

# A Time Series Analysis Method for Geophysical Sensor Networks

Robert Granat

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109

Email: granat@jpl.nasa.gov

**Abstract**—Modern sensor networks are collecting enormous volumes of observational information about geophysical systems. These sensor arrays, including GPS networks, ocean buoys, and seismic networks, are often designed to detect one particular phenomenon, but also capture signatures generated by a wide variety of other processes. We present a method for identifying, detecting, and cataloging these signals.

Our method is based on the use of hidden Markov models (HMMs). HMMs are statistical models for time series data; fitting an HMM allows use to describe the statistics of the data in a simple way that ascribes discrete modes of behavior to the system. Our HMM fitting algorithm, which uses a regularized deterministic annealing expectation-maximization (RDAEM) procedure, allows for far greater solution stability than other methods, even when the model is unconstrained. This has two benefits. First, the method can be used in situations in which the underlying system model is unavailable or unreliable. Second, the method can be used in applications where the stability and reliability of results is paramount.

This approach has been implemented in our software, called RDAHMM (Regularized Deterministic Annealing Hidden Markov Models), and integration with the SCIGN and SOPAC GPS networks through the QuakeSim project web portal environment is ongoing.

## I. INTRODUCTION

In the course of scientific investigation there are often situations in which the quantity of observational data greatly outpaces corresponding explanatory theory. In such scenarios statistical models become a valuable step in the research cycle. In this work we focus on the use of hidden Markov models (HMMs). HMMs are an attractive tool for the analysis of scientific data because many physical systems are time-ordered and exhibit evidence of discrete modes of behavior. For example, in seismology, the system undergoes distinct changes before and after an earthquake.

One challenge of HMM optimization is the problem of local maxima in the objective function, which can lead to divergent outcomes of the model fitting procedure. In application domains in which HMMs have previously enjoyed significant success, such as speech processing and recognition, the problem of local maxima is overcome primarily by the application of various constraints. In some cases, these

constraints are motivated by other concerns, such as ensuring robust estimates of output distributions, but nevertheless result in a reduction of the number of free parameters. These constraints include restrictions on the form of the state-to-state transition probability matrix [1]–[3], restrictions on the form of the output distributions [4], and parameter tying [5]–[7]. At their core, all of these approaches capitalize on the availability of extensive prior information about the underlying system. In our motivating scenario, such information is usually unavailable and so an alternative approach to learning the parameters of the HMM is required.

We address this problem by using the regularized deterministic annealing expectation-maximization (RDAEM) procedure [8] to perform the HMM fitting. This method constructs high-quality, self-consistent model fits without using a priori information (although it does not exclude the use of such information where available), at the cost of some additional computation time. By building on this basic approach, we can provide a number of analysis capabilities for time series data. These capabilities include:

- **Mode Segmentation:** a time series containing unknown activity is segmented into statistically significant behavior modes.
- **Sensor Network Change Detection:** mode segmentation analysis is applied to each sensor in a network; when modes change simultaneously across the entire network or within a significant sub-network, an alert is triggered.
- **Signal Search:** a user identifies a signal of interest in streaming or stored data, and matches to this data ranked according to relevance are retrieved from stored data.

We begin in Section II with a description of the RDAEM algorithm for fitting hidden Markov models (HMMs) and provide an example of its benefits as compared to standard methods on a sample GPS data set. In the subsequent sections we describe in more detail the three capabilities discussed above.

## II. FITTING HIDDEN MARKOV MODELS

We start our description of the RDAEM algorithm for fitting HMMs by establishing some notation: a hidden Markov model  $\lambda$  with  $N$  states is composed of a vector of initial state probabilities  $\pi = (\pi_1, \dots, \pi_N)$ , a matrix of state-to-state transition probabilities  $A = (a_{11}, \dots, a_{ij}, \dots, a_{NN})$ , and the observable output probability distributions  $B = (b_1, \dots, b_N)$ . The observable outputs can be either discrete or continuous.

Here we are concerned with continuous valued outputs with probability distributions denoted by  $b_i(y, \theta_i)$  where  $y$  is the real-valued observable output (scalar or vector) and the  $\theta_i$ s are the parameters describing the output probability distribution. For the normal distribution we have  $b_i(y, \mu_i, \Sigma_i)$ . An observation sequence  $O$  of length  $T$  is denoted  $O_1 O_2 \cdots O_T$  and a state sequence  $Q$  of the model is denoted  $q_1 q_2 \cdots q_T$ .

#### A. Deterministic Annealing

Deterministic annealing is a technique inspired by statistical mechanics, that can be used to reduce the sensitivity to initial conditions of the well-known expectation-maximization (EM) method. Deterministic annealing uses the principle of maximum entropy to specify an alternative posterior probability density for the hidden variables, allowing us to define a new effective cost function depending on a temperature parameter. This new cost function is analogous to the thermodynamic free energy. Maximization of the likelihood of the model at a given temperature is achieved via minimization of this cost function. Deterministic annealing attempts to avoid the standard problems associated with simulated annealing [9], such as relaxation from initial conditions and false local minima in the free energy, through deterministic optimization of the cost function at each temperature. In theory, it provides similarly global solutions with less variance, while reducing average computational cost.

We can apply the deterministic annealing method to HMM optimization as follows (for more details, see [8], [10], and [11]): at each computational temperature we follow an EM-like procedure to minimize the free energy; at the  $k$ th iteration of we optimize over the function

$$U(\gamma, \lambda | \lambda^{(k)}) = \sum_{i=1}^N \tau_{i1}^{(k)}(\gamma) \log \pi_i + \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{T-1} \tau_{ijt}^{(k)}(\gamma) \log a_{ij} + \sum_{i=1}^N \sum_{t=1}^T \tau_{it}^{(k)}(\gamma) \log b_i(O_t), \quad (1)$$

where  $1/\gamma$  is the computational temperature and

$$\tau_{it}(\gamma) = \frac{\alpha_t(i, \gamma) \beta_t(i, \gamma)}{\sum_{i=1}^N \alpha_t(i, \gamma) \beta_t(i, \gamma)}, \quad (2)$$

and

$$\tau_{ijt}(\gamma) = \frac{\alpha_t(i, \gamma) a_{ij}^\gamma b_j^\gamma(O_{t+1}) \beta_{t+1}(j, \gamma)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i, \gamma) a_{ij}^\gamma b_j^\gamma(O_{t+1}) \beta_{t+1}(j, \gamma)}. \quad (3)$$

The modified forward and backward variables  $\alpha(\gamma)$  and  $\beta(\gamma)$  are calculated by a modification of the standard iterative procedure:

##### 1) Initialization:

$$\alpha_1(i, \gamma) = \pi_i^\gamma b_i^\gamma(O_1), \quad i = 1, \dots, N. \quad (4)$$

$$\beta_T(i, \gamma) = 1, \quad i = 1, \dots, N. \quad (5)$$

##### 2) Induction:

$$\alpha_{t+1}(j, \gamma) = \left[ \sum_{i=1}^N \alpha_t(i, \gamma) a_{ij}^\gamma \right] b_j^\gamma(O_{t+1}), \quad t = 1, \dots, T-1, \quad j = 1, \dots, N. \quad (6)$$

$$\beta_t(i, \gamma) = \sum_{j=1}^N a_{ij}^\gamma b_j^\gamma(O_{t+1}) \beta_{t+1}(j, \gamma), \quad t = T-1, \dots, 1, \quad i = 1, \dots, N. \quad (7)$$

#### B. Regularized Deterministic Annealing

We use a regularization approach to improve upon the optimization results provided by the deterministic annealing EM (DAEM) algorithm alone. Regularization terms, which are added to the so-called  $Q$ -function maximized during the M-step of the EM algorithm, can be cast as statistical priors that modify the HMM objective function. Priors used in previous work have included Dirichlet-type priors used to prevent over-training of discrete output distributions [12] and maximum entropy priors used for pruning the HMM structure [13]. In addition, statistical priors have been used more generally to provide upper bounds on the objective function of models with continuous output distributions. For instance, the conjugate prior for Gaussian output distributions described by Ormonet and Tresp [14] wards against solutions with infinitely narrow Gaussian outputs.

Our regularized DAEM method is designed to discourage HMM local optima in which multiple HMM states have the same output distributions. These solutions occur quite often in practice, not only when the method is applied to HMMs both also for other model types [15]. In order to address this problem, we chose an improper prior on the HMM likelihood based on the squared Euclidean distance between the means of the state output distributions:

$$P(B) = \prod_{i=1}^N \prod_{j=1}^N \exp\left(\frac{\omega_{Q_3}}{2} (\mu_i - \mu_j)^T (\mu_i - \mu_j)\right), \quad (8)$$

where  $\omega_{Q_3} > 0$  is a weighting term. This prior rewards solutions with widely separated output distributions. Recall that for an HMM, the  $Q$ -function is

$$Q(\lambda, \lambda^{(k)}) = \sum_{i=1}^N \tau_{i1}^{(k)} \log \pi_i + \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{T-1} \tau_{ijt}^{(k)} \log a_{ij} + \sum_{i=1}^N \sum_{t=1}^T \tau_{it}^{(k)} \log b_i(O_t).$$

Since this is separable in  $\pi$ ,  $A$ , and  $B$ , we can divide it into the sum of three functions:  $Q_1(\pi)$ ,  $Q_2(A)$ , and  $Q_3(B)$ . From

our prior we derive the modified  $Q$ -function with

$$Q'_3 = \sum_{i=1}^N \sum_{t=1}^T \tau_{it}^{(k)} \left( \log n - \frac{1}{2} \log \det(\Sigma_i) - \frac{1}{2} (m_i - \mu_i)^T \Sigma_i^{-1} (m_i - \mu_i) - \frac{1}{2} (O_t - m_i)^T \Sigma_i^{-1} (O_t - m_i) + \frac{\omega_{Q_3}}{2} \sum_{j=1}^N (\mu_i - \mu_j)^T (\mu_i - \mu_j) \right) \quad (9)$$

where  $m_i = \sum_{t=1}^T \tau_{it}^{(k)} O_t / \sum_{t=1}^T \tau_{it}^{(k)}$ . Note that for the ease of subsequent manipulation and computation, we do not properly account for the independence of the prior from the hidden state variable. In theory, systems with highly skewed populations of observations from each of the output distributions may be misrepresented in this regularization scheme. In practice, we have observed no evidence of any serious consequence; nevertheless we keep in mind that the regularized  $Q$ -function (9) is only an approximation.

To derive the revised EM update rule we first take the vector derivative of  $Q'_3$  in the means:

$$\frac{\partial Q'_3}{\partial \mu_i^T} = \Sigma_i^{-1} (m_i - \mu_i) - \omega_{Q_3} \sum_{j=1}^N \mu_j + N \omega_{Q_3} \mu_i. \quad (10)$$

Setting the derivative to zero we have

$$\Sigma_i^{-1} m_i + (N \omega_{Q_3} I - \Sigma_i^{-1}) \mu_i = \omega_{Q_3} \sum_{j=1}^N \mu_j, \quad (11)$$

for  $i = 1, \dots, N$ , which leads us to the system of equations

$$\left( \begin{array}{c} \left[ \begin{array}{ccc} \Sigma_1^{-1} & & \\ & \ddots & \\ & & \Sigma_N^{-1} \end{array} \right] \\ + \omega_{Q_3} \left[ \begin{array}{ccc} I_{D \times D} & \cdots & I_{D \times D} \\ \vdots & \ddots & \vdots \\ I_{D \times D} & \cdots & I_{D \times D} \end{array} \right] U \\ - N \omega_{Q_3} U = \\ \left[ \begin{array}{ccc} \Sigma_1^{-1} & & \\ & \ddots & \\ & & \Sigma_N^{-1} \end{array} \right] M, \end{array} \right) \quad (12)$$

where  $M = [m_1; \dots; m_N]$ ,  $U = [\mu_1; \dots; \mu_N]$ , and  $I_{D \times D}$  is a  $D \times D$  identity matrix. This system can be solved by any standard linear method given the inverse covariances  $\Sigma_i^{-1}$ . In order for the problem to remain concave, and thus have a global maxima in the means, we require

$$\omega_{Q_3} \leq \|\Sigma_i^{-1}\| / 2N. \quad (13)$$

To find the maximum in the covariances, we take the derivatives of (9) in the components of  $\Sigma_i^{-1}$  and set them

equal to zero, which gives us

$$\Sigma_i = \sum_{t=1}^T \tau_{it}^{(k)} (O_t - \mu_i)(O_t - \mu_i)^T / \sum_{t=1}^T \tau_{it}^{(k)}. \quad (14)$$

This is a global maximum since the  $Q$ -function is concave as a function of the covariances  $\Sigma_i$ . To find the means and covariances we need to solve equations (12) and (14) simultaneously. This can be done by using the approximation  $\Sigma_i = S_i$ , where

$$S_i = \sum_{t=1}^T \tau_{it}^{(k)} (O_t - m_i)(O_t - m_i)^T / \sum_{t=1}^T \tau_{it}^{(k)}, \quad (15)$$

in equation (12) as an initial guess and then iterating between equations (12) and (14) until the solution converges. In practice, it is usually sufficient merely to approximate  $\Sigma_i$  as  $S_i$  when calculating the means without any attempt at iterative convergence whatsoever. This iterative rather than direct maximization forces us to characterize the method as a generalized EM algorithm rather than a pure EM approach.

### III. MODE SEGMENTATION

The first and most straightforward application of hidden Markov models is to perform mode segmentation of time series. Given the fitted model parameters, we can determine the individually most likely state assignments of the observation sequence,

$$q_t = \operatorname{argmax}_{1 \leq i \leq N} (\tau_{it}), \quad t = 1, \dots, T \quad (16)$$

The resulting sequence of state assignments indicates the modes of the system through time. As an example of this sort of analysis, we present the results of fitting an HMM to a time series collected by a GPS station in Claremont, California.

This time series consists of relative displacement measurements in three dimensions (north-south, east-west, and vertical) collected daily over about two years spanning 1998-1999. In Figure 1 we show this time series, with observations color coded according to the state assignments provided by fitting a seven state HMM to the data.

This data set contains certain clear signals of deformation processes; these include a vertical dip and rebound around days 100-200 resulting from the pumping and subsequent refilling of a local aquifer, as well as a sudden east-west jump caused by the magnitude 7.1 Hector Mine earthquake in 1999. We see that the fitting method has identified not only these mode changes but a number of more subtle signals, including periods of increased noise in the beginning and end of the series.

This mode segmentation analysis tool has been integrated into the QuakeSim web portal [16] and linked via web services to both archived and streaming GPS data. In this environment, analysis can be performed not only on daily integrated solution data, as in the example above, but also on real-time position information collected at 1Hz.

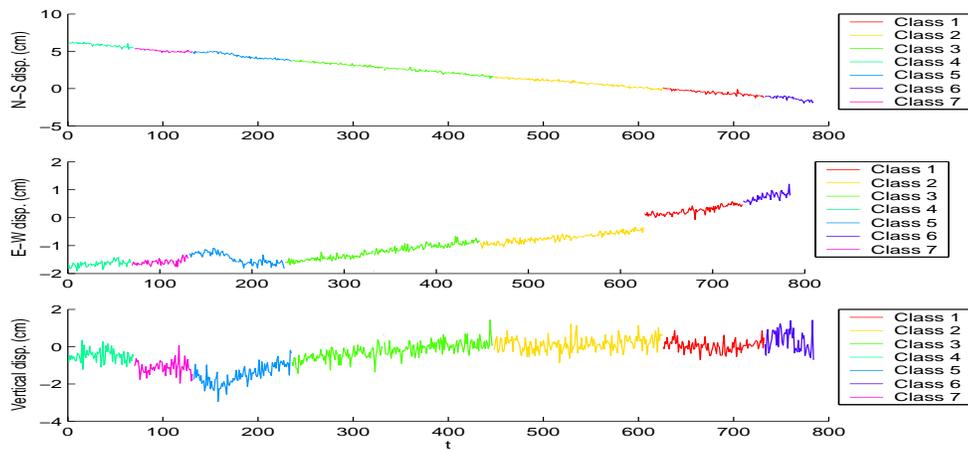


Fig. 1. Classification results for a seven-state HMM applied to GPS displacement data from Claremont, California. Automatically detected mode changes include a period of aquifer pumping (blue) and a large jump caused by the Hector Mine earthquake (yellow to red transition).

#### IV. SENSOR NETWORK CHANGE DETECTION

In this section we extend the techniques used to analyze data from a single sensor to the analysis of entire networks or subnetworks. We present a case study in which we are interested in detecting geophysical events with geographically disperse signatures using GPS data, including not only earthquakes but also aseismic events linked to crustal block motion or stress transfer between earthquake faults. These types of events have been observed in a few instances [17]–[23], but detections remain rare due to the subtlety of the signals. We hope to observe evidence of these types of aseismic events in the GPS data.

For this experiment we used daily displacement solutions for all 127 available SCIGN stations in a 820 day window beginning Jan 1st, 1998. When GPS displacement values for a given station were not available on a particular day due to signal dropout or incomplete installation, we assumed a zero displacement measurement for that day. We note that since actual measurements are almost never of zero displacement, this in effect adds an additional “dropout” class to the data. Our next step was to fit a separate hidden Markov model to each of these GPS signals. Once these models were trained, they provided us with the state sequences for each GPS time series. We suspected that interesting geophysical events would manifest themselves as changes in the signals across multiple GPS stations, thus we looked for correlations in state changes across the network.

Figure 2 shows the number of same-day state changes across all stations using classifications given by six-state models. There are a number of strong peaks indicating correlated state changes; of note is the strong peak on day 652, which corresponds to the Hector Mine earthquake. We observe that there is an increasing trend in the average number of coincident state transitions; this is because of the increasing number of stations installed and activated during the observation period. When we compare these network change correlations with the seismic record, we find that the only large event during that

time period was the Hector Mine earthquake; other correlation peaks are not associated with strong seismic events, and thus indicate possible aseismic events.

#### V. SIGNAL SEARCH

By providing high-quality, reliable HMM fitting results, the RDAEM method also allows us to perform time series search operations. These operations allow a human investigator to identify a signal of interest and search over a data base of time series to find matches to that signal. This capability allows the user to easily find and catalog multiple instances of an unusual signal in a large volume of data on the basis of a single human observation; a bootstrapping approach seems particularly appropriate in such a scenario. If the database supports annotation, new investigators can easily access the observations of previous researchers about similar signals that were previously observed.

Our approach to performing time series search was as follows. In the first step, the user selects a particular time series snippet, either from a stream or database, and an HMM is fit to that snippet. Once the model was trained on the selected data, it is then used to classify the selected observations into sequence of discrete states as described in Section III. That same HMM is then used to similarly classify time series in the database. The state sequence pattern of the search snippet is matched against all possible state sequence patterns of equal length in the database, and a Hamming distance metric is used to rank matches. The top results are then returned to the user.

An screenshot of this time series search procedure is shown in Figure 3. In this example, the method is being applied to sensor data collected from the International Space Station. A strip chart displays various sensor values to the flight controller (figure upper left), from which the flight controller can select a signal of interest (in this case, a saw-toothed pattern in the beta gimbal array motor current). Signal matches are shown on the right.

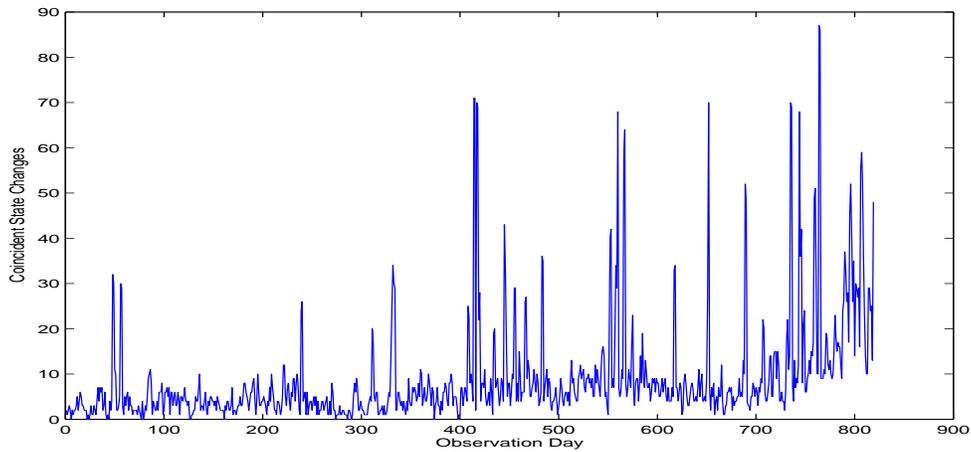


Fig. 2. Coincident state changes for six-state HMMs trained on daily position solutions from each of 127 SCIGN GPS stations.

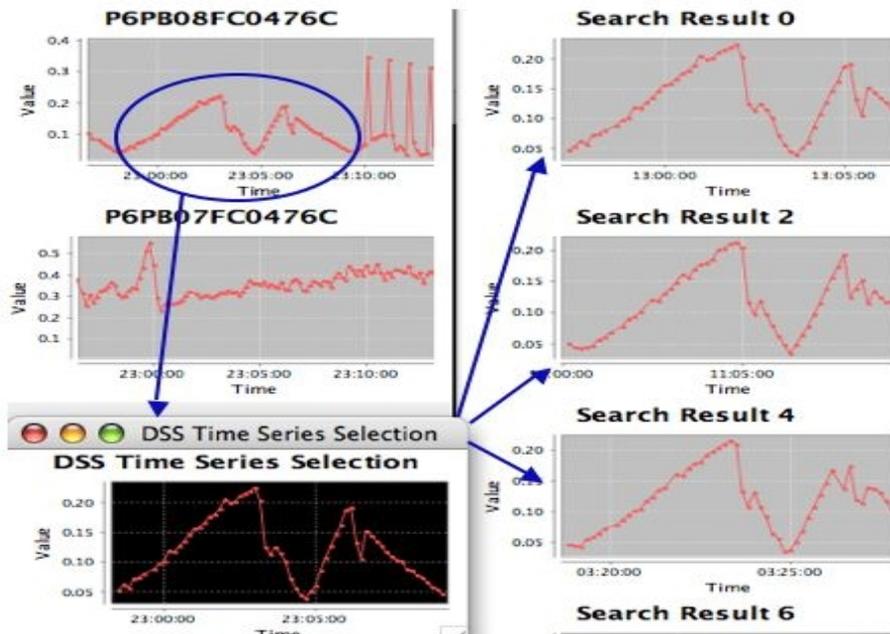


Fig. 3. A screenshot of a time series signal search application, implemented for use with sensor data from the International Space Station. The flight controller selects a signal of interest from one of a number of sensors displayed in strip charts (left), to which is fit an HMM model. The model is used to find matches in the database (right). Some results are off screen.

## VI. CONCLUSIONS

We have presented a technology based on hidden Markov models (HMMs) that provides a number of capabilities for the analysis of sensor web time series data, including segmentation, detection of network-wide phenomena, and time series signal search. These capabilities have been made possible by use of the regularized deterministic annealing expectation maximization (RDAEM) algorithm for HMM fitting. This algorithm provides robust, reliable model fits even in the absence of a priori information about the system.

This algorithm has been implemented and integrated into both the QuakeSim web portal environment for analysis of

geophysics data and a Decision Support System for NASA flight controllers. In the former case, an interactive GUI enables exploratory data analysis via the mode segmentation technology, and network change detection capability enables detection of regional events, including not only earthquakes but also atmospheric and aseismic events. In the latter case, the signal search capability allows flight controllers to select signals of ongoing events and recover records of similar activity that occurred previously. This enables the quick assessment of events based on previous experience as well as rapid recovery of records of relevant corrective actions taken by flight controllers in similar situations in the past.

We envision a wide variety of applications of this technology, including analysis of seismic networks, ocean buoy data analysis, analysis of image sequences via summary statistics such as vegetation indices, and analysis of flight sensor data from aircraft and spacecraft.

## REFERENCES

- [1] B.H. Juang and L.R. Rabiner, "Mixture autoregressive hidden Markov models for speech signals," *IEEE Transactions On Acoustics Speech And Signal Processing*, vol. 33, no. 6, pp. 1404–1413, 1985.
- [2] A. Farago and G. Lugosi, "An algorithm to find the global optimum of left-to-right hidden Markov model parameters," *Problems Of Control And Information Theory-Problemy Upravleniya I Teorii Informatsii*, vol. 18, no. 6, pp. 435–444, 1989.
- [3] G. McGuire, F. Wright, and M.J. Prentice, "A Bayesian model for detecting past recombination events in DNA multiple alignments," *Journal Of Computational Biology*, vol. 7, no. 1-2, pp. 159–170, 2000.
- [4] Y. Ephraim, A. Dembo, and L.R. Rabiner, "A minimum discrimination information approach for hidden Markov modeling," *IEEE Transactions On Information Theory*, vol. 35, no. 5, pp. 1001–1013, 1989.
- [5] J.R. Bellegarda and L.R. Nahamoo, "Tied mixture continuous parameter modeling for speech recognition," *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, vol. 38, no. 12, pp. 2033–2045, 1990.
- [6] S.J. Young and P.C. Woodland, "State clustering in hidden Markov model-based continuous speech recognition," *Computer Speech And Language*, vol. 8, no. 4, pp. 369–383, 1994.
- [7] E. Bocchieri and B.K.W. Mak, "Subspace distribution clustering hidden Markov model," *IEEE Trans. on Speech and Audio Proc.*, vol. 9, no. 3, pp. 264–275, 2001.
- [8] R. A. Granat, *Regularized Deterministic Annealing EM for Hidden Markov Models*, Ph.D. thesis, University of California, Los Angeles, 2004.
- [9] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
- [10] K. Rose and A. V. Rao, "Deterministically annealed design of hidden Markov model speech recognizers," *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 2, pp. 111–126, 2001.
- [11] R. Granat and A. Donnellan, "Deterministic annealing hidden Markov models for geophysical data exploration," *Proc. 3rd Workshop on Mining Scientific Data*, pp. 1–8, 2001.
- [12] A. Krogh, M. Brown, I.S. Mian, K. Sjolander, and D. Haussler, "Hidden Markov models in computational biology - applications to protein modeling," *Journal Of Molecular Biology*, vol. 235, no. 5, pp. 1501–1531, 1994.
- [13] M. Brand, "Structure learning in conditional probability models via an entropic prior and parameter extinction," *Neural Computation*, vol. 11, pp. 1155–1182, 1999.
- [14] D. Ormonet and V. Tresp, "Improved Gaussian mixture density estimates using Bayesian penalty terms and network averaging," *Advances in Neural Information Processing Systems 8*, pp. 542–548, 1996.
- [15] M. Whitley and D. M. Titterton, "Applying the deterministic annealing expectation maximization algorithm to naive Bayesian networks," *Univ. Glasgow Tech. Report*, 2002.
- [16] A Donnellan, J Rundle, G Fox, D McLeod, L Grant, T Tullis, M Pierce, J Parker, G Lyzenga, R Granat, and M Glasscoe, "QuakeSim and the Solid Earth Research Virtual Observatory," *PAGEOPH*, vol. 163, pp. 1–17, 2006.
- [17] T.I. Melbourne and F.H. Webb, "Slow but not quite silent," *Science*, vol. 300, no. 5627, pp. 1886–1887, 2003.
- [18] G. Rogers and H. Dragert, "Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip," *Science*, vol. 300, no. 5627, pp. 1942–1943, 2003.
- [19] T.I. Melbourne and F.H. Webb, "Precursory transient slip during the 2001  $m_w=8.4$  Peru earthquake sequence from continuous gps," *Geophysical Research Letters*, vol. 29, no. 21, pp. art. no.–2032, 2002.
- [20] T.I. Melbourne, F.H. Webb, J.M. Stock, and C. Reigber, "Rapid post-seismic transients in subduction zones from continuous gps," *Journal Of Geophysical Research – Solid Earth*, vol. 107, no. B10, pp. art. no.–2241, 2002.
- [21] M.M. Miller, T. Melbourne, D.J. Johnson, and W.Q. Sumner, "Periodic slow earthquakes from the cascadia subduction zone," *Science*, vol. 295, no. 5564, pp. 2423–2423, 2002.
- [22] H. Hirose, K. Hirahara, F. Kimata, N. Fujii, and S. Miyazaki, "A slow thrust slip event following the two 1996 Hyuganada earthquakes beneath the Bungo Channel, southwest Japan," *Geophysical Research Letters*, vol. 26, no. 21, pp. 3237–3240, 1999.
- [23] K. Heki, S. Miyazaki, and H. Tsuji, "Silent fault slip following an interplate thrust earthquake at the Japan Trench," *Nature*, vol. 386, no. 6625, pp. 595–598, 1997.