2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)

Bayesian Sparse Wideband Source Reconstruction of Japanese 2011 Earthquake

Christoph F. Mecklenbräuker
Institute of Telecommunications
Vienna University of Technology
1040 Vienna, Austria
cfm@nt.tuwien.ac.at

Peter Gerstoft, Huajian Yao
Scripps Institution of Oceanography
La Jolla, CA 92093-0238, USA
gerstoft@ucsd.edu

***Abstract*—We consider the sparse inversion of seismic recordings from a Bayesian perspective. We have a prior belief that the spatially distributed seismic source should be sparse in the spatial domain. In a Bayesian framework, we assume a Laplace-like prior for a distributed wideband source and derive the corresponding objective function for minimization. We solve a sequence of convex minimization problems for finding a sparse seismic source representation from an underdetermined system of linear measurement equations using teleseismic P waves recorded by an array of sensors. The root mean square reconstruction error for the source distribution is evaluated through numerical simulations.**

I. INTRODUCTION

Array processing methods for source localization have been used for the analysis of seismic recordings for many decades. It was this field of application which originally motivated Capon's beamformer [1], [2].

In a Bayesian perspective, an inversion result is the *a posteriori* probability density for the parameters of interest, from which all information regarding the sources and propagation environment can be inferred [3]. Many inversion experiments have relied on the *a priori* knowledge that only a single point source is present. Such inversion approaches can be interpreted as optimization problems with an extreme sparsity constraint: The spatial source distribution is a single Dirac-δ distribution.

Tibshirani [4] proposed the *least absolute shrinkage and selection operator* (lasso) to overcome drawbacks from ordinary least squares, lack of prediction accuracy and interpretability. In effect, the lasso is a least squares regression constrained by an upper bound on the ℓ_1-norm of the solution. He shows that this technique sets many coefficients of the regression solution to zero, and therefore favors *sparse* solutions. Daubechies et al. [5] recommend the use of ℓ_p-norm penalized inversion problems with $p < 2$ when one expects the solution to be sparse. Since then numerous papers have appeared in the areas *compressive sensing* and *sparse reconstruction* which employ the penalization by a ℓ_1-norm to promote sparsity of the solution, e.g. [6], [7], [8].

Here, we extend the Bayesian approach, e.g. [9], [10] to a wideband array model for application to seismic recordings.

II. SEISMIC RECORDING MODEL

Let $(\theta_1, \ldots, \theta_M)^{\mathrm{T}}$ be a vector of M *potential* earthquake source locations on a suitably chosen grid on the earth's surface. Further, let $\boldsymbol{x} = (x_1, \ldots, x_M)^{\mathrm{T}}$ be the asociated complex-valued source vector. We observe time-sampled seismic waveforms on an array of N sensors which are stacked in a vector. After a short-time Fourier transform, we obtain the following linear model which relates the Fourier transformed sensor array output $\boldsymbol{y}(\omega)$ to the source vector,

$$\boldsymbol{y}(\omega) = \boldsymbol{A}(\omega)\boldsymbol{x}(\omega) + \boldsymbol{n}(\omega) \ . \tag{1}$$

The mth column of the transfer matrix $\boldsymbol{A}(\omega)$ is the steering vector $\boldsymbol{a}_m(\omega)$ for candidate source location θ_m. We model the (n, m)-element of $\boldsymbol{A}(\omega)$ by $\mathrm{e}^{-\mathrm{j}\omega\tau_{nm}}$. Here τ_{nm} is the traveltime from location θ_m to sensor n,

$$\tau_{nm} = \int_{\mathcal{C}_{nm}} \frac{\mathrm{d}s}{c_{\mathrm{p}}(s)} \ , \tag{2}$$

where we integrate along the curved ray $\mathcal{C}_{nm}$ which connects location θ_m to sensor n, and $c_{\mathrm{p}}(s)$ is the phase velocity of compressional ("P") waves along the ray in the earth's mantle. Travel times are calculated from the IASP91 1-D earth reference model [11]. The additive noise vector $\boldsymbol{n}(\omega)$ is assumed to be spatially uncorrelated and follows the zero-mean complex normal distribution with diagonal covariance $\nu(\omega)\boldsymbol{I}$. Further, the noise is assumed to be uncorrelated across frequency.

Let us stack all individual single-frequency observations $\boldsymbol{y}(\omega_j)$ for $j = 1, 2, \ldots, J$ into a long $NJ \times 1$ vector $\boldsymbol{Y}$, and similarly for the $MJ \times 1$ source vector $\boldsymbol{X}$ and the $NJ \times 1$ additive noise vector $\boldsymbol{N}$. Further, we define the $NJ \times MJ$ block diagonal matrix

$$\boldsymbol{A} = \mathrm{diag}\left(\boldsymbol{A}(\omega_1),\ \boldsymbol{A}(\omega_2),\ \ldots, \boldsymbol{A}(\omega_J)\right),$$

then the multi-frequency model becomes

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{N} \ . \tag{3}$$

III. BAYESIAN FORMULATION

First, we select a set of frequencies $\{\omega_1, \ldots, \omega_J\}$ at which we observe measurement data $\boldsymbol{Y}$ according to the model (3). For the linear model (3), we arrive at the following conditional

978-1-4577-2105-2/11/$26.00 ©2011IEEE

probability density for multi-frequency observations given the wideband source vector $\boldsymbol{X}$,

$$p(\boldsymbol{Y}|\boldsymbol{X}) = \prod_{j=1}^{J} \frac{\exp\left(-\frac{1}{\nu_j}\|\boldsymbol{y}(\omega_j) - \boldsymbol{A}(\omega_j)\boldsymbol{x}(\omega_j)\|_2^2\right)}{(\pi\nu_j)^N} \tag{4}$$

where we have used the shorthand notation $\nu_j = \nu(\omega_j)$ and $\|\cdot\|_p$ is the ℓ_p-norm. In our setting, the number of candidate source locations M is larger than the number of sensors N, i.e. $N < M$ and the linear models (1) and (3) are underdetermined. To reconstruct a physically meaningful source vector, we exploit its sparsity through the introduction of a prior probability density on $\boldsymbol{X}$ which promotes sparsity. A widely used sparseness prior is the Laplace probability density [9], [12] which puts a higher probability mass than the normal distribution both for *large* and *small* absolute deviations. We assume the multivariate complex Laplace-like probability density [13],

$$p(\boldsymbol{X}) = \prod_{j=1}^{J} \left(\frac{\lambda_j}{\sqrt{2\pi}}\right)^{2M} \exp\left(-\lambda_j \|\boldsymbol{x}(\omega_j)\|_1\right) \ . \tag{5}$$

The parameters λ_j are positive scale parameters which are suitably chosen later. For the posterior probability density function $p(\boldsymbol{X}|\boldsymbol{Y})$, we follow Bayes' rule and apply the logarithm. We arrive at the following cost function to be minimized,

$$\begin{aligned} \Lambda &= -\ln p(\boldsymbol{X}|\boldsymbol{Y}) = -\ln p(\boldsymbol{Y}|\boldsymbol{X}) - \ln p(\boldsymbol{X}) + \ln p(\boldsymbol{Y}) \\ &= \sum_{j=1}^{J} \Bigg(\frac{\|\boldsymbol{y}(\omega_j) - \boldsymbol{A}(\omega_j)\boldsymbol{x}(\omega_j)\|_2^2}{\nu_j} + \lambda_j \|\boldsymbol{x}(\omega_j)\|_1 \\ &\qquad + N \ln \nu_j - 2M \ln \lambda_j \Bigg) + \text{const.} \end{aligned} \tag{6}$$

In (6) the term $\ln p(\boldsymbol{Y})$ is absorbed into the additive constant which will be neglected in the following.

A. Arbitrary noise spectrum: Colored noise

For the case of colored noise, we assume that the $\nu_1, \ldots, \nu_J$ parameters are unknown non-negative constants. First, we minimize Λ with respect to ν_j and λ_j:

$$\begin{aligned} \frac{\partial \Lambda}{\partial \nu_j} = 0 \quad &\Rightarrow \quad \hat{\nu}_j = \frac{1}{N}\|\boldsymbol{y}(\omega_j) - \boldsymbol{A}(\omega_j)\boldsymbol{x}(\omega_j)\|_2^2 \\ \frac{\partial \Lambda}{\partial \lambda_j} = 0 \quad &\Rightarrow \quad \hat{\lambda}_j = \frac{2M}{\|\boldsymbol{x}(\omega_j)\|_1} \ . \end{aligned}$$

Inserting $\hat{\nu}_j$ and $\hat{\lambda}_j$ into (6) leads to the following wideband cost function,

$$\begin{aligned} \Lambda_2(\boldsymbol{X}) \ = \ 2 \sum_{j=1}^{J} \big(& N \ln \left(\|\boldsymbol{y}(\omega_j) - \boldsymbol{A}(\omega_j)\boldsymbol{x}(\omega_j)\|_2\right) \\ & + M \ln \|\boldsymbol{x}(\omega_j)\|_1 \big) \, . \end{aligned} \tag{7}$$

for minimization with respect to $\boldsymbol{X}$. The minimizing solution $\hat{\boldsymbol{X}}$ is the Bayesian sparse wideband maximum a posteriori source estimate.

B. Flat noise spectrum: White noise

Consider now the special case where ν and λ is constant across frequency, i.e., $\nu_j = \nu$ and $\lambda_j = \lambda$ whereby Λ becomes

$$\Lambda = JN \ln \nu - 2JM \ln \lambda + \lambda \|\boldsymbol{X}\|_1 + \frac{1}{\nu}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\|_2^2. \tag{8}$$

If we regard ν and λ as given constants then the cost function can be cast into the form

$$\Lambda_3 = \|\boldsymbol{X}\|_1 + \mu \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\|_2^2 \tag{9}$$

with $\mu = (\lambda\nu)^{-1}$. On the other hand, if we regard ν as an unknown parameter, the minimizing $\hat{\nu}$ turns out to be

$$\hat{\nu} = \frac{1}{JN}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\|_2^2 = \frac{1}{J}\sum_{j=1}^{J} \hat{\nu}_j, \tag{10}$$

i.e. $\hat{\nu}$ is the arithmetic mean over all single-frequency estimates $\hat{\nu}_j$. Next, we minimize (8) with respect to λ,

$$\frac{\partial \Lambda}{\partial \lambda} = 0 \quad \Rightarrow \quad \hat{\lambda} = \frac{2JM}{\|\boldsymbol{X}\|_1} \ , \tag{11}$$

and it turns out that $\hat{\lambda}$ is the harmonic mean over all single-frequency estimates $\hat{\lambda}_j$. Finally, we insert $\hat{\nu}$ and $\hat{\lambda}$ into (8). which leads to the following cost function for minimization

$$\Lambda_5 = N \ln \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\|_2 + M \ln \|\boldsymbol{X}\|_1 \ . \tag{12}$$

IV. COMPUTATION OF SPARSE REPRESENTATION

The cost function (9) is convex for $\mu \geq 0$ and this leads to a second order cone problem which is efficiently solvable with interior points solvers [14].

Unfortunately, the cost functions (7) and (12) are non-convex due to the logarithm [14]. We minimize these by a sequence of convex minimization problems by locally linearizing the logarithm via its Taylor expansion,

$$\ln z \approx \ln z_0 + \frac{1}{z_0}(z - z_0) \ . \tag{13}$$

Given an approximate solution vector $\boldsymbol{X}_0$ to the minimizer of Λ_5, this leads to the convex approximation valid in the neighborhood of $\boldsymbol{X}_0$

$$\Lambda_5 \approx a_1 \|\boldsymbol{X}\|_1 + a_2 \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\|_2 \ , \tag{14}$$

with the coefficients

$$a_1 = \frac{M}{\|\mathbf{X}_0\|_1} \qquad a_2 = \frac{N}{\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}_0\|_2}. \tag{15}$$

By applying the locally convex approximation (14) repeatedly, we arrive at the iterative procedure in Table I which uses the convex optimization package `CVX` [15], [16].

given: Arrays $\boldsymbol{A}$ of size $NJ \times MJ$ and $\boldsymbol{Y}$ of size $NJ \times 1$

```
a1=M
a2=N
for cnt = 1:ncnt
   cvx_begin
      variables L1 L2
      variable X(M) complex
      minimize(a1*L1 + a2*L2)
      subject to
         L1 > norm(X,1)
         L2 > norm(Y-A*X,2)
   cvx_end
   a1 = M/norm(X,1)
   a2 = N/norm(Y-A*X,2)
end
```

output: $\boldsymbol{X}$ of size $MJ \times 1$

TABLE I
COMPUTATIONAL PROCEDURE IN MATLAB/CVX FOR MINIMIZING THE COST FUNCTION Λ_5 EQ.(12) USING [15], [16].

V. NUMERICAL RESULTS

A. *Reconstruction performance for seismic array*

First, we report on numerical simulations with synthetic data for modeling the USArray [18] with $N = 409$ wideband seismic sensor stations. Figure 1 shows the geometry of the sensor array which has a rather random geometry. Instead of a single source location (as indicated by the red star $\star$ at the center of Fig. 1), we assume a sparsely distributed source. Based on the simulations, we evaluate the Root Mean Squared Reconstruction Error (RMSRE) for the source vector.

We create a regular grid of 40×40 *candidate* source locations covering the latitudes and longitudes shown in Fig. 3 and resulting in $M = 1600$. We synthetically generated a sparse source vector $\boldsymbol{X}_\star$ with $K = 6$ randomly selected elements representing equal power sources with uniformly distributed random phases. The remaining $M - K = 1594$ elements are set to zero (almost surely). The traveltimes τ_{nm} which enter into $\boldsymbol{A}$ are calculated according to (2), cf. [11], [17].

For sparse source reconstruction, we minimize the cost function Λ_5 defined in (12) via the computational procedure in Table I. For simplicity, we use 5 iterations only. Figure 2 shows the RMSRE from the corresponding simulation runs,

$$\mathrm{RMSRE} = \sqrt{\mathrm{E}\, \|\boldsymbol{X} - \boldsymbol{X}_\star\|_2^2} \ , \tag{16}$$

versus Signal-to-Noise Ratio (SNR) in dB,

$$\mathrm{SNR} = -10 \log_{10} \nu \ , \tag{17}$$

for $\frac{\omega_1}{2\pi} = 0.3125\,\mathrm{Hz}$, $\frac{\omega_2}{2\pi} = 0.6250\,\mathrm{Hz}$ and $\frac{\omega_3}{2\pi} = 0.9375\,\mathrm{Hz}$. From Fig. 2, we conclude that the RMSRE for source distributions which emit high frequency P waves can be reconstructed more precisely than low frequency sources.

B. *Sparse source reconstruction for Tohoku-Oki earthquake*

We apply sparse source reconstruction to image the temporal evolution of the Tohoku-Oki earthquake source distribution on March 11, 2011. The first $200\,\mathrm{s}$ of the teleseismic P waves recorded by over 500 stations in the US are used [17]. The seismic sensor output $y_n(t)$ for $(n = 1, \ldots, N)$ is sampled at $10\,\mathrm{Hz}$, subsequently 0.05–$2\,\mathrm{Hz}$ bandpass filtered, and finally normalized by its peak amplitude. We use $N = 476$ stations (Fig.1) with good data quality and estimate the sparse spatial source distribution using $10\,\mathrm{s}$ sliding windows. Before the earthquake, we only observe noise (Fig. 3a). Initially the earthquake is weak causing some source ambiguity (Fig. 3b), which disappeared as the earthquake gained strength (Fig. 3c–d).

VI. CONCLUSION

The wideband cost functions Λ_2 and Λ_5 for sparse source reconstruction are derived from a Bayesian perspective using a Laplace-like prior for the source. These cost functions are not convex, but they can be efficiently minimized by solving a sequence of convex problems. Numerical simulations with synthetic data indicate that sparse source reconstruction is applicable to seismic array data. Source distributions which emit high frequency P waves can be reconstructed more precisely than low frequency sources. Finally, we image the source distribution in space and time for the Tohoku-Oki earthquake.

ACKNOWLEDGMENT

The authors like to thank Gerald Matz and Florian Xaver for valuable discussions and comments on the manuscript. Work funded by Grants ICT08-44 of Wiener Wissenschafts–, Forschungs– und Technologiefonds and National Science Foundation EAR-0944109.

REFERENCES

[1] J. Capon: High resolution frequency-wavenumber spectrum analysis. Proc. IEEE, vol. 57, no. 8, pp. 1408-1418, 1969.

[2] H. L. Van Trees: Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory, Wiley, 2002.

[3] P. Gerstoft and C. F. Mecklenbräuker: Ocean acoustic inversion with estimation of a posteriori probability distributions, J. Acoust. Soc. Am., vol. 104, no. 2, pp. 808-819, 1998.

[4] R. Tibshirani: Regression Shrinkage and Selection via the Lasso, J. R. Statist. Soc. B, vol. 58, no. 1, pp. 267–288, 1996.

[5] I. Daubechies, M. Defrise, C. De Mol: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Communications on Pure and Applied Mathematics, vol. 57, no. 11, pp. 14131457, Nov. 2004.

[6] E. J. Candès, J. Romberg, T. Tao: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory, vol. 52, no. 2 489–509, Feb. 2006.

[7] D. Donoho: Compressed sensing, IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[8] R. G. Baraniuk, Compressive sensing, IEEE Signal Processing Mag., vol. 24, no. 4, pp. 118120, Jul. 2007, 124.

[9] S. Ji, Y. Xue, L. Carin, Bayesian Compressive Sensing, IEEE Trans. Sig. Proc., vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[10] S. D. Babacan, R. Molina, A. K. Katsaggelos: "Bayesian Compressive Sensing Using Laplace Priors," IEEE Trans. Image Proc., vol. 19, no. 1, pp. 53–63, Jan. 2010.

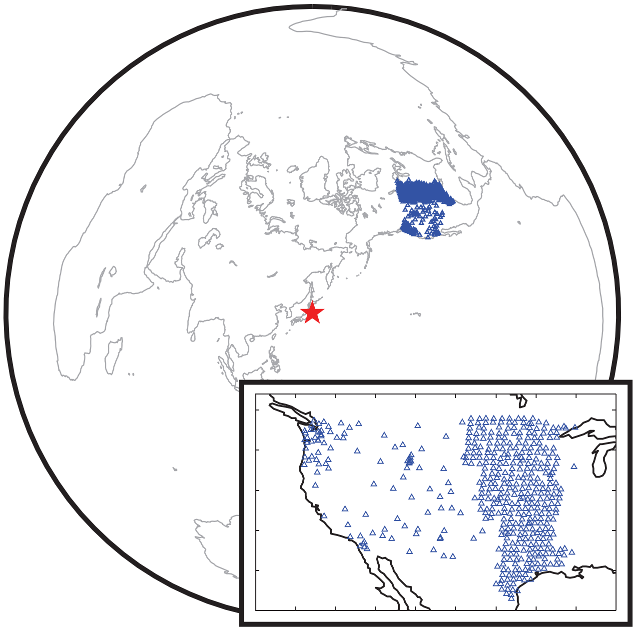

Fig. 1. Location of the 2011 Tohoku-Oki earhthquake (red star $\star$) and array geometry of seismic stations in the central and western US [17].

[11] B. L. N. Kennett, E. R. Engdahl: Traveltimes for global earthquake location and phase identification, Geophysical Journal International, vol. 105, no. 2, pp. 429–465, May 1991.

[12] J. M. Bernardo and A. F. M. Smith, Bayesian Theory. New York: Wiley, 1994.

[13] Z. He, S. Xie, S. Ding, A. Cichocki: Convolutive Blind Source Separation in the Frequency Domain Based on Sparse Representation, IEEE Trans. Audio, Speech, and Language Proc., vol. 15, no. 5, Jul. 2007.

[14] S. P. Boyd, L. Vandenberghe, *Convex optimization*, Chapters 1–7 (Cambridge University Press, 2004).

[15] M. Grant, S. Boyd, CVX: Matlab Software for Disciplined Convex Programming, version 1.21, http://cvxr.com/cvx, Apr. 2011.

[16] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, pp. 95–110 (Springer-Verlag Limited, 2008).

[17] H. Yao, P. Gerstoft, P. M. Shearer, C. F. Mecklenbräuker: Compressive sensing of Tohoku-Oki M9.0 Earthquake: Frequency-dependent Rupture Modes, Geophys. Res. Lett., doi:10.1029/2011GL049223.

[18] USArray `http://www.usarray.org/`

[19] L. Meng, A. Inbal, and J.-P. Ampuero: A window into the complexity of the dynamic rupture of the 2011 Mw 9 Tohoku-Oki earthquake, Geophys. Res. Lett., Vol. 38, L00G07, No. 6, pp. , 2011. doi:10.1029/2011GL048118.

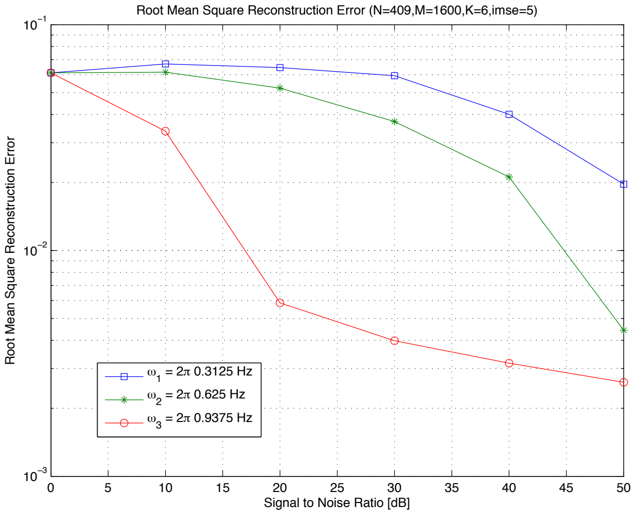

Fig. 2. RMSRE versus SNR at three frequencies of interest: Cost function Λ_5 in (12) is minimized by the computational procedure in Table I with `ncnt`=5. The USArray [18] has lower RMSRE at higher frequencies.

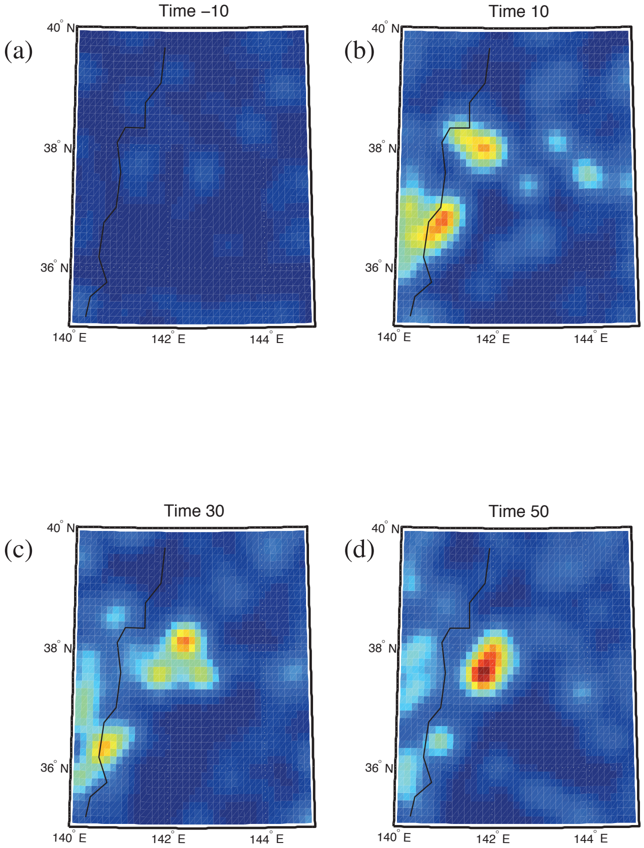

Fig. 3. Temporal evolution of spatial source power distribution for Tohoku-Oki Earthquake 2011: (a) $t = -10s$, (b) $t = 10s$, (c) $t = 30s$, (d) $t = 50s$.