GRAPH CLUSTERING FOR LOCALIZATION WITHIN A SENSOR ARRAY

Nima Riahi and Peter Gerstoft

Scripps Institution of Oceanography,
University of California San Diego,
La Jolla, CA 92093

Christoph F. Mecklenbräuker

Institute of Telecommunications,
Vienna University of Technology
1040 Vienna, Austria

ABSTRACT

We develop a model-free technique to identify weak sources within dense sensor arrays using graph clustering. No knowledge about the propagation medium is needed except that signal strengths decay to insignificant levels within a scale that is shorter than the aperture. We then reinterpret the spatial coherence matrix of a wave field as a matrix whose support is a connectivity matrix of a graph with sensors as vertices. In a dense network, well-separated sources induce clusters in this graph. The support of the covariance matrix is estimated from limited-time data using a hypothesis test with a robust phase-only coherence test statistic combined with a physical distance criterion. The method is applied to a dense 5200 element geophone array that blanketed 7 km $\times$10 km of the city of Long Beach (CA). The analysis exposes a helicopter traversing the array.

1. INTRODUCTION

Large and dense sensor arrays are becoming more common as the cost for sensor and communications hardware decreases. Examples of such arrays are the USArray initiative in seismology with 500 stations covering large parts of the continental US or the seismic exploration array with 5200 sensors as presented here. As array sizes increase the occurrence of within-aperture source events that cause coherent signals over only a small fraction of all sensors becomes more common.

This work addresses the problem of localizing such weak sources in a complex and unknown environment. For known or well characterized media this problem has been addressed using frameworks such as matched field processing (MFP), maximum likelihood methods, or migration techniques in, e.g., acoustics [1, 2, 3], seismology [4, 5, 6], and electromagnetics [7, 8]. The eigen-structure of the array covariance matrix or its inverse plays an important role in these approaches, in particular for data-adaptive implementations using, e.g., MVDR or MUSIC. A possible solution to locate sources in dense networks without relying on medium information is the spatiotemporal analysis of signal intensity [9, 10, 11]. However, using power alone cannot detect sources near or below the noise floor.

Table 1. Three scenarios considered for $M = 19$ snapshots of the noise processes x_i and x_j (see Figure 1). $x_i(q)$ referes to the q-th snapshot of process x_i.

Scenario	Variance x_i	Variance x_j
Stationary	$\sigma_i^2(1\dots19)=1$	$\sigma_j^2(1\dots19)=1$
Heteroscedastic 1	$\sigma_i^2(1\dots5)=10$	$\sigma_j^2(1\dots5)=10$
	$\sigma_i^2(6\dots19)=1$	$\sigma_j^2(6\dots19)=1$
Heteroscedastic 2	$\sigma_i^2(1\dots5)=10$	$\sigma_j^2(1\dots14)=1$
	$\sigma_i^2(6\dots19)=1$	$\sigma_j^2(15\dots19)=10$

We present a model-free analysis approach that can work also for weak signals through the use of coherent averaging. The only assumption made is that source signals enter the noise floor within a distance that is much smaller than the array aperture. That requirement is realistic for large arrays based on wave propagation in moderately attenuating media such as the earth. We follow a graph-based analysis paradigm [12, 13]: The sensors are considered vertices in a graph with edges between vertices existing if the corresponding sensors signals are coherent. Thus the graph connectivity matrix is defined as the support of the array coherence matrix.

We demonstrate that finding weak within-aperture sources is tantamount to identifying connected components in this graph (here referred to as clusters). Such clusters can be found through an eigenvalue decomposition of a matrix (the graph Laplacian) that is derived from the connectivity matrix [14]. For more details on the approach see the full paper [15].

2. GRAPH DEFINED BY COHERENCE MATRIX

Consider a large aperture array with N sensors distributed densely over spatial locations $\{\mathbf{r}_i\}_{i=1,\dots,N}$. The arrays we consider are quasi-uniform, but different configurations are applicable. This section first describes a hypothesis test to find the support of the coherence matrix of these sensors and then describes how to use this support to construct a graph and find its clusters.

ISBN 978-0-9928626-7-1

2.1. Coherence hypothesis test

2.1.1. Robust coherence

We compare the behavior of two definitions of magnitude of coherence for uncorrelated and heteroscedastic signals, i.e. stochastic signals with time-varying variance. Consider a zero-mean signal $u_j(t)$ observed at location $\mathbf{r}_j$ and captured by a window of Q samples at intervals Δt. Its discrete Fourier transform over a period $T_W = Q\Delta t$ during the m-th window (snapshot) is defined as:

$$x_j(m,f) = \sum_{q=0}^{Q-1} w_q \, u_j(mQ\Delta t + q\Delta t)\, e^{-\imath 2\pi (q\Delta t) f} \;, \quad (1)$$

where the frequencies are discretized $f = \frac{k}{T_W}, k = 0, \ldots, Q/2$ and the weights w_q control spectral leakage. All the considerations in this article are made in the Fourier domain.

Let $x_i(m)$ and $x_j(m)$ be a sequence of snapshots of two such signals in the frequency domain, $m = 1, \ldots, M$. The sample covariance is defined as

$$\hat{\Gamma}_{ij} = \frac{1}{M} \sum_{m=0}^{M-1} x_i(m) x_j^*(m) \;, \quad (2)$$

with the actual covariance reached for infinite snapshots: $\hat{\Gamma}_{ij} \xrightarrow{M\to\infty} \Gamma_{ij}$.The sample covariance estimate will be affected by the variances in x_i and x_j that are unrelated to any physical relation between the two sensors. A customary attempt to reduce the impact of such variations is to compute the coherence as a normalized covariance. We compare a full-sample normalized coherence:

$$\hat{C}_{ij}^c = \left| \frac{\sum_{m=0}^{M-1} x_i(m) x_j^*(m)}{\left(\sum_{m=0}^{M-1} |x_i(m)|^2\right)^{1/2} \left(\sum_{m=0}^{M-1} |x_j(m)|^2\right)^{1/2}} \right| \;, \quad (3)$$

and a version relying only on phase-information:

$$\hat{C}_{ij} = \left| \frac{1}{M} \sum_{m=0}^{M-1} \frac{x_i(m)}{|x_i(m)|} \frac{x_j^*(m)}{|x_j(m)|} \right| \;, \quad (4)$$

The coherence (4) only relies on phase information and is invariant against heteroscedasticity, i.e. signals with time-varying intensity [16] which are common in seismic and acoustic time-series. To illustrate the effect of heteroscedasticity we consider three scenarios for x_i and x_j as shown in Table 1, where x_i and x_j are uncorrelated and each i.i.d. complex Gaussian with a variance that depends on the snapshot index. Figure 1A shows the simulated pdf of $\hat{C}_{ij}^c$ for the three scenarios (based on 10^6 realizations). The pdf of the sample coherence $\hat{C}_{ij}^c$ substantially deviates from the stationary case for the two non-stationary scenarios considered. This instability with respect to heteroscedasticity makes $\hat{C}_{ij}^c$ a poor choice for a hypothesis test against independence. Figure 1B shows the pdf of $\hat{C}_{ij}$ for the same scenarios as before and demonstrates how the distribution of this statistic is invariant for the considered heteroscedasticity scenarios.

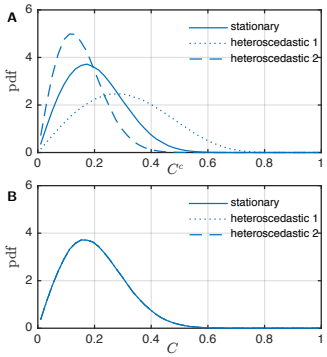

Fig. 1. *(A)* The PDF of the sample coherence (3) for uncorrelated signals x_i and x_j. A stationary and two heteroscedastic scenarios are considered (see Table 1). *(B)* The pdf of sample coherence (4) for the same three scenarios (lines overlap).

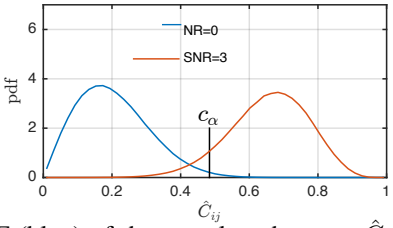

Fig. 2. PDF (blue) of the sample coherence $\hat{C}_{ij}$ (4) for uncorrelated noise, $\Gamma_{ij} = 0$, for M=19. The decision threshold c_α=0.484 is exceeded with probability α=0.01. The PDF (orange) of $\hat{C}_{ij}$ for the case of $\Gamma_{ij} \neq 0$ with a common signal present in the noise of the two recordings ($\mathsf{SNR} = 3$).

2.1.2. Hypothesis test

We describe a hypothesis test with robust test-statistics to establish the support of the array coherence matrix Γ_{ij}. We test for the two alternative hypotheses:

$$H_0 : \Gamma_{ij} = 0 \qquad \text{and} \qquad H_1 : \Gamma_{ij} \neq 0 \;, \quad (5)$$

i.e. the signals observed at locations $\mathbf{r}_i$ and $\mathbf{r}_j$ are uncorrelated (H_0) or correlated (H_1). We use the magnitude of the robust sample coherence $\hat{C}_{ij}$ (4) to test the hypothesis. If H_0 is true, then $\hat{C}_{ij}$ will be distributed according to a precomputable PDF (Figure 2 shows the PDF for M=19 derived by simulation). The hypothesis H_0 is accepted if $\hat{C}_{ij} \leq c_\alpha$ and rejected otherwise. The threshold coherence magnitude c_α is set such that the probability of falsely rejecting the hypothesis is α, formally $c_\alpha = \text{cdf}^{-1}(1-\alpha)$, where $\text{cdf}^{-1}(\cdot)$ is the inverse of the cumulative distribution function of $\hat{C}_{ij}$ estimated by simulation (blue plot in Figure 2).

ISBN 978-0-9928626-7-1 © EURASIP 2017

2.2. Graph preliminaries

An undirected (symmetric) and unweighted graph G consists of N vertices $\{v_i\}_{i=1\ldots N}$ and edges $\{e_{ij}\}_{j\geq i} \in \{0,1\}$ where $e_{ij} = 1$ means that v_i and v_j are connected. The edges define a binary and symmetric connectivity matrix $\mathbf{E}$ with $E_{ij} = e_{ij}$. The number of edges connecting to vertex v_i is its degree d_i. The mean vertex degree γ of a graph is the average over the vertex degrees of all its vertices. If γ approaches a constant as N increases the graph is sparse [17].

A connected component $U \subset G$ is a subset of vertices and edges in G for which every pair of vertices $v, v' \in U$ is connected directly or indirectly through a sequence of edges in U Finding connected components is a basic task in graph analysis [18] and is an example of spectral clustering for non-overlapping clusters. In this context the connected components are found using the eigen-vectors of the graph Laplacian, which is derived from the connectivity matrix $\mathbf{E}$ [14]. We assume now that there are K fully connected components U^k in G, i.e. each vertex in U^k connects to all other vertices in U^k. Let $\mathbf{u}^k = [u_1^k, \ldots, u_N^k]^T$ be the vertex indicator vector of U^k ($u_i^k = 1$ if $v_i^k \in U^k$ and 0 otherwise). The connectivity matrix of G is then (see [15]):

$$\mathbf{E} = \sum_k \mathbf{u}_k \mathbf{u}_k^T \,, \tag{6}$$

where k indexes the set of connected components of G.

Finally, consider the random, unweighted graph $G_0(N,p)$ with N vertices where all pairs of vertices have the same probability p of being connected. The mean vertex degree in $G_0(N,p)$ is therefore $\gamma = (N-1)p$ because every vertex can connect with all $N-1$ other vertices with equal probability. A large fraction of vertices in a random graph tend to be connected when $\gamma > 1$ with about 90% being connected when $\gamma > 2.5$ [17]. Such a high connectivity, e.g. 90%, will thus occur above a threshold probability of

$$p_0 = 2.5/(N-1) \,, \tag{7}$$

i.e. for an edge probability above p_0 most vertices is connected.

2.3. Constructing an array graph

Armed with the hypothesis test in Section 2.1.2 we construct a coherence graph G_0 with the following connectivity matrix:

$$E_{ij}^0 = \begin{cases} 1 & \text{if } \hat{C}_{ij} > c_\alpha \\ 0 & \text{otherwise} \,, \end{cases} \tag{8}$$

i.e. two vertices are connected if the corresponding signals exhibit significant coherence. This straight-forward construction of an array graph, however, is insufficient because of the statistical fluctuations of the hypothesis test. Even if the array is sensing N uncorrelated noise signals the probability of observing $\hat{C}_{ij} > c_\alpha$ is α for all receiver pairs. This means that G_0 is a random graph $G_0(N,\alpha)$. A graph with, say, 300 sensors will likely have a giant connected component if $\alpha > 2.5/(300-1) = 0.008$. For graphs constructed with a less conservative threshold any attempt to find smaller connected components that are not due to chance is thus futile.

We modify (8) to define a localized coherence graph $G(c_\alpha)$ with connectivity matrix:

$$E_{ij} = \begin{cases} 1 & \text{if } \hat{C}_{ij} > c_\alpha \text{ and } i \in \mathsf{N}(j) \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

where $\mathsf{N}(j)$ is the index set of the nearest neighbors of sensor j. For a regular lattice, the nearest neighbors are here limited to eight sensors. Besides being coherent any two connected sensors are thus also required to be spatial neighbors.

Enforcing spatially short connections limits the number of neighbors any vertex can connect to in a way that is independent of the global graph size and the graph remains sparse for large arrays. The criterion (9) thus reduces the chance of forming clusters by chance, even for values of α that are above the threshold suggested by (7). Sensor clusters can still have a spatial extent beyond that given by the nearest neighbors as long as the vertices in the cluster are contiguous in space.

To characterize the spatial extent of each connected component U^k a two-dimensional Gaussian probability density function is estimated from the sensor locations of the vertices of U^k with mean and covariance, respectively:

$$\begin{aligned} \mathbf{m}_k &= \frac{1}{|U^k|} \sum_{i \in U^k} \mathbf{r}_i \\ \mathbf{\Sigma}_k &= \frac{1}{|U^k|} \sum_{i \in U^k} (\mathbf{r}_i - \mathbf{m}_k)(\mathbf{r}_i - \mathbf{m}_k)^T, \end{aligned} \tag{10}$$

where $|U^k|$ is the number of vertices in U^k. The source area is the region where the point source is likely located and is here defined as the ellipse that contains a probability mass p of the Gaussian defined in (10):

$$\Omega_k(p) = \{\mathbf{r} \mid (\mathbf{r} - \mathbf{m}_k)^T \mathbf{\Sigma}_k^{-1} (\mathbf{r} - \mathbf{m}_k) < \chi_I^2(p)\} \,, \tag{11}$$

where χ_I^2 is the cumulative inverse χ^2-distribution with two degrees of freedom (because the Gaussian is 2D).

For a source within the array aperture Ω is the geographic area within which the source is estimated to be. Source directionality, physical obstacles or attenuation heterogeneities in the propagation medium can cause Ω to be not centered around a source. Sensor geometry such as array gaps and boundaries will also cause a cluster to move away from its source. In those special cases the identified clusters can, however, still serve to select a data subset for follow-up analysis with other array processing methods since by definition its sensors contain significant signal levels from a common source. Conventional beamforming using just the data from the vertices in U^k could be used.

ISBN 978-0-9928626-7-1 © EURASIP 2017

3. SOURCES INDUCE GRAPH CLUSTERS

The relation between sources within an array and the clusters of a graph constructed from the array data has been presumed so far. Combining signal features for clustering purposes to analyze sources was used implicitly in a heuristic approach in [19] for the difficult case of an ad-hoc and dynamic sensor network with communication constraints. In this section we make the relation between sources and network clusters explicit for the asymptotic case of infinite observation time without communications constraints but under the assumption that source-to-receiver coherence is insignificant after some physical distance.

Consider again a large aperture array with N sensors distributed densely over spatial locations $\{\mathbf{r}_i\}_{i=1,\dots,N}$. It is assumed that there are weak sources within the aperture that produce signals that propagate through space. For a given frequency the channel between any such source location $\boldsymbol{\rho}$ and sensor location $\mathbf{r}_i$ is characterized by a Green's function, $g(\mathbf{r}_i, \boldsymbol{\rho})$. Let the vector $\mathbf{g}(\boldsymbol{\rho}) = [g(\mathbf{r}_1, \boldsymbol{\rho}), \dots, g(\mathbf{r}_N, \boldsymbol{\rho})]^T \in \mathbb{C}^N$ be the frequency domain response of the array to a source at location $\boldsymbol{\rho}$. Consider then $\boldsymbol{\rho}_k$ to be the location of K sources $\{\boldsymbol{\rho}_k\}_{k=1,\dots,K}$ with an associated response $\mathbf{g}(\boldsymbol{\rho}_k) \equiv \mathbf{g}_k$ and source signals s_k. The measured signal at the N array sensors is thus modeled as:

$$\mathbf{x} = \sum_{k=1}^{K} s_k \, \mathbf{g}_k + \mathbf{n} \, . \tag{12}$$

where $\mathbf{n} = [n_1 \dots n_N]^T \in \mathbb{C}^N$ is a multivariate i.i.d. noise process. From (12) the covariance matrix is:

$$\boldsymbol{\Gamma} = \langle \mathbf{x}\mathbf{x}^H \rangle = \left\langle \sum_{k,l=1}^{K} s_k s_l^* \, \mathbf{g}_k \mathbf{g}_l^H \right\rangle + \langle \mathbf{n}\mathbf{n}^H \rangle$$

$$= \sum_{k,l=1}^{K} \langle s_k s_l^* \rangle \, \mathbf{g}_k \mathbf{g}_l^H + \mathbf{D} = \sum_{k=1}^{K} \langle |s_k|^2 \rangle \, \mathbf{g}_k \mathbf{g}_k^H + \mathbf{D} \, , \tag{13}$$

Here we exploit the mutual independence between the source and noise processes $\langle s_i n_j^* \rangle = 0, \langle n_i n_j^* \rangle = \langle s_i s_j^* \rangle = a_i \delta_{ij}, \ \forall i, j$ (a_i is an unknown power), and $\mathbf{D}$ is a diagonal matrix with D_{ii} the noise variance of sensor i.

We assume that there is a distance δ smaller than the array dimensions such that $g(\mathbf{r}_i, \boldsymbol{\rho})$ is small if $\|\boldsymbol{\rho} - \mathbf{r}_i\|_2 > \delta$, i.e. signals cannot be detected beyond δ. This is used to form connected graph for each source. If all sources are separated by at least 2δ, $\|\boldsymbol{\rho}_k - \boldsymbol{\rho}_l\|_2 > 2\delta \forall k \neq l$, then the corresponding support sets of the sources do not overlap.

Let $\mathcal{I}(\mathbf{v})$ be the support indicator function of a vector or matrix $\mathbf{v}$. The lack of overlap of the $\mathbf{g}_k$ and the support-indicator function properties (see [15]) allow us to write the support of the sum in (13) as

$$\mathcal{I}(\boldsymbol{\Gamma}) = \mathcal{I}\left(\sum_{k=1}^{K} |s_k|^2 \mathbf{g}_k \mathbf{g}_k^H \right) = \sum_{k=1}^{K} \mathcal{I}(\mathbf{g}_k) \mathcal{I}(\mathbf{g}_k)^T \, . \tag{14}$$

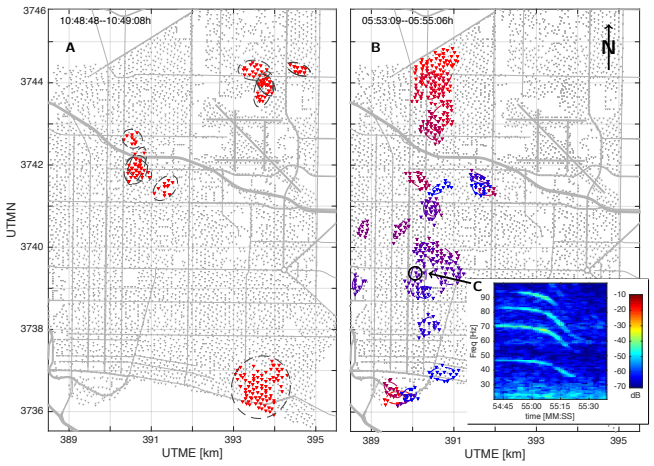

Fig. 3. Connected components of the array graph are used to find coherent sensor clusters in the Long Beach array. *(A)* The clusters at 20 Hz from four 10.2 s windows after 10:48h on March 11th. The spatial extent of the clusters $\Omega(0.5)$ is indicated by dashed ellipses (11). *(B)* A North-to-South helicopter transect is captured in a sequence of coherent clusters at 47 Hz over consecutive analysis time periods (starting at 05.53h). The colors change from red to blue as the analysis windows advance in time. The arrow points to the receiver from which the spectrogram *(C)* was computed around the time of increased coherence.

We now define $E_{ij} = \mathcal{I}(\boldsymbol{\Gamma})$ as the connectivity matrix of a graph G with N vertices (sensors), i.e. there is an edge between vertices i and j if $E_{ij} = \mathcal{I}(\mathbf{C})_{ij} = 1$. Such a graph G will have exactly K connected components, i.e. K non-overlapping subsets $S^k \in G$ whose vertices are indicated by $\mathcal{I}(\mathbf{g}_k)$. The connected components thus correspond to the sensor clusters that sensed the K sources.

Finding connected components is a standard task in graph analysis [18]. In section 4 we use an approach from spectral clustering [14] which uses the eigendecomposition of the graph Laplacian, which is derived from the connectivity matrix as

$$\mathbf{L} = \mathbf{K} - \mathbf{E} = \mathbf{U}\,\mathbf{S}\,\mathbf{U}^T \quad , \tag{15}$$

with $\mathbf{K}$ a diagonal matrix with $\mathbf{K}_{ii} = \sum_{j=1}^{N} E_{ij}$ and the latter is the eigen-decomposition of $\mathbf{L}$. Following [14, Proposition 2] there will be exactly K eigenvectors with eigenvalue 0 and the column vectors in $\mathbf{U}$ corresponding to those eigenvectors will each indicate one of the K connected components through its non-zero entries.

4. LONG BEACH (CA) GEOPHONE ARRAY

We apply the above technique on a geophone array that was deployed over an area of 7 × 10 km in Long Beach (California, US) as part of an industrial seismic survey [10, 6]. The array consisted of more than 5200 geophones (OYO CT32D

ISBN 978-0-9928626-7-1 © EURASIP 2017

vertical velocity sensors with 10 Hz corner frequency) sampling at a period of Δt=4 ms (array configuration shown in Figure 3A). For the most part the array had a quasi-regular layout with a relatively even spatial sampling.

The ground velocity data stream of each geophone is transformed following (1) into a sequence of Fourier coefficients $x_i(m, f)$ using $Q = 256$ samples ($T_W = 1.02$ s) and a Hanning window w_j with time windows overlapping by $T_W/2$. The coherence matrix $\hat{\mathbf{C}}_{ij}$ (4) is computed for 41 frequency bins from 9.8-48.8 Hz using $M = 19$ snapshots ($T_W/2 \times (M + 1)$=10.2 s). A matrix with about $5200^2 \approx 27 \cdot 10^6$ entries is therefore computed for every frequency bin and time period. In a 24 hour analysis period there are about 9400 time windows.

Continuing with the coherence matrices at 20 Hz a localized array graph $G(c_{\alpha=0.01})$ is defined and all connected components are identified. Figure 3A shows the coherent groups found over four consecutive 10.2 s analysis windows starting on March 11th, 10:48:48h. For clarity, only clusters with more than nine vertices are shown in order to focus on larger phenomena. The period contains a 40 s stretch during which a seismic vibrotruck is known to have been operating in the Southeast of the array, which is confirmed by a cluster in that area. This source dominated over the background and was also broad-band (about 10-80 Hz) and therefore corresponds to the type of source that was also detected with the energy-only approach reported in [10].

Figure 3B shows a sequence of coherent groups at 47 Hz for consecutive windows starting March 11th at 05.53h. They show a north-south transect over 6 km during the course of about 95 s. The average velocity along the trajectory is 60 m/s (134 mph). Figure 3C shows a spectrogram from a receiver within the trajectory of the moving source computed around the time the coherence was observed. The observed Doppler shifts of $f_{\text{high}}/f_{\text{low}} \simeq 1.4 \simeq (1 + \frac{60}{340})/(1 - \frac{60}{340})$ are consistent with the approximate velocity estimate. The narrow-band harmonics at multiples of 12 Hz suggest that the passage of a helicopter was captured.

5. REFERENCES

[1] A.B. Baggeroer, W.A. Kuperman, and P.N. Mikhalevsky. An overview of matched-field methods in ocean acoustics. *IEEE J. Ocean. Eng.*, 18(4):401–424, Oct 1993.

[2] JC. Chen, K Yao, and RE. Hudson. Acoustic source localization and beamforming: Theory and practice. *EURASIP J Adv Signal Process*, 2003(4):1–12, 2003.

[3] A. Beck, Petre Stoica, and Jian Li. Exact and approximate solutions of source localization problems. *Signal Processing, IEEE Transactions on*, 56(5):1770–1778, May 2008.

[4] DB Harris and T Kvaerna. Superresolution with seismic arrays using empirical matched field processing. *Geophys. J. Int.*, 182(3):1455–1477, 2010.

[5] M. Corciulo, P. Roux, M. Campillo, D. Dubucq, and WA. Kuperman. Multiscale matched-field processing for noise-source localization in exploration geophysics. *Geophys.*, 77(5):KS33–KS41, Sep 2012.

[6] A. Inbal, JP Ampuero, and RW Clayton. Localized seismic deformation in the upper mantle revealed by dense seismic arrays. *Science*, 354(6308):88–92, 2016.

[7] DF Gingras, P Gerstoft, and NL Gerr. Electromagnetic matched-field processing: Basic concepts and tropospheric simulations. *IEEE Trans Ant. Prop*, 45(10):1536–1545, 1997.

[8] P. Valtr, P. Pechac, V. Kvicera, and M. Grabner. Estimation of the refractivity structure of the lower troposphere from measurements on a terrestrial multiple-receiver radio link. *IEEE Transactions on Antennas and Propagation*, 59(5):1707–1715, May 2011.

[9] Dan Li and Yu Hen Hu. Energy-based collaborative source localization using acoustic microsensor array. *EURASIP Journal on Advances in Signal Processing*, 2003(4):985029, 2003.

[10] N Riahi and P Gerstoft. The seismic traffic footprint: Tracking trains, aircraft, and cars seismically. *Geophys. Res. Lett.*, 42(8):2674–2681, 2015.

[11] CD de Groot-Hedlin and MAH Hedlin. A method for detecting and locating geophysical events using groups of arrays. *Geophys. J. Int.*, 203(2):960–971, 2015.

[12] DI Shuman, SK Narang, P Frossard, A Ortega, and P Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process Mag.*, 30(3):83–98, 2013.

[13] A Sandryhaila and JMF Moura. Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure. *IEEE Signal Process Mag*, 31(5):80–90, 2014.

[14] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, DEC 2007.

[15] N. Riahi and P. Gerstoft. Using graph clustering to locate sources within a dense sensor array. *Signal Processing*, 132:110–120, 2017.

[16] AD Chave, DJ Thomson, and ME Ander. On the robust estimation of power spectra, coherences, and transfer functions. *J Geophys Research:*, 92(B1):633–648, 1987.

[17] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[18] Satu Elisa Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27 – 64, 2007.

[19] D Friedlander, C Griffin, N Jacobson, S Phoha, and RR Brooks. Dynamic agent classification and tracking using an ad hoc mobile acoustic sensor network. *EURASIP J Adv Signal Process*, 2003(4):1–7, 2003.

ISBN 978-0-9928626-7-1 © EURASIP 2017