NORSAR Scientific Report No. 2-2012

Semiannual Technical Summary

1 July – 31 December 2012

Tormod Kværna (Ed.)

Kjeller, June 2013

6.5 Classifying Seismic Signals at Small–Aperture Arrays via Stochastic Modeling of F-K Image Sequences

6.5.1 Automatic seismic event classification systems

For any seismic monitoring task (volcano, geothermal, microseismic, …) it is desirable to detect and classify seismic events consistently (objectively and time-invariantly) and with little/no need for (costly) expert intervention. The seismic event categorization should be based on wave field properties including temporal structure and context. A similar problem in the realm of speech recognition applications is termed the word spotting problem. Thus, Ohrnberger (2001) transferred successfully hidden Markov model (HMM) techniques from the field of acoustic level speech recognition to tackle the seismic event spotting task applied to volcano-seismic signal observations. Since then, a number of HMM based classification approaches have been reported in seismology (*e.g.*, Beyreuther and Wassermann, 2008; Benitez *et al.*, 2007; Hammer *et al.*, 2012).

6.5.2 Seismic events casted as doubly stochastic process

HMMs are doubly stochastic processes that allow modeling of trajectories of observed sequences forming a seismic event (signal of interest) in a representative multidimensional space, which is referred to as feature space. In Fig. 6.5.1 (top) we show examples of trajectories for two distinct events in a two-dimensional feature space (*e.g.*, this may be amplitude and dominant frequency). Different signal properties result in clearly distinct paths traversing the feature space in time.

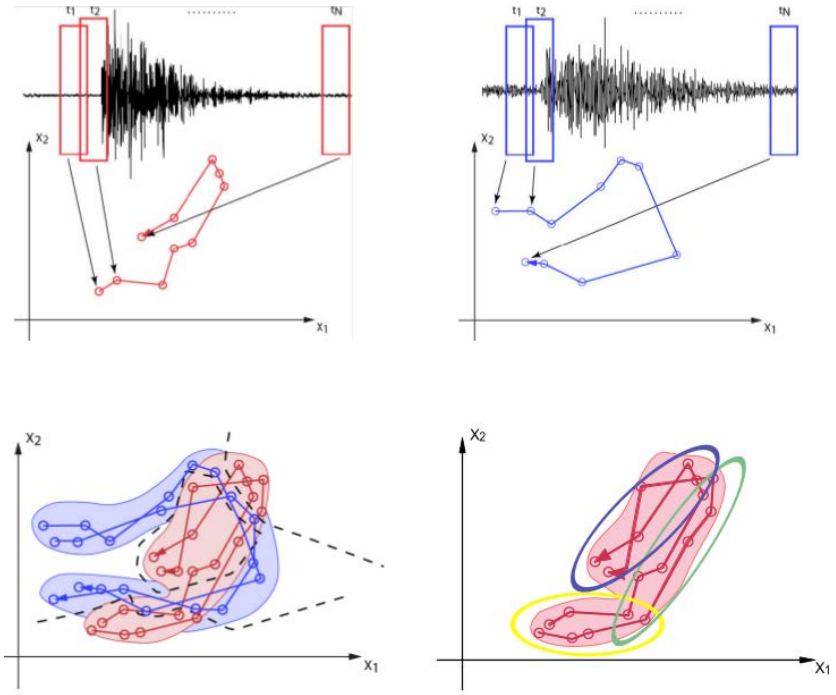

Fig. 6.5.1 Modeling trajectories of feature vector sequences as doubly stochastic processes.

When there are several examples for one event class that share similar behavior in feature space, similar trajectories will be observed. The distribution of observed trajectories for one and the same event type may then be described by a dynamically evolving probability density function of

observations (Fig. 6.5.1, bottom right). One particularly well suited stochastic model for describing such dynamically evolving probability density functions are Hidden Markov Models (HMM). Thus, the distribution of trajectories for any event class can be modeled as outcome of a HHM. The first underlying (hidden) process of a HMM creates a sequence of discrete states obeying Markov properties. At each time step, the system emits an observation drawn from a state-dependent multivariate Gaussian (see Fig. 6.5.2). For details about the underlying mathematics and the main standard computational issues about HMMs (*i.e.*, parameter learning) we refer here to the tutorial paper by Rabiner (1989), to fundamental work on the speech processing by Rabiner and Juang (1993) or the summary given in Ohrnberger (2001).

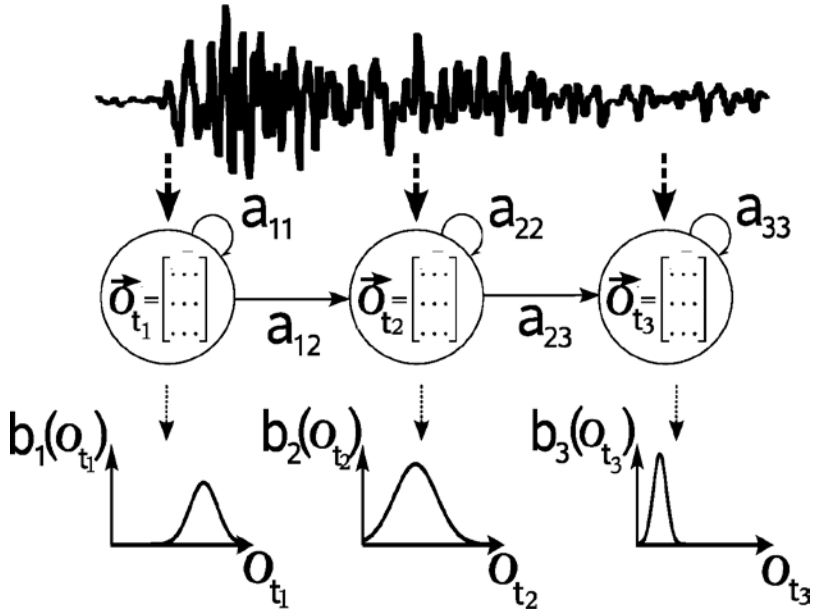

Fig. 6.5.2 HMM as a double stochastic process. The underlying state sequence of hidden state variables creatse an observation (seismic time series attributes) at each time step. The sequences of seismic time series attributes are then interpreted as having been created by a HMM. Model parameters are transition probabilities (a_{ij}) and output probability density functions (PDFs) related to a current state index (b_i).

The parameters of HMMs (state transition probabilities and output power-density-functions) are usually learned for each class of interest from a large training set (several tens of examples needed for each class). Here, we follow an alternative approach recently proposed by Hammer *et al.* (2012), which avoids the costly preparation of training sets. Indeed it is possible to start off with a single reference event by exploiting abundant information from unlabeled and therefore cheap training data (mostly continuous background/noise wave field information). The main idea for the proposed training procedure is to estimate with high confidence parameters of the probability density functions describing the general wave field characteristics in feature space (see Fig. 6.5.3). The parameters of the event class are then estimated from adjustments to these probability density functions based on a single/few reference event(s) for a specific class (see Fig. 6.5.4).

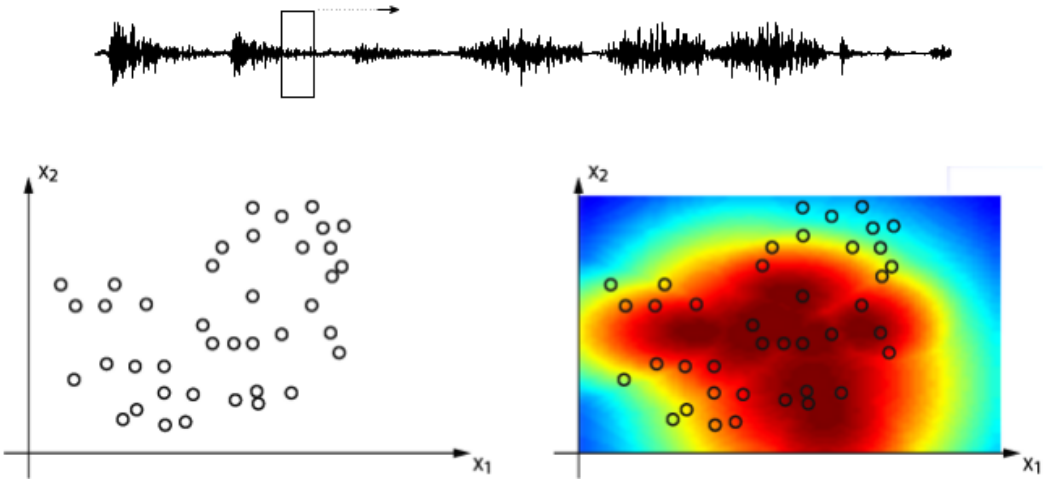

Fig. 6.5.3 Learning general background model from continuous data (unsupervised clustering from "cheap" training data).

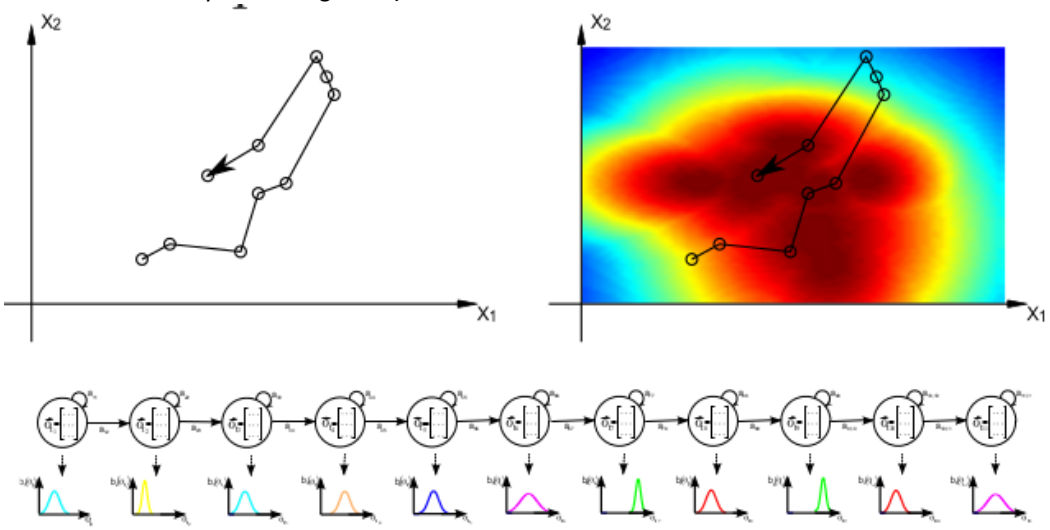

Fig. 6.5.4 Single reference waveform to update an event-specific classifier. The reference pattern is modeled as sequence of "visited" Gaussians.

6.5.3 Sliding window multi-broadband f-k array analysis as representative feature sequence

Seismic arrays are superior to single station observations allowing for determination of wave field direction and apparent speed along the earth surface. Three-component broadband arrays like SPITS (located on the Svalbard Archipelago, Norway) even enable a quite complete decomposition of the entire seismic wave field. We apply a sliding window f-k analysis in three overlapping, broad frequency bands to the vertical component recordings of SPITS. The example given in Fig. 6.5.5 shows that blurry f-k images indicate ”noise” or non-plane wave / multiple wave arrivals. Clear arrivals are easily detected by eye as the focused f-k image resembles the theoretical broadband array response pattern shifted to the wave group arrival's slowness vector position.

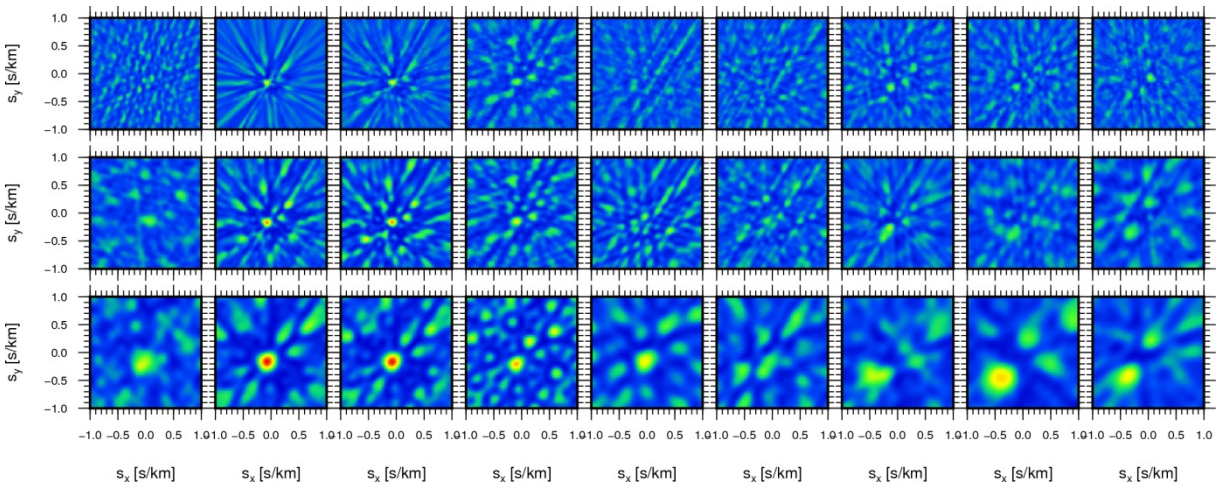

Fig. 6.5.5 Temporal sequence of f-k images computed in sliding window analysis for multiple broad frequency bands (from bottom to top row evaluated frequency bands are 1.5-4.5 Hz, 3-9 Hz, and 6-18 Hz; horizontal axis indicates time).

In order to encode the visual impression of an f-k image, we compute additionally the residual f-k image by subtracting the theoretical array response centered on the maximal coherent plane wave (see Fig. 6.5.6). Then we derive robust statistical parameters from broadband f-k and residual f-k images computed for bands 1.5 − 4.5 Hz, 3 − 9 Hz, and 6 − 18 Hz (101 × 101 grid in [-1, 1] s/km in 1.5 s windows and 0.1 s steps). In total we obtain a set of 51 parameters, including: μ, σ, rms, med, L1-scale, sx, sy, max/min-coherence, max-power. Certainly this selection is just one possible encoding of the f-k image information that would span a 6 x 101 x 101 = 61206 dimensional feature space in its raw representation.

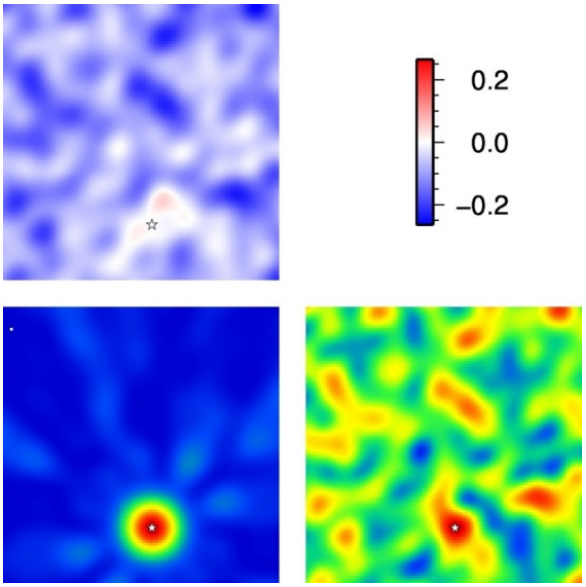

Fig. 6.5.6
Top left: Residual image resulting from subtracting the theoretical response from the observed f-k image. Bottom right: example of observed f-k image for a broad frequency band. Bottom left: Theoretical array response function computed for the same frequency band and centered on the most coherent plane wave arrival (maximum of the f-k image on the bottom right).

The 51-dimensional feature space ($f \in R^{51}$) has been examined visually. The discriminative power of features was judged when reviewing the temporal structure of feature series for selected events (Figs. 6.5.7 and 6.5.8).

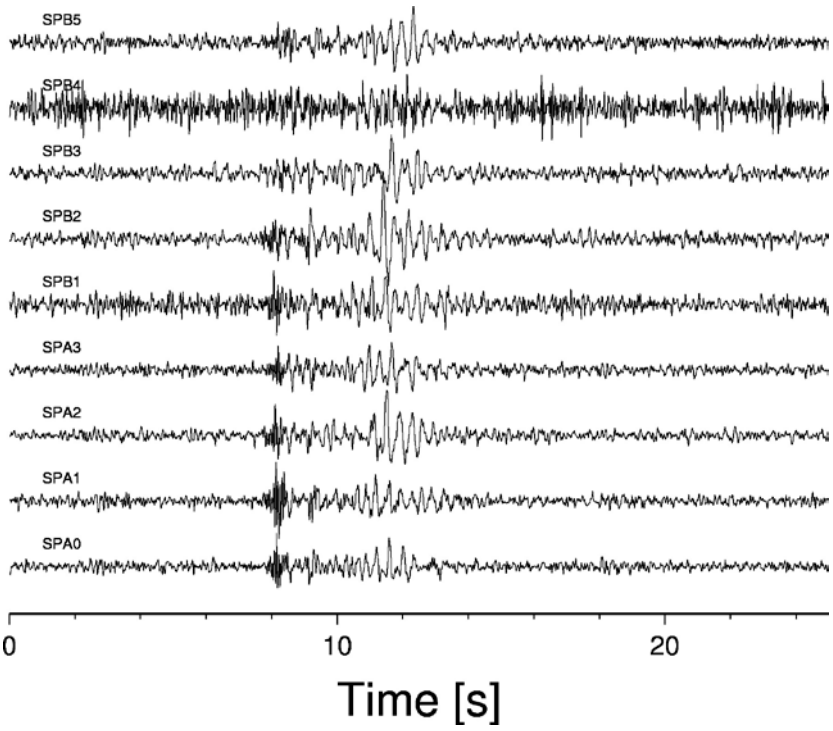

Fig. 6.5.7
The reference event observed at SPITS.

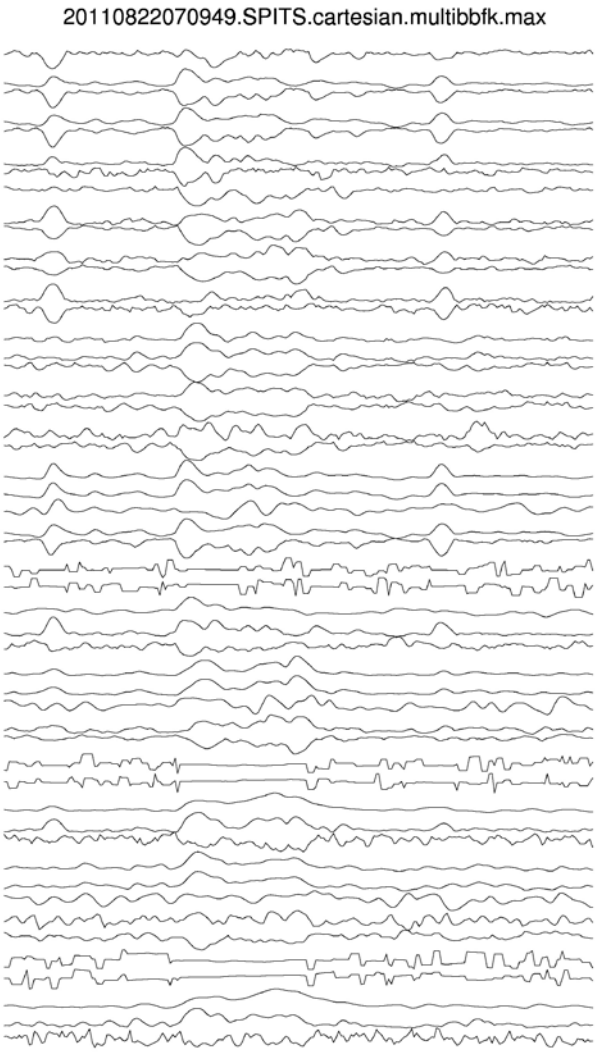

Fig. 6.5.8 Time series of 51 feature components for the reference event shown in Fig. 6.5.7.

Simple crossplots - omitting the temporal context - were used for finding correlated feature components (Fig. 6.5.9) for a large number of observations (6 hours of background data). Finally, we selected feature subsets with dimensions of 18, 21, 27, 33, and 39, based on the findings of the visual control and by seismological expertise. Those feature sets were then tested for the classifier design and the classification of continuous data.

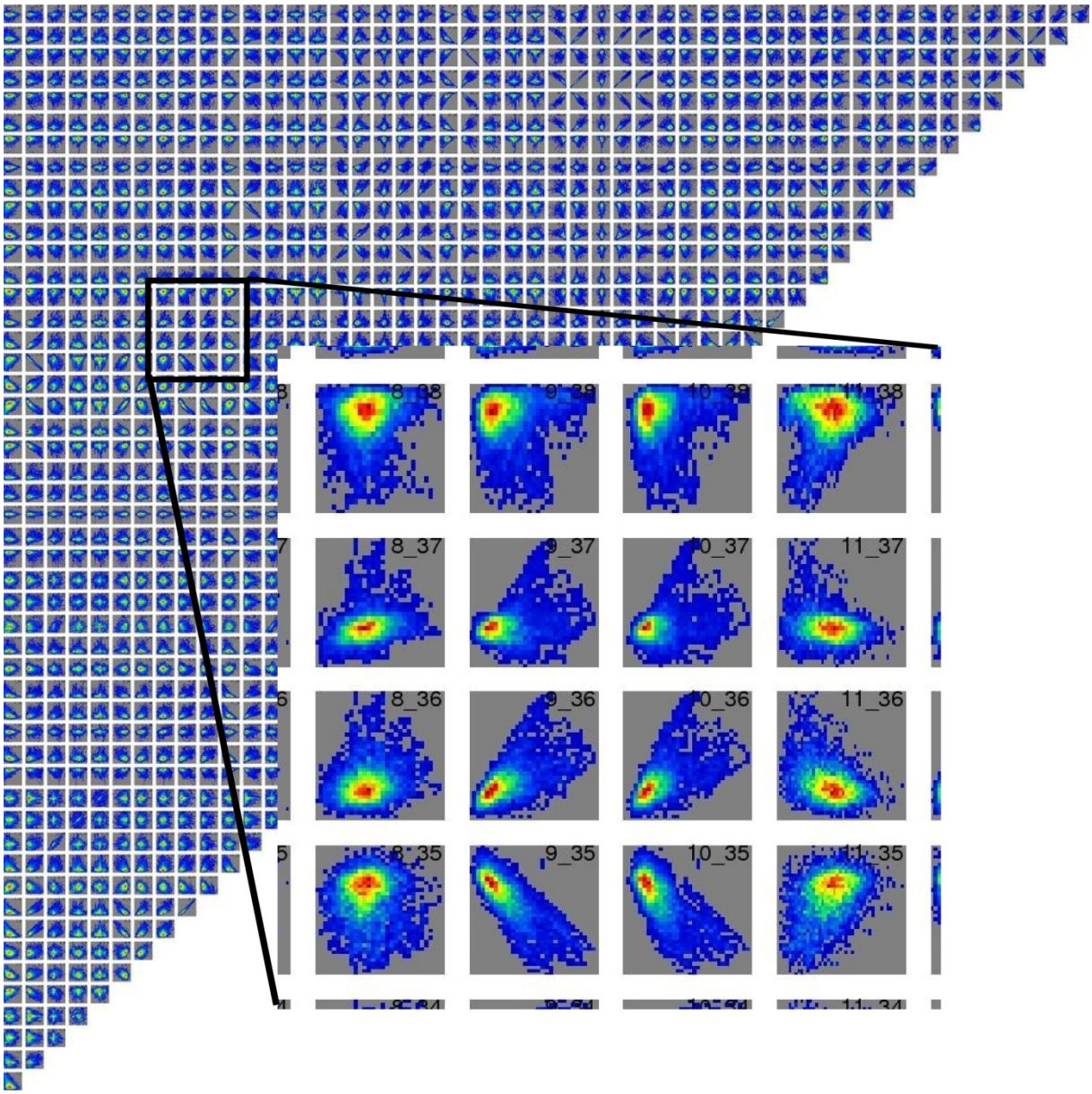

Fig. 6.5.9 Density crossplot of all pairs of feature components (see Fig. 6.5.8) for 6 hours of background data.

A large set of (cheap) unlabeled feature vectors are the ingredients for learning the background wave field properties as multivariate Gaussian mixture (Gaussian Mixture Model - GMM) in unsupervised fashion (k-means algorithm). Event HMM classifiers are then constructed on the base of the GMM using a single reference feature vector sequence for updating the HMM parameters (compare also Fig. 6.5.3 and 6.5.4). Both background and event models are finally integrated as a parallel grammar network (see Fig. 6.5.10). At each time step, a partial sequence of feature vectors is presented to the classifier network and the most likely hidden state path is decoded using the Viterbi-algorithm (Viterbi, 1967; Forney, 1973). Whenever the Viterbi-path is passing states related to the event HMM, a detection of this event-class is declared. Note that also multi-class systems can be constructed easily by adding more classifiers in parallel to the grammar network.

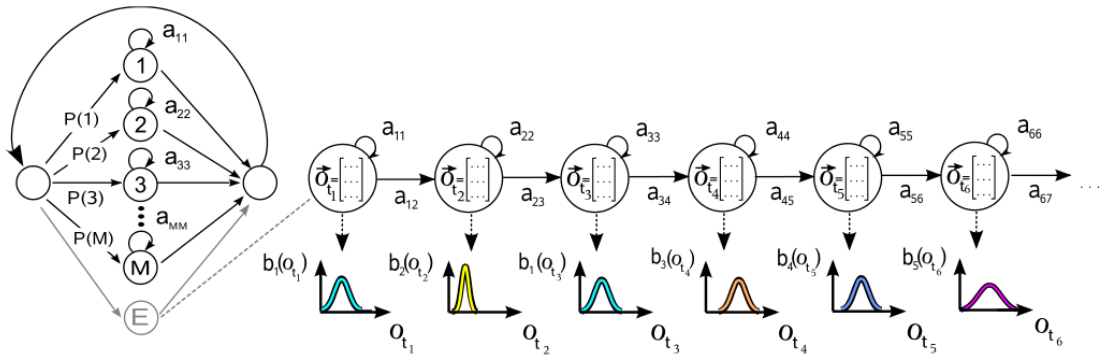

Fig. 6.5.10 Parallel grammar network in which each Gaussian mixture is part of the background model description and the event model HMM is connected in parallel.

An alternative approach for applying the HMM based detector to continuously recorded data is the use of running classifier and in parallel keeping track at each time step of the likelihood scores (Fig. 6.5.11). We may choose to use a noise only, or noise-event-noise grammar types and compare the average log likelihoods for the models in short time windows. This approach enables a judgment of the quality of detection by the log-likelihood difference as a confidence measure, but usually shows less accuracy in the segmentation of the event start and end times.

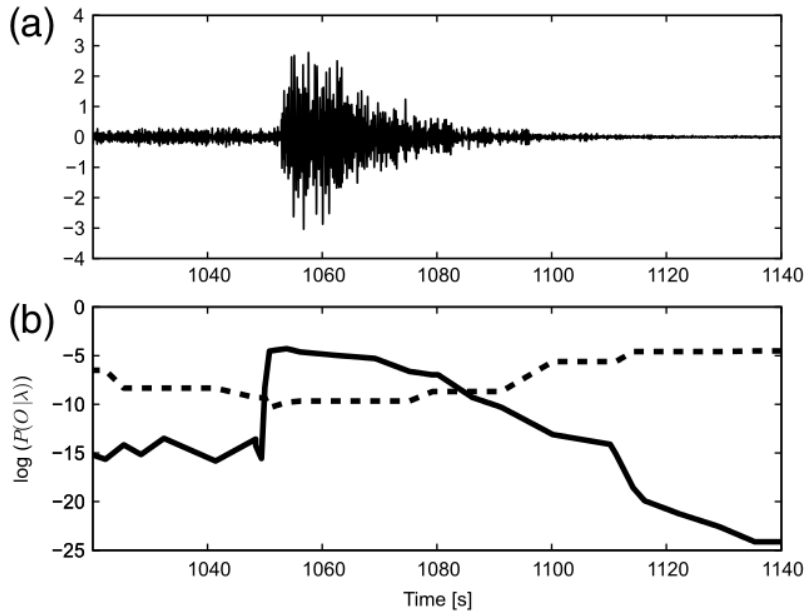

Fig. 6.5.11 Likelihood scoring of alternative HMM models (background and event models).

6.5.4 Test data at SPITS array and preliminary results

We use data from the SPITS array as a test environment to adapt the recently proposed classification technique by Hammer *et al.* (2012) for array settings. Our main goal is to:

- screen out events not interesting to analysts in NDC/CTBT context
- count glacier related events as a link to climatic change in arctic regions

We used 6 hours of continuous recordings for learning the background model (typical example of background activity shown in Fig. 6.5.12) and used a single record of an event which is assumed to be related to glacier activity (see Fig. 6.5.7). We expect glacier related events to show quite some variation in wave field appearance due to the variability of source location and complexity of source processes. We consider the HMM-based approach to be particularly suited for this task. One (rather typical) restriction for evaluating the quality of the classifier approach in a quantitative way is the lack of real ground truth for the observed seismicity.
There are only two tunable parameters for the classifier:

i) the number of Gaussians;
ii) feature vectors pre-whitened or not.

In combination with the distinct feature subsets there are then numerous settings to be considered. In Fig. 6.5.13 we show brief examples of classifier outcomes for the same waveform portion (only the vertical seismogram at the array site SPA0 is shown). Note the distinct scoring techniques (parallel classifiers vs. network).

6.5.5 Concluding remarks

- Preliminary classification results obtained for selected time windows allow the detection of a predefined event class (glacier related short and high frequency events).
- However, the comparison of results for distinct feature sets and parameters indicates sub-optimal performance. Spurious detections occur frequently.
- We conclude that there is a need for retraining and testing other selections of background training data.
- The final aim is to run the classifier for a longer time span of data with confidence. This ultimate goal requires ground truth data, which is at the moment lacking.

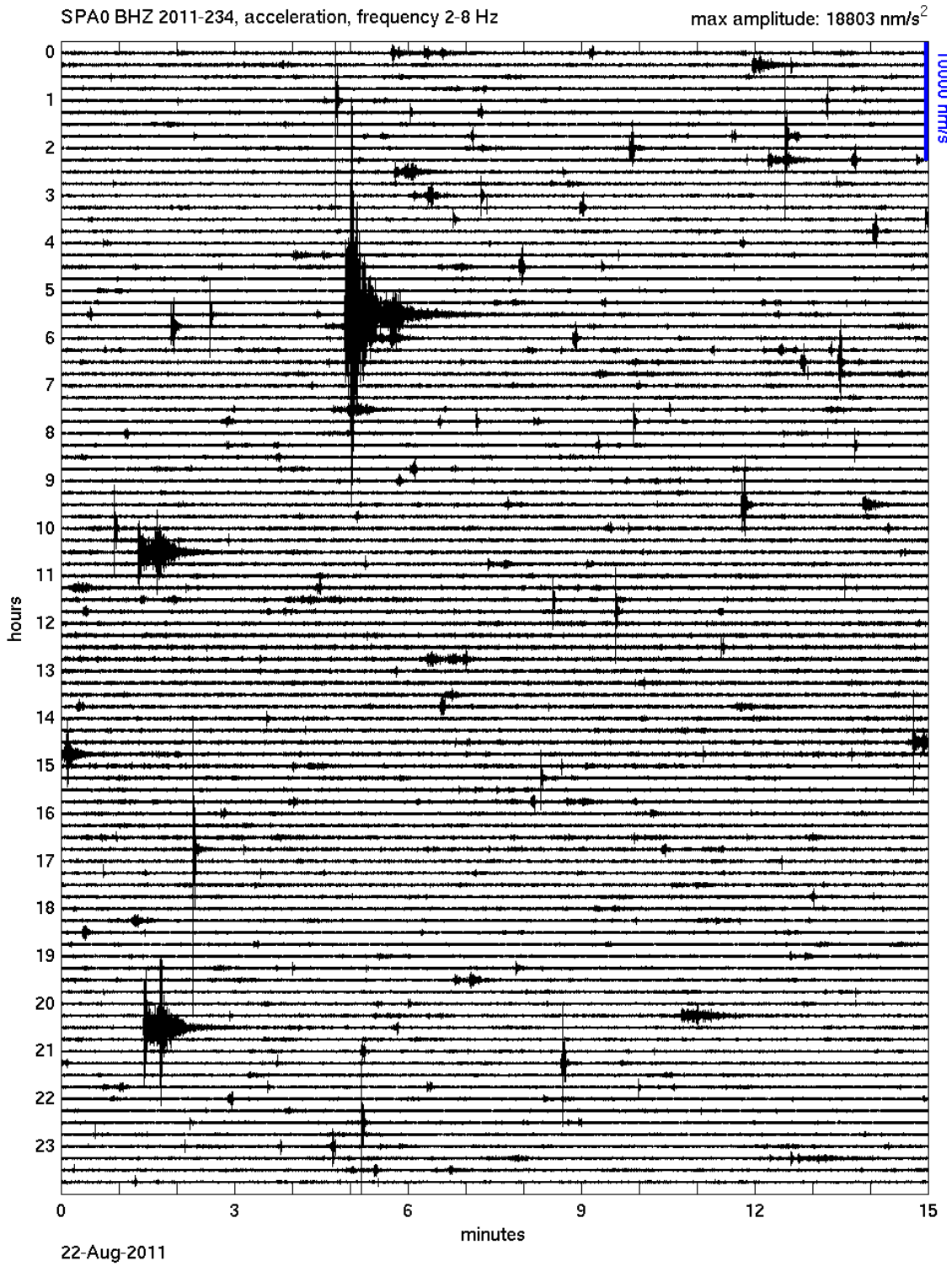

Fig. 6.5.12 A typical daily helicorder plot simulation for station SPA0 provides insight of the normal seismic activity levels at this arctic island region.

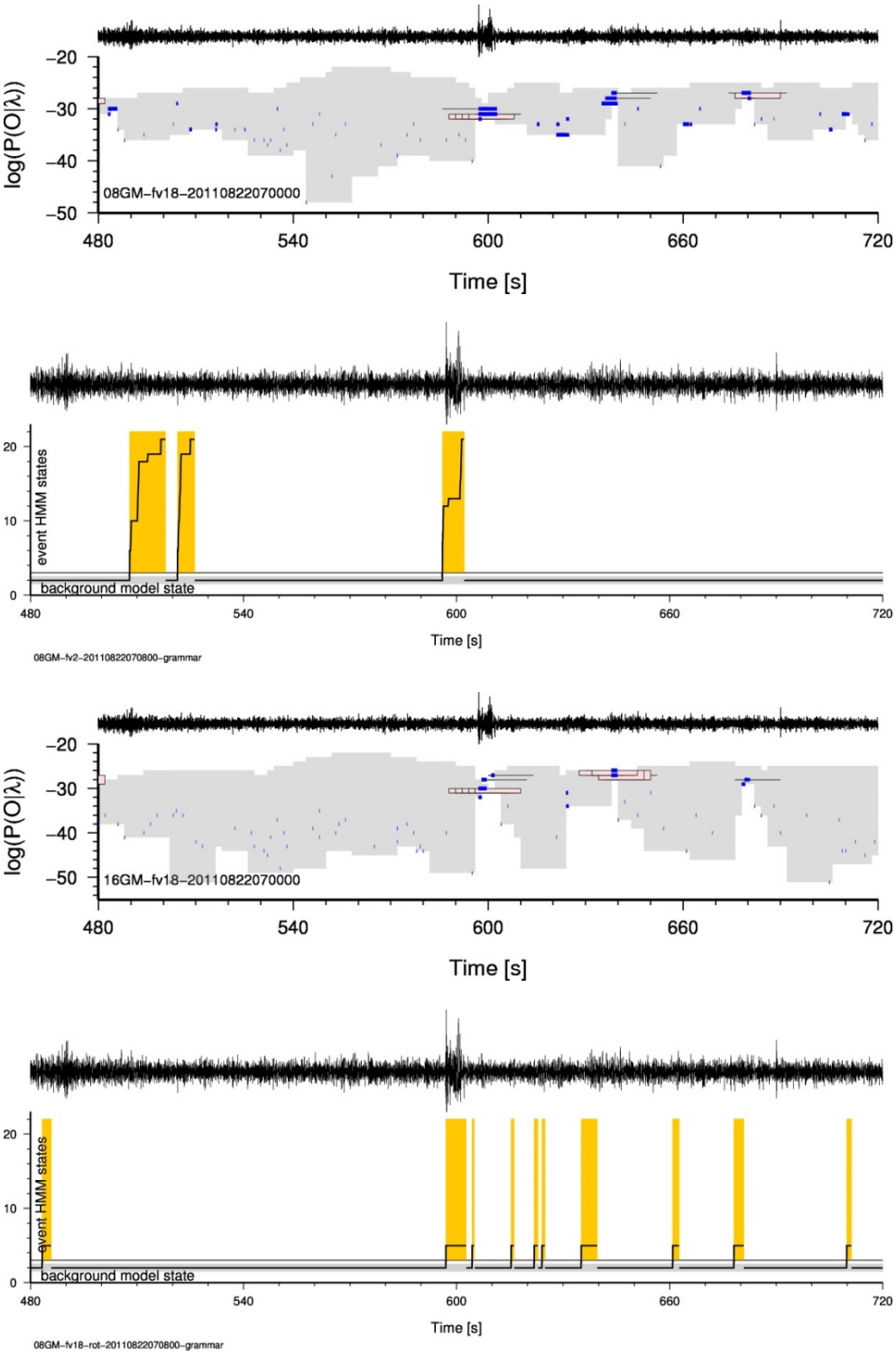

Fig. 6.5.13 Results of continuous classification (just a short example window for visualization) using either the grammar network or the relative likelihood scoring approach. Spurious detections are visible in both approaches and indicate not yet optimal classification performance.

Acknowledgements

This project has been initiated through a mobility program for researchers within the EU-project NERA (#262330) and provided funds for a one month stay of the first author at NORSAR.

Matthias Ohrnberger, University of Potsdam
Conny Hammer, University of Potsdam
Nikos Giannotis, University of Potsdam
Johannes Schweitzer

References

Benitez, M.C., J. Ramirez, J.C. Segura, J.M. Ibanez, J. Almendros, A. Garcia-Yeguas and G. Cortes (2007). Continuous HMM-based seismic-event classification at Deception Island, Antarctica. IEEE Trans. Geosciences Remote Sensing, 45(1), 138–146.

Beyreuther, M. and J. Wassermann (2008). Continuous earthquake detection and classification using discrete Hidden Markov Models. Geophys. J. Inter., 175(3), 1055–1066.

Forney, G. (1973). The Viterbi algorithm. Proc. IEEE, 61(3), 268 -278.

Hammer, C., M. Beyreuther and M. Ohrnberger (2012). A seismic event spotting system for volcano fast response systems. Bull. Seism. Soc. Amer., 102(3), 948-960.

Ohrnberger, M. (2001). Continuous automatic classification of seismic signals of volcanic origin at Mt. Merapi, Java, Indonesia. Ph.D. Thesis, Universität Potsdam.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE, 77(2), 257 -286.

Rabiner, L. and B. Juang (1993). Fundamentals of speech recognition. Prentice-Hall, XXXV+507 pp.

Viterbi, A. (1967). Error bounds for convolutional codes and asymptotically optimal decoding algorithm. IEEE Trans. Inf. Theory, IT-13(2), 260 -269.