for the GMM and HMM involves costly computa-

tional techniques (EM algorithm), whereas it is straightforward for the naive Bayes classifier or the logistic regression.

This paper demonstrates that low-level array processing can be fruitfully associated with high-level statistical modelling techniques, which opens perspectives for operational seismic monitoring.

To continue this work, other machine learning techniques can be considered to further improve the classification performance, such as artificial neural networks or support vector machines. Besides, more difficult classification problems can be considered, such as the classification of events with respect to their epicenter distance (e.g., regional and teleseismic events).

Acknowledgement

The authors wish to thank Munkhuu Ulzibat from RCAG, Ulaanbaatar, Mongolia, who has provided the regional event bulletin used for this study.

REFERENCES

- [1] C. Benítez, J. Ramírez, J. C. Segura, J. M. Ibáñez, J. Almendros, A. García-Yeguas, G. Cortés, Continuous HMM-Based Seismic-Event Classification at Deception Island, Antarctica, *IEEE Transactions on Geoscience and Remote Sensing*, 45:138-146, 2007.
- [2] C. M. Bishop. *Pattern Recognition and Machine Learning*, Springer, 2006.
- [3] N. Brachet, D. Brown, R. Le Bras, Y. Cansi, P. Mialle, J. Coyne, Monitoring the Earth's Atmosphere with the Global IMS Infrasound Network. In A. Le Pichon, E. Blanc and A. Hauchecorne, editors, *Infrasound Monitoring for Atmospheric Studies*, Springer, 2010.
- [4] Y. Cansi, An automatic seismic event processing for detection and location: The P.M.C.C. method, *Geophysical Research Letters*, 22:1021-1024, 1995.
- [5] S. Daskalaki, I. Kopanas, N. Avouris. Evaluation of classifiers for an uneven class distribution problem, *Applied Artificial Intelligence*, 20:381-417, 2006.
- [6] M. Ohrnberger, Continuous Automatic Classification of Seismic Signals of Volcanic Origin at Mt. Merapi, Java, Indonesia. Ph.D. thesis, Universität Potsdam, 2001.
- [7] L. R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE*, 77:257-286, 1989.
- [8] P. M. Shearer. *Introduction to Seismology*, Cambridge University Press, 1999.
- [9] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*, Wiley, 2001.