Efficient Pseudo Gauss-Newton FWI in the time-ray parameter domain

Wenyong Pan, Kristopher A. Innanen, Gary F. Margrave

CREWES, University of Calgary

- **Introduction**
  - General Principle of FWI
  - Why FWI fails in industry practice ?
- **Theory and methods**
  - Gradient
  - Approximate Hessian
  - Pseudo-Hessian
  - Phase Encoded Hessian
  - Multiscale Approach
- **Numerical Example**
- **Conclusions**

General Principle of FWI

$$\phi\left(s^{(n)}(\mathbf{r})\right) = \frac{1}{2}\int d\omega \left(\sum_{\mathbf{r}_s,\mathbf{r}_g} \|\delta P\left(\mathbf{r}_g, \mathbf{r}_s, \omega | s^{(n)}(\mathbf{r})\right)\|_2\right)$$

Least-squares misfit function for full waveform inversion

General Principle of FWI

$$s^{(n+1)}(\mathbf{r}) = s^{(n)}(\mathbf{r}) + \mu^{(n)}\delta s^{(n)}(\mathbf{r})$$

The model can updated iteratively

$$\delta s^{(n)} = -\int d\mathbf{r}' H^{(n)-}(\mathbf{r},\mathbf{r}')g^{(n)}(\mathbf{r})$$

The model perturbation can be constructed by gradient and inverse Hessian

Why FWI fails in industry practice?

- ❑ Extensively computational burden
- ❑ Slow convergence rate
- ❑ Cycle skipping problem

Strategies

- ❑ Source encoding method
- ❑ Hessian approximations
- ❑ Multiscale approach

Gradient

$$g^{(n)}(\mathbf{r}) = -\sum_{\mathbf{r}_s,\mathbf{r}_g} \int d\omega \Re \left(\frac{\delta G\left(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)}\right)}{\delta s_0^{(n)}(\mathbf{r})} \delta P^* \right)$$

Gradient

$$g^{(n)}(\mathbf{r}) = -\sum_{\mathbf{r}_s,\mathbf{r}_g} \int d\omega \Re \left(\omega^2 \mathcal{F}_s(\omega) G(\mathbf{r}, \mathbf{r}_s, \omega) G(\mathbf{r}_g, \mathbf{r}, \omega) \delta P^* \right)$$

Gradient can be constructed using adjoint state method

Gradient

$$g^{(n)}(\mathbf{r}) = -\sum_{\mathbf{r}_s,\mathbf{r}_g}\int d\omega \Re\left(\frac{\delta G\left(\mathbf{r}_g,\mathbf{r}_s,\omega|s_0^{(n)}\right)}{\delta s_0^{(n)}(\mathbf{r})}\delta P^*\right)$$

Gradient

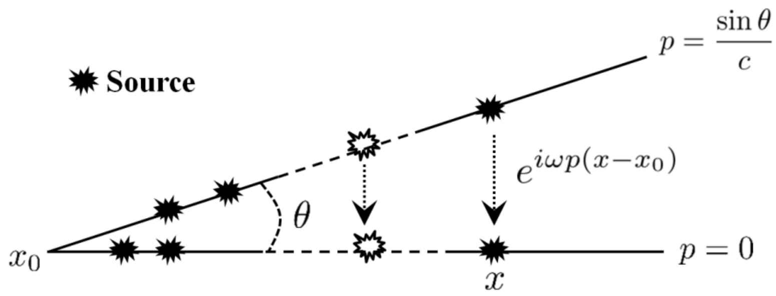
$$g^{(n)}(\mathbf{r}) = -\sum_{\mathbf{r}_s,\mathbf{r}_g}\int d\omega \Re\left(\omega^2\mathcal{F}_s(\omega)G(\mathbf{r},\mathbf{r}_s,\omega)G(\mathbf{r}_g,\mathbf{r},\omega)\delta P^*\right)$$

RTM image based on crosscorrelation imaging condition

$$g^{(n)}(\mathbf{r}) = -\int d\omega \Re\left\{\omega^2 \mathcal{F}_s(\omega)\tilde{G}(\mathbf{r},\mathbf{r}_s,\omega)G(\mathbf{r}_g,\mathbf{r},\omega)\delta P^*\right\}$$

Linear phase encoded gradient

where $\tilde{G}(\mathbf{r},\mathbf{r}_s,\omega) = G(\mathbf{r},\mathbf{r}_s,\omega)e^{i\omega p_s(x_s'-x_s)}$**,** p_s **is the ray-parameter,** x_s' **and** x_s **are the sources' coordinates.**



Linear phase encoding strategy

- ❑ **Integration over ray parameter can disperse the crosstalk noise.**
- ❑ **Different ray parameters can balance the gradient.**

The Functions of Hessian in Least-Squares Inverse Problem

Hessian Matrix ?

- ❑ **Improve the convergence rate**

The Functions of Hessian in Least-Squares Inverse Problem

Hessian Matrix = Nonstationary Deconvolution Operator

- Improve the convergence rate

- Compensate the geometrical spreading effects and balance the amplitude
- Suppress the multiple scattering effects and improve the resolution

Approximate Hessian

$$H^{(n)}(\mathbf{r}',\mathbf{r}) = \cancel{H_1^{(n)}} + H_2^{(n)}$$

Full Hessian

$$H_a^{(n)} = H_2^{(n)} \simeq \sum_{\mathbf{r}_s,\mathbf{r}_g} \int d\omega \Re\left\{\omega^4 G\left(\mathbf{r}_g,\mathbf{r}',\omega|s_0^{(n)}\right) G\left(\mathbf{r}',\mathbf{r}_s,\omega|s_0^{(n)}\right) G^*\left(\mathbf{r}_g,\mathbf{r}'',\omega|s_0^{(n)}\right) G^*\left(\mathbf{r}'',\mathbf{r}_s,\omega|s_0^{(n)}\right)\right\}$$

Approximate Hessian in Gauss-Newton Method by Gary and Kris (2011)

$$H_a^{(n)} \simeq diag\left(H_a^{(n)}\right)$$

Diagonal part of the approximate Hessian

Pseudo-Hessian

$$f_{virtual} = -\omega^2 \mathcal{F}_s(\omega) G\left(\mathbf{r}, \mathbf{r}_s, \omega | s_0^{(n)}(\mathbf{r})\right)$$

Virtual Source

$$H_{p\_a}^{(n)} = f_{virtual} f_{virtual}^* = \sum_{\mathbf{r}_s} \int d\omega \Re\{\omega^4 |\mathcal{F}_s(\omega)|^2 G(\mathbf{r}', \mathbf{r}_s, \omega) G^*(\mathbf{r}'', \mathbf{r}_s, \omega)\}$$

Pseudo-Hessian

$$I_{dec} = \frac{\sum_{\mathbf{r}_s, \mathbf{r}_g} \int d\omega \Re\left\{\omega^2 \mathcal{F}_s(\omega) G(\mathbf{r}, \mathbf{r}_s, \omega) G(\mathbf{r}_g, \mathbf{r}, \omega) \delta P^* \left(\mathbf{r}_g, \mathbf{r}_s, \omega\right)\right\}}{\sum_{\mathbf{r}_s} \int d\omega \omega^4 \Re\left\{|\mathcal{F}_s(\omega)|^2 G\left(\mathbf{r}', \mathbf{r}_s, \omega\right) G^*\left(\mathbf{r}', \mathbf{r}_s, \omega\right)\right\} + \lambda I}$$

Deconvolution imaging condition

Phase Encoded Hessian

$$H_{encoded} = \sum_{\mathbf{r}_s} \int d\omega \Re \left\{ \omega^4 G(\mathbf{r}', \mathbf{r}_s, \omega) G^*(\mathbf{r}'', \mathbf{r}_s, \omega) \right\}$$
$$\times \sum_{\mathbf{p}_g} \int d\omega \Re \left\{ G(\mathbf{r}', \mathbf{r}'_g, \omega) e^{i\omega p_g (x'_g - x_{initial})} G^*(\mathbf{r}'', \mathbf{r}_g, \omega) e^{-i\omega p_g (x_g - x_{initial})} \right\}$$

Receiver-side linear phase encoded Hessian

$$H_{encoded} = H_{exact} + \cancel{H_{crosstalk}}$$

By Tang (2009)

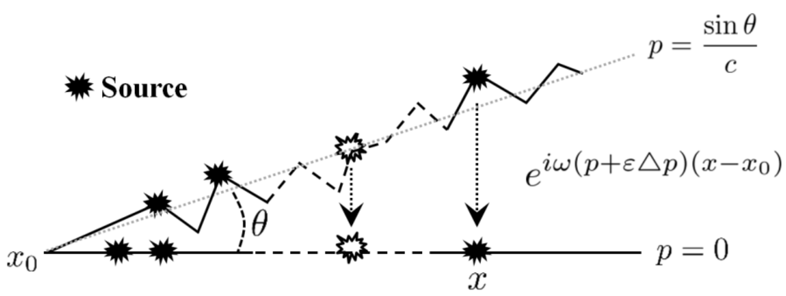
$$H_{encoded} = H_{exact}, \mathbf{p}_g \in (-\infty, +\infty)$$

By Tao and Sen (2013)

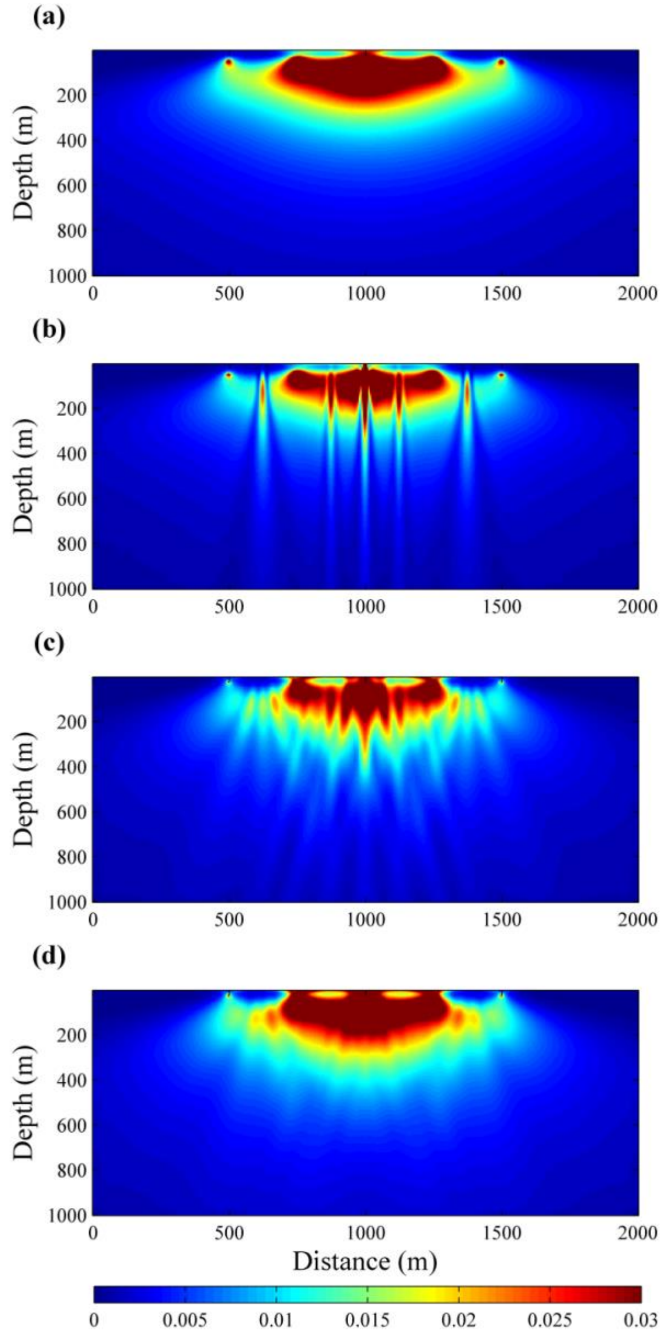
Phase Encoded Hessian

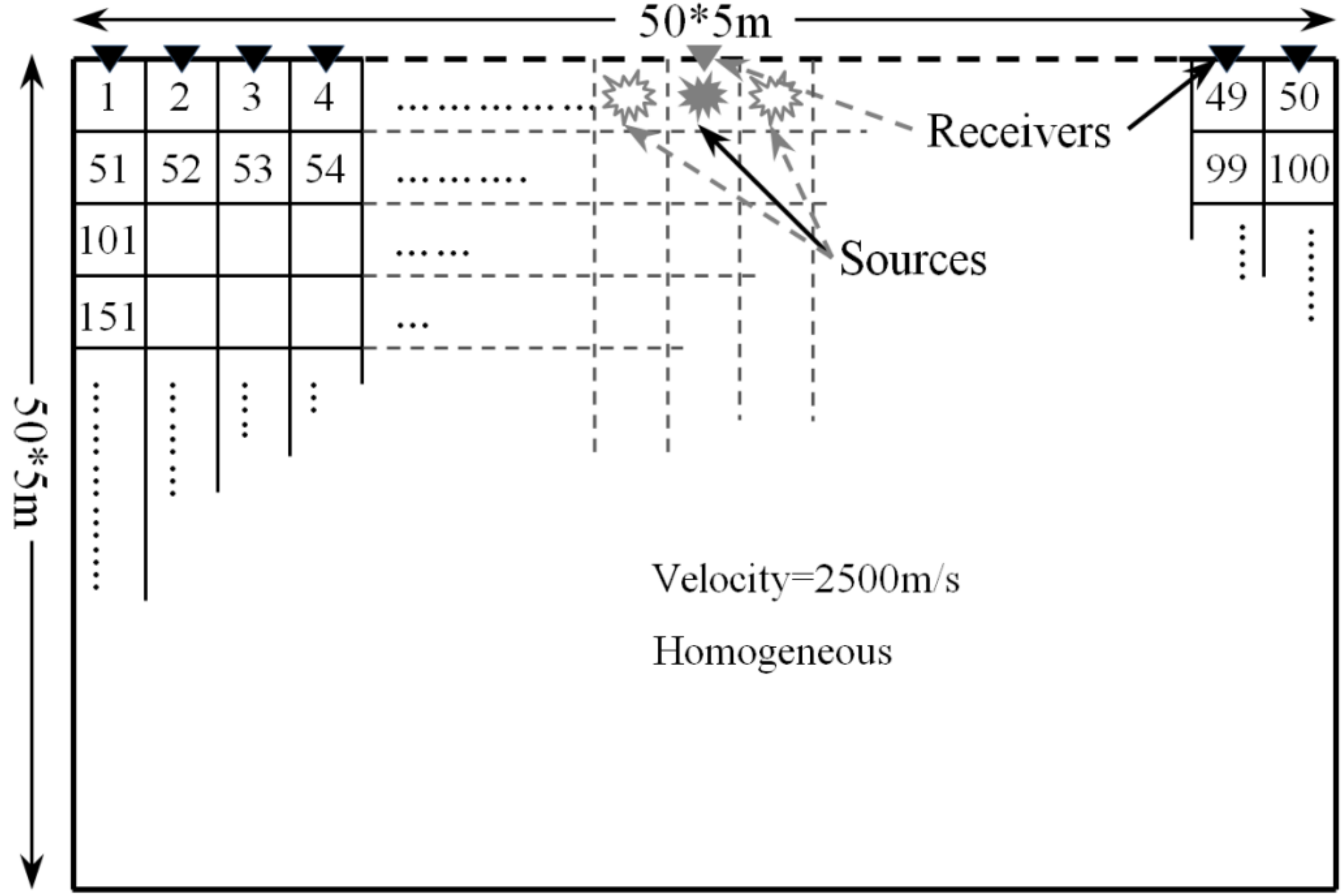
$$H_{chirp\_encoded} = \sum_{\mathbf{r}_s} \int d\omega \Re \left\{ \omega^4 G(\mathbf{r}', \mathbf{r}_s, \omega) G^*(\mathbf{r}'', \mathbf{r}_s, \omega) \right\}$$
$$\times \sum_{\mathbf{p}_g} \int d\omega \Re \left\{ G(\mathbf{r}', \mathbf{r}'_g, \omega) G^*(\mathbf{r}'', \mathbf{r}_g, \omega) e^{i\omega(p_g + \varepsilon \triangle p)(x'_g - x_g)} \right\}$$

Chirp phase encoded Hessian



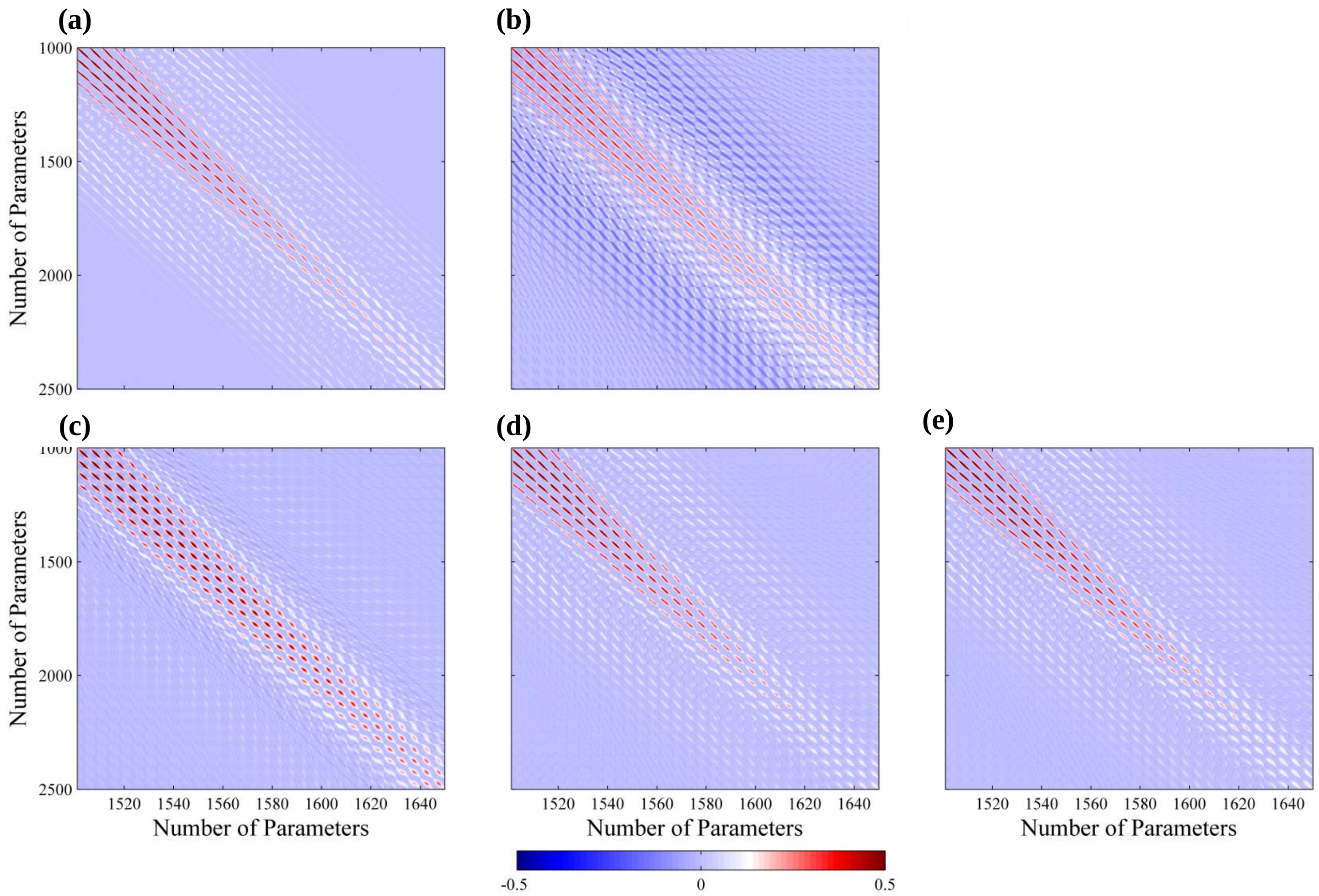
Chirp phase encoding strategy


(a)
(b)
(c)
(d)
Depth (m)
Distance (m)
0
500
1000
1500
2000
200
400
600
800
1000
0.005
0.01
0.015
0.02
0.025
0.03


50*5m
1
2
3
4
49
50
51
52
53
54
99
100
101
151
Receivers
Sources
50*5m
Velocity=2500m/s
Homogeneous

Hessian Approximations Comparison

CREWES

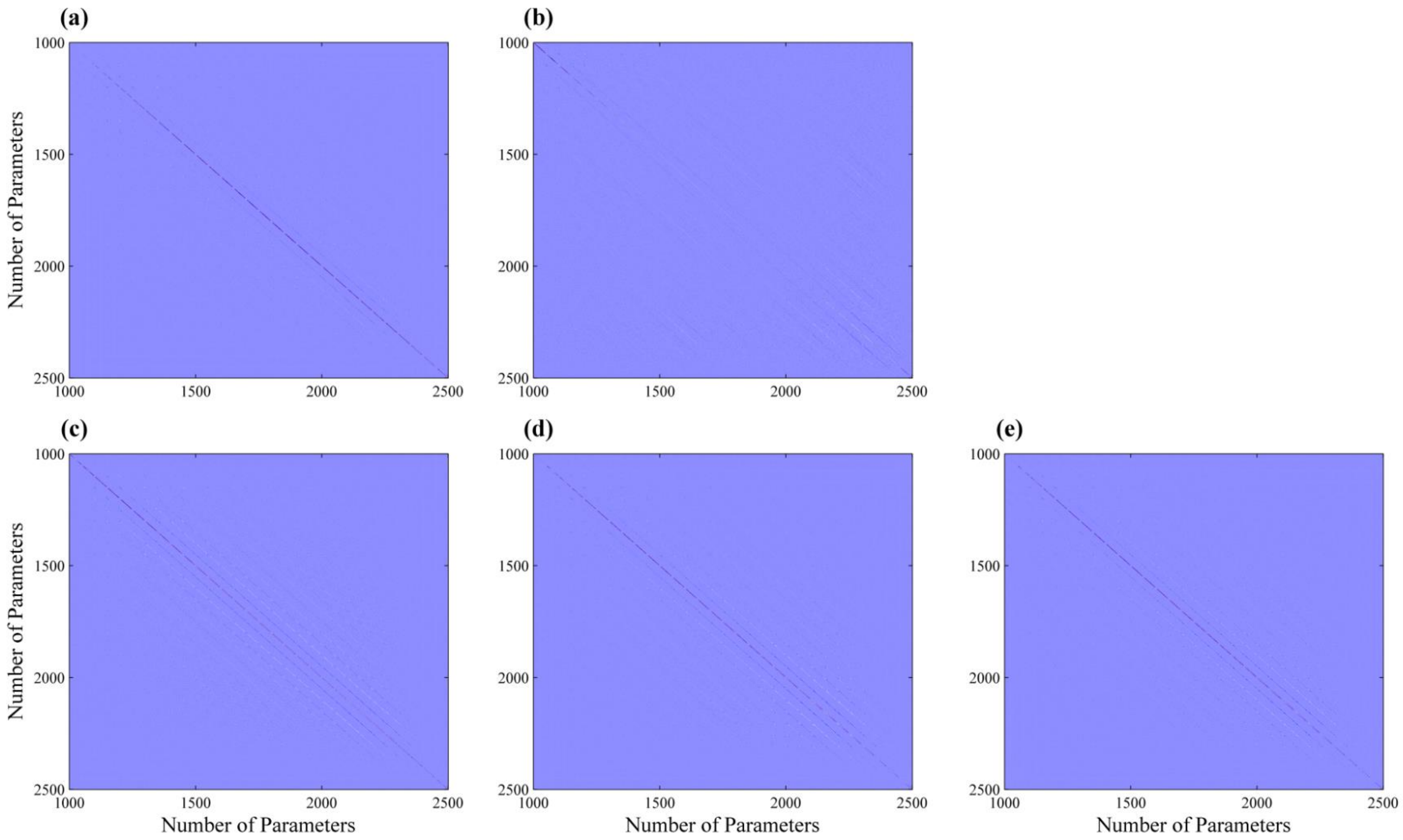


UNIVERSITY OF CALGARY

GeoScience

Inverse Hessian Comparison

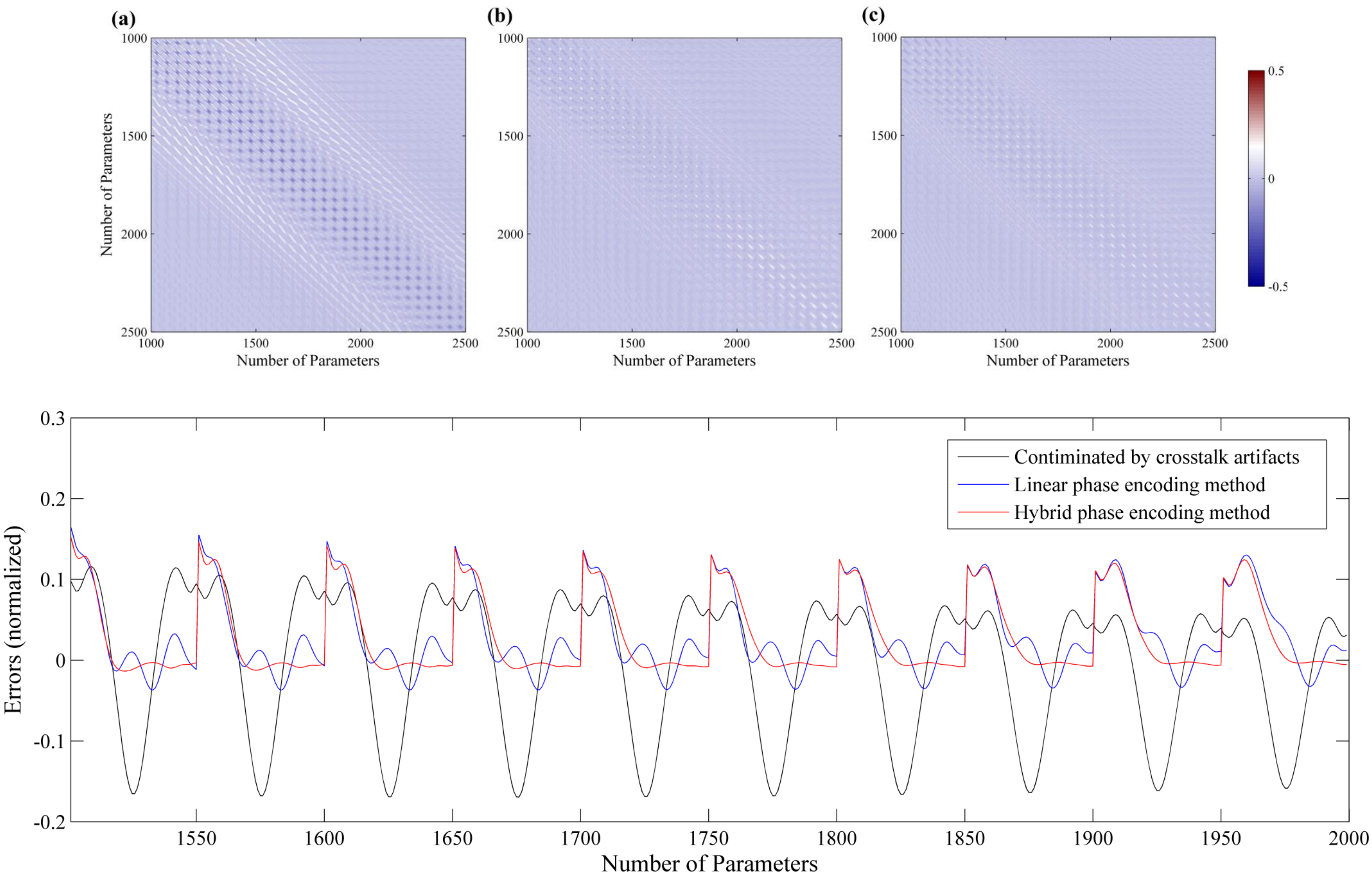
CREWES



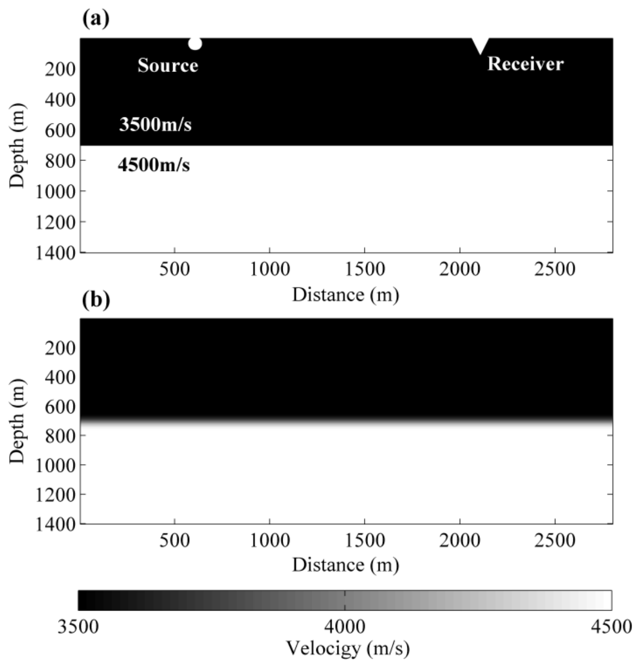
UNIVERSITY OF CALGARY

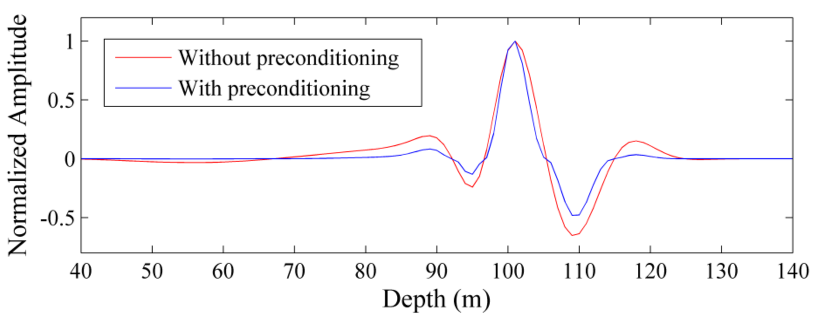
GeoScience

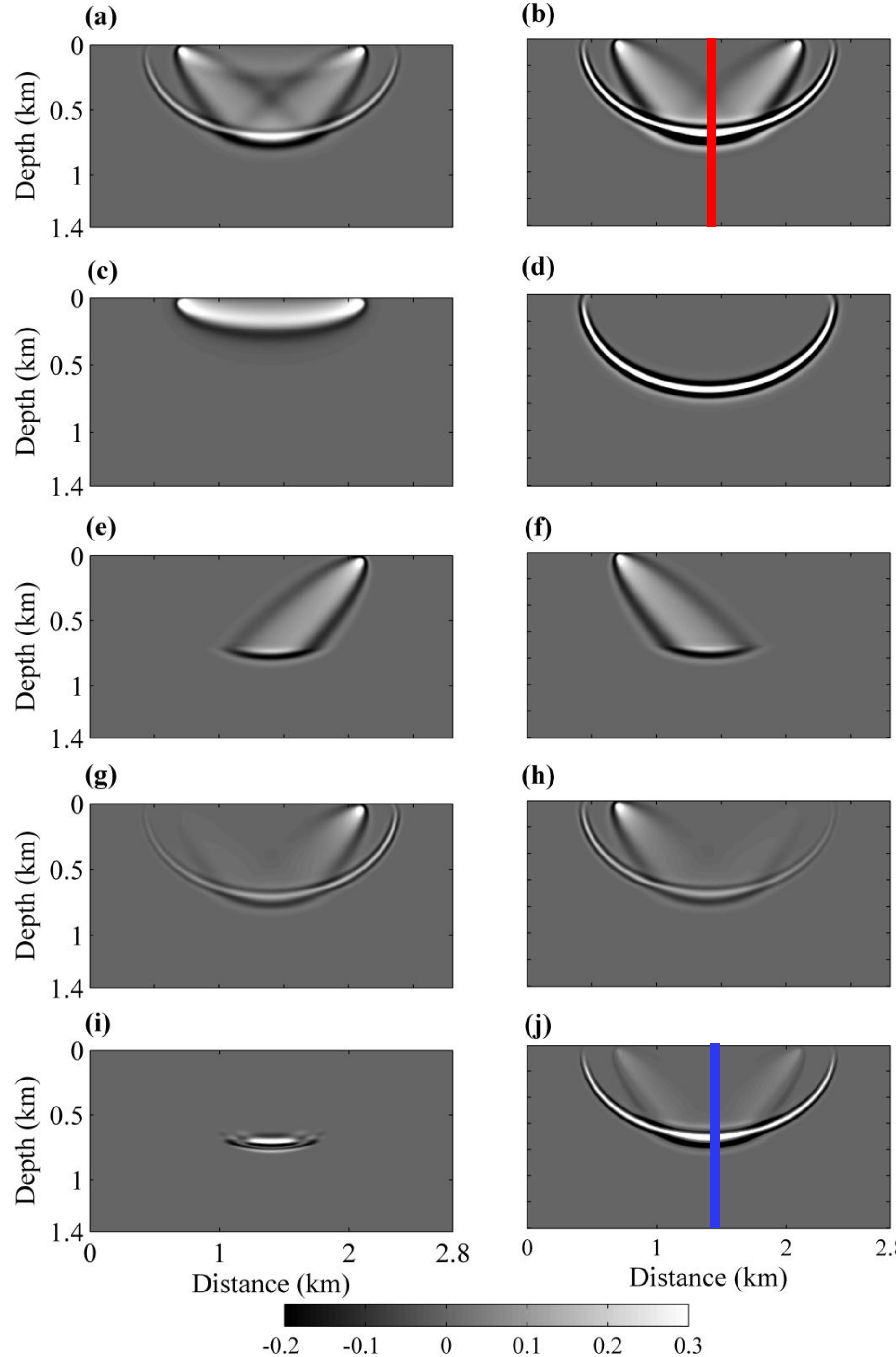
Error Comparison



Gradient Contribution Analysis

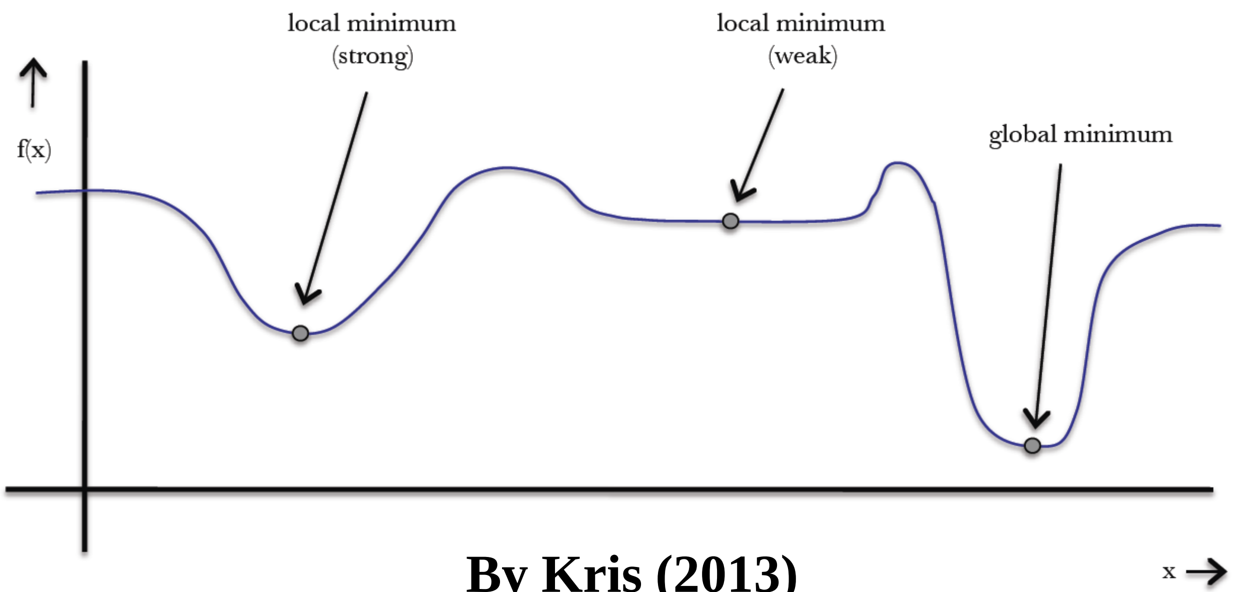






UNIVERSITY OF CALGARY

Multiscale Approach



By Kris (2013)

Multiscale Approach

- **Low frequency is responsible to catch the low wavenumber component**
- **High frequency is responsible to add detailed information**

$$s(\mathbf{r}) = s(\mathbf{r})^{low} + s(\mathbf{r})^{high}$$

Pseudo Gauss-Newton Step

To reduce the computational cost further, we proposed to use one ray parameter in one FWI iteration but change the ray parameter for different iterations.

$$\delta s\left(\mathbf{r}\right)=\frac{\int d\omega\Re\left\{\omega^{2}\mathcal{F}_{s}(\omega)\tilde{G}(\mathbf{r},\mathbf{r}_{s},\omega)G(\mathbf{r}_{g},\mathbf{r},\omega)\delta P^{*}\right\}}{diag\left(H_{chirp\_encoded}\right)+\lambda I}$$

Pseudo Code for PGN method

BEGIN $\leftarrow s_0$, initial model;
WHILE $\varepsilon \leq \varepsilon_{min}$ or $n \leq n_{max}$

Identify the ray parameter $p_s^{(n)}$

Identify the frequency band $f^{(n)} = f_0 \rightarrow f_{max}, f_{interval}$, every k iterations

Generate the data residual δP and apply low-pass filtering $\delta \tilde{P} = \textbf{low\_pass}\left(\delta P, f^{(n)}\right)$

Generate the linear phase encoded gradient $g^{(n)}\left(p_s^{(n)}\right)$

FOR $i = 1$ to $\mathbf{p}_s^H, \mathbf{p}_r^H$, every 1 or m iterations

Construct the diagonal part of the hybrid phase encoded Hessian $diag\left(H_{en\_a}^{(n)}\right)$

END FOR

Calculate the step length $\mu^{(n)}$ using the line search method

update the velocity model:

$$s^{(n+1)}(\mathbf{r}) = s^{(n)}(\mathbf{r}) - \mu^{(n)} \left\{ diag\left(H_{en\_a}^{(n)}\right) + \lambda I \right\}^{-1} g^{(n)}\left(p_s^{(n)}\right)$$

Calculate the relative least-squares error:

$$\varepsilon = \frac{\|s^{(n)}(\mathbf{r}) - s^{true}(\mathbf{r})\|_2}{\|s^{true}(\mathbf{r})\|_2}$$

END WHILE

CREWES

UNIVERSITY OF CALGARY

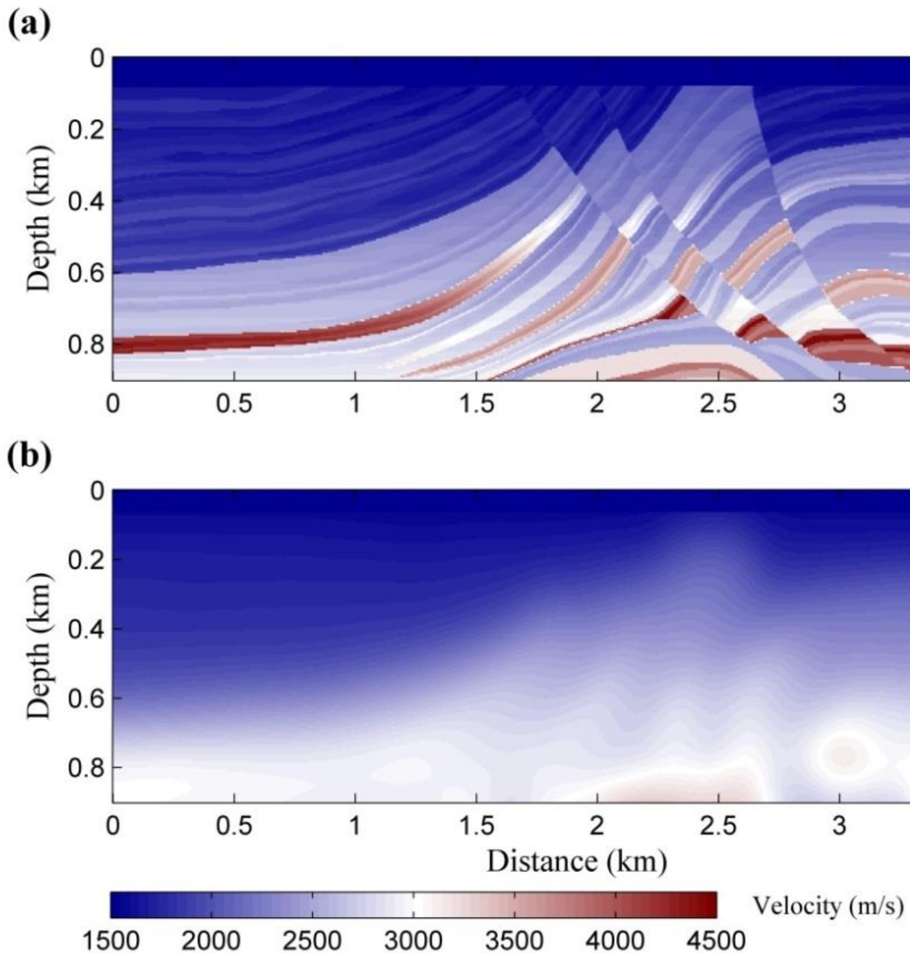
GeoScience

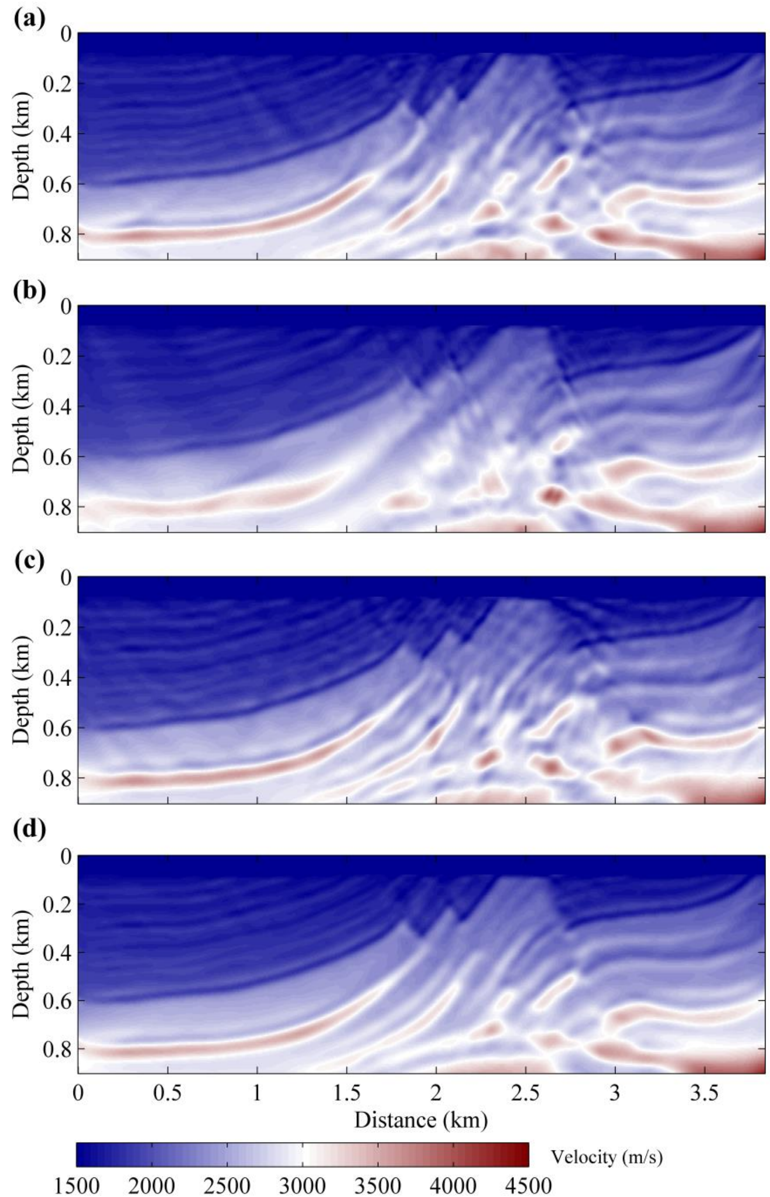
Computational Cost Comparison

Table2. Computational cost comparison for different strategies

Methods	Gradient	H_a	$diag\,(H_{en\_a})$	Step length	Cost for one iteration
TGN Method	$2N_s$	$N_s \times N_r$	\	1	$2N_s + N_s \times N_r$
SEGN Method	$2N_p^g$	\	$N_{ps}^H + N_{pr}^H$	1	$2N_p^g + N_{ps}^H + N_{pr}^H$
PGN Method	2	\	$N_{ps}^H + N_{pr}^H$	1	$N_{ps}^H + N_{pr}^H + 2$

Numerical Experiment





$p = 0s/km$

$p = -0.2s/km$

$p = 0.2s/km$

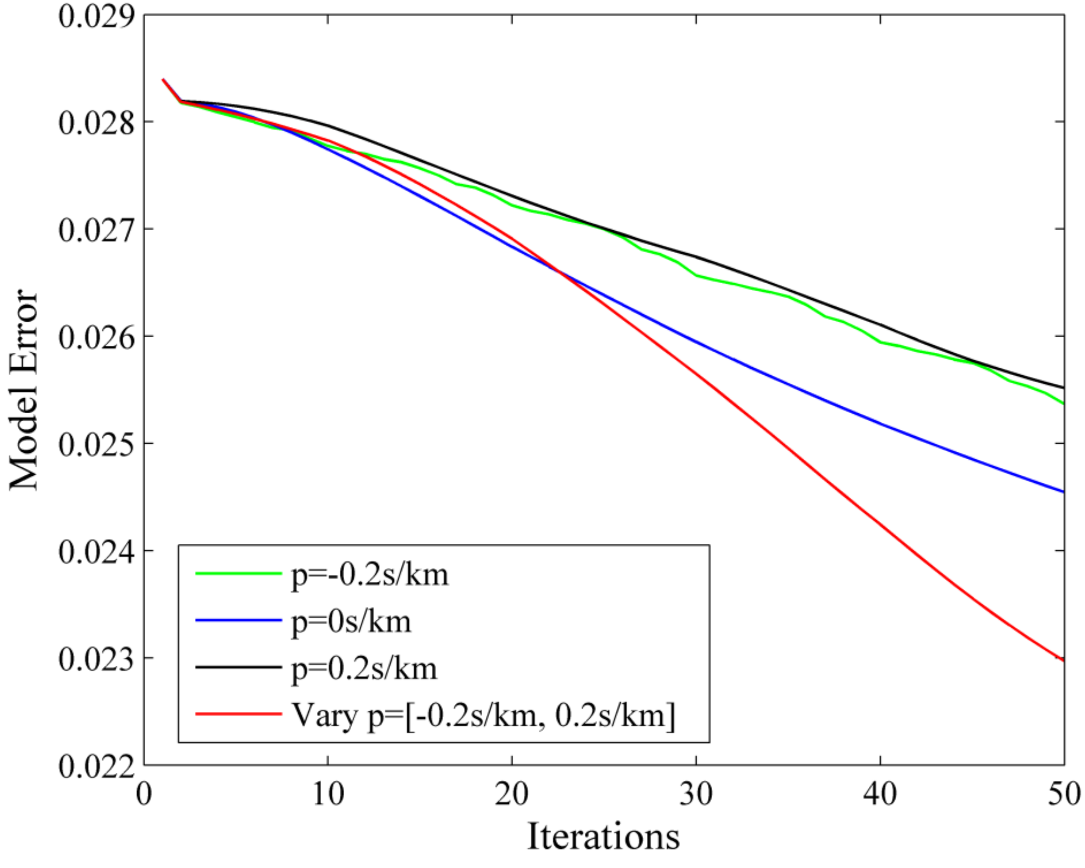
$p = [-0.2s/km, 0.2s/km]$

CREWES

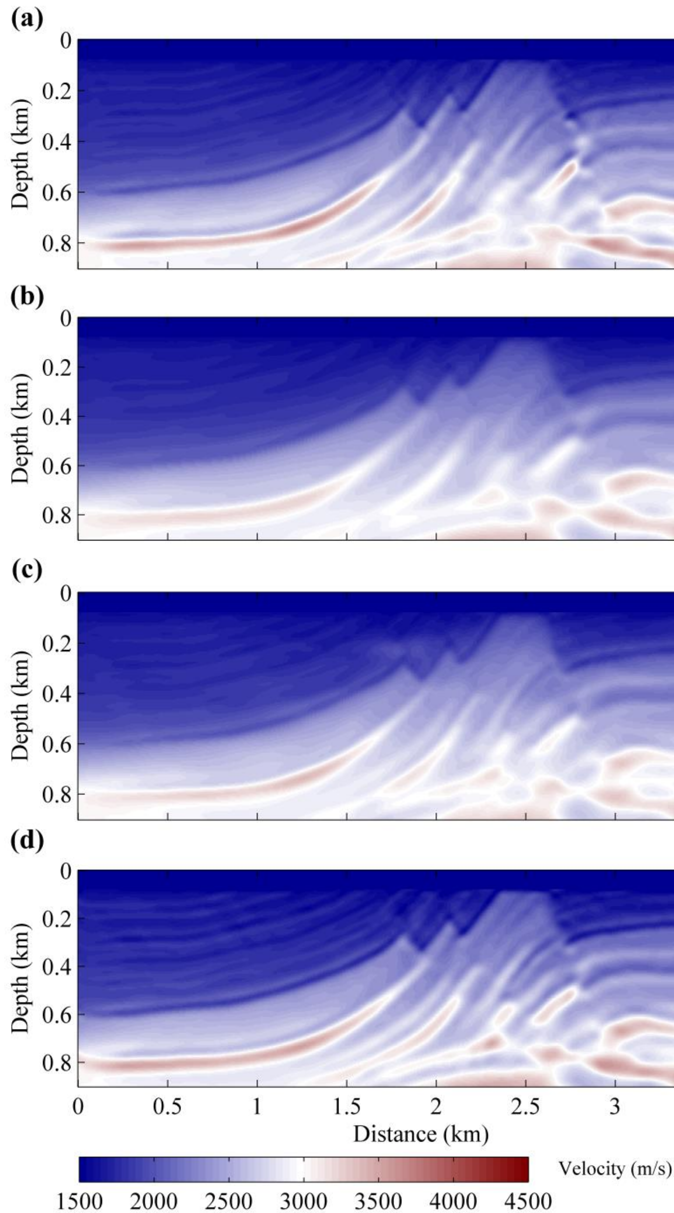
UNIVERSITY OF CALGARY

Effects for Fixed and Varying Ray Parameter

GeoScience



Effects for Varying Ray Parameter



$$p = [-0.1s/km, 0.1s/km]$$
$$step = 0.02s/km$$

$$p = [-0.3s/km, 0.3s/km]$$
$$step = 0.06s/km$$

$$p = [-0.6s/km, 0.6s/km]$$
$$step = 0.12s/km$$

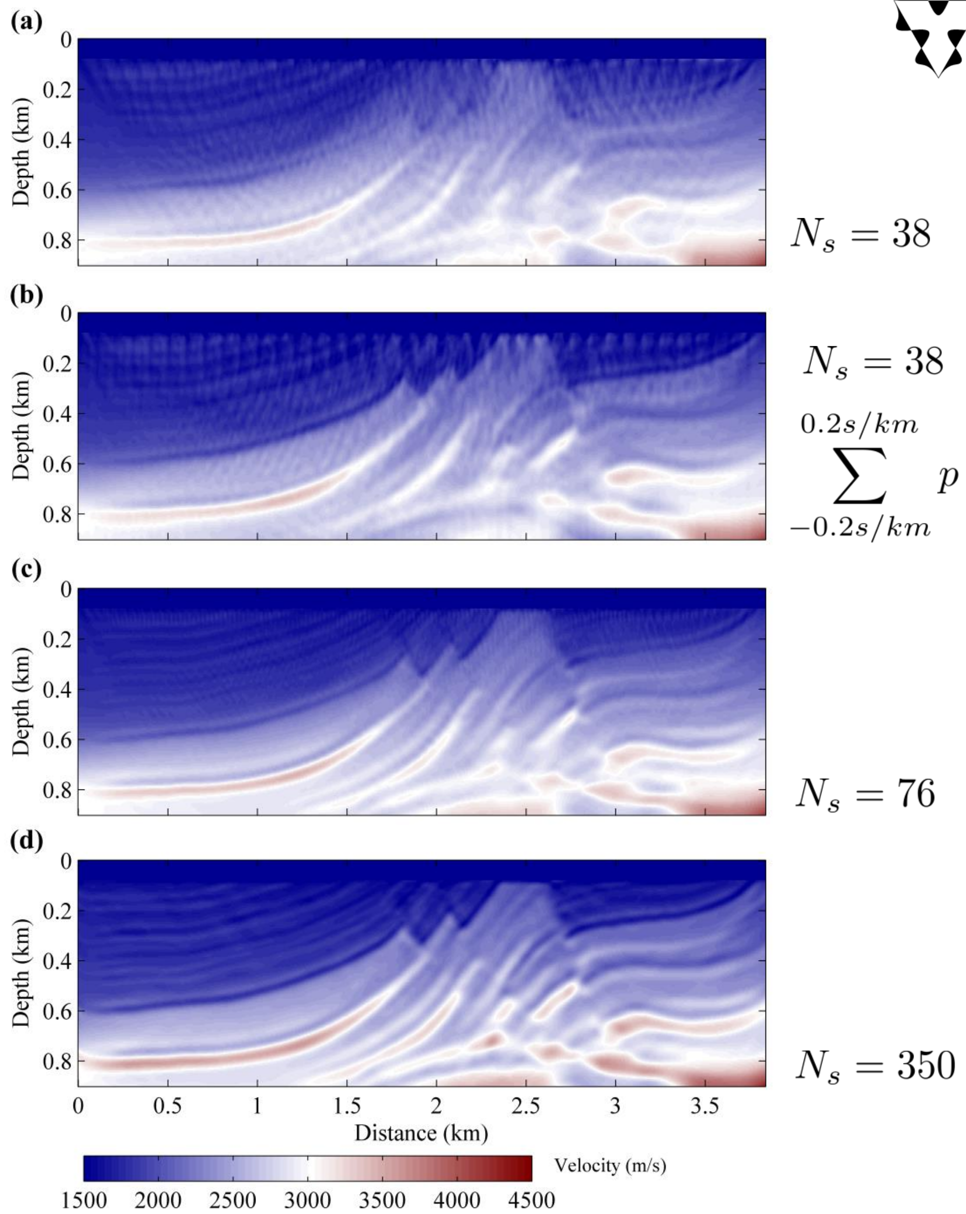
$$p = [-0.2s/km, 0.2s/km]$$
$$step = 0.04s/km$$

CREWES

UNIVERSITY OF CALGARY

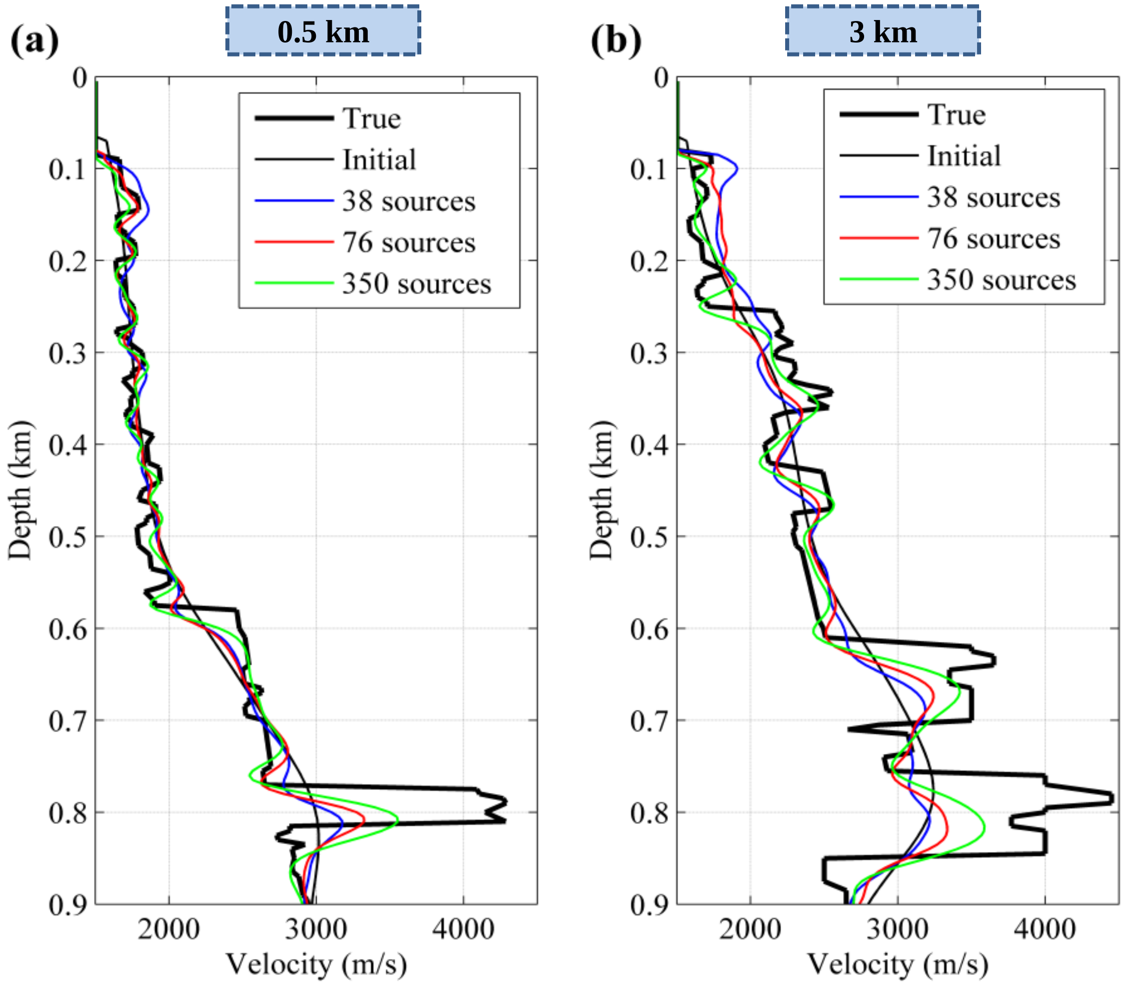
Sensitivity to the Ray-Parameter Range

GeoScience



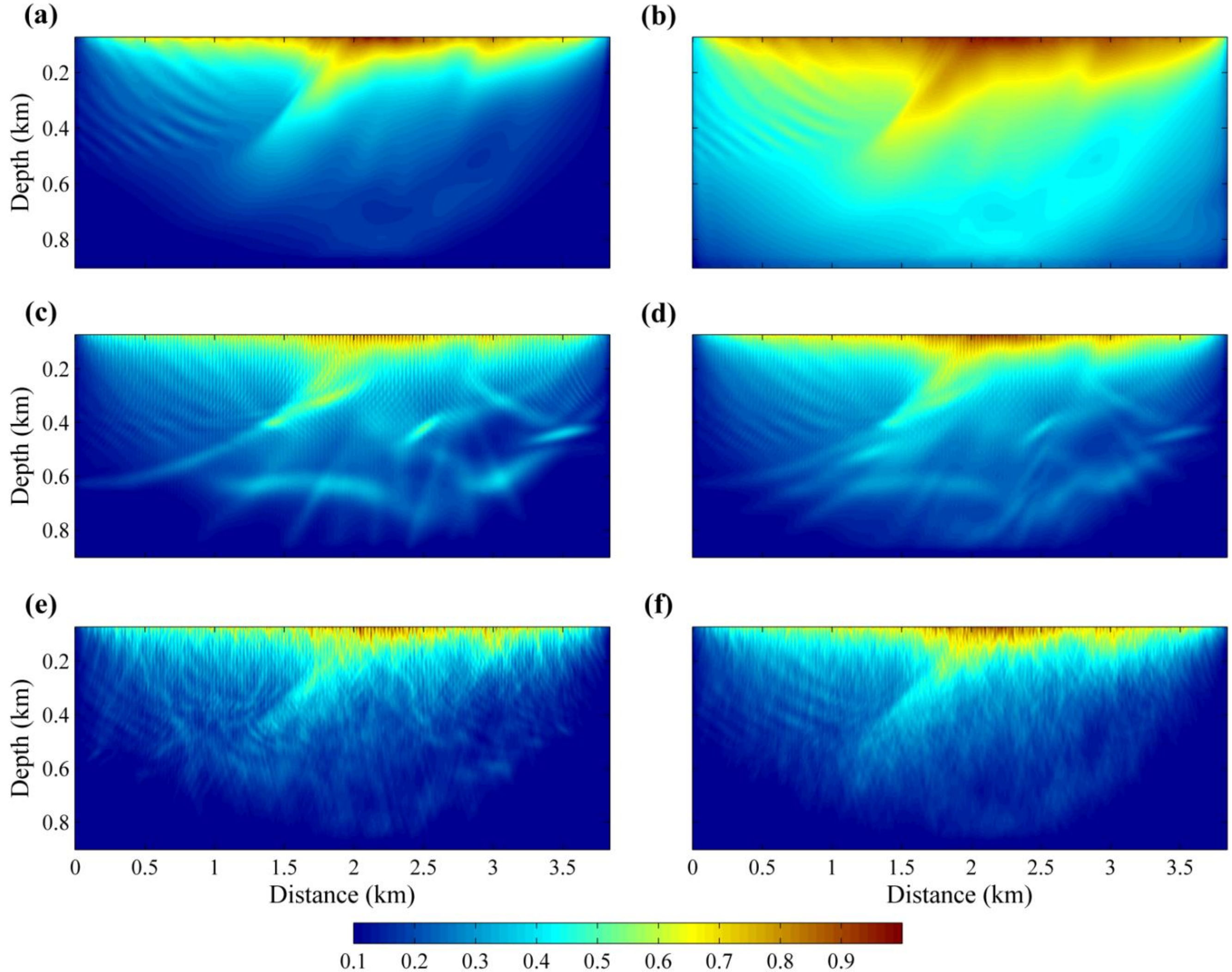
UNIVERSITY OF CALGARY

Sensitivity to the Encoded Sources

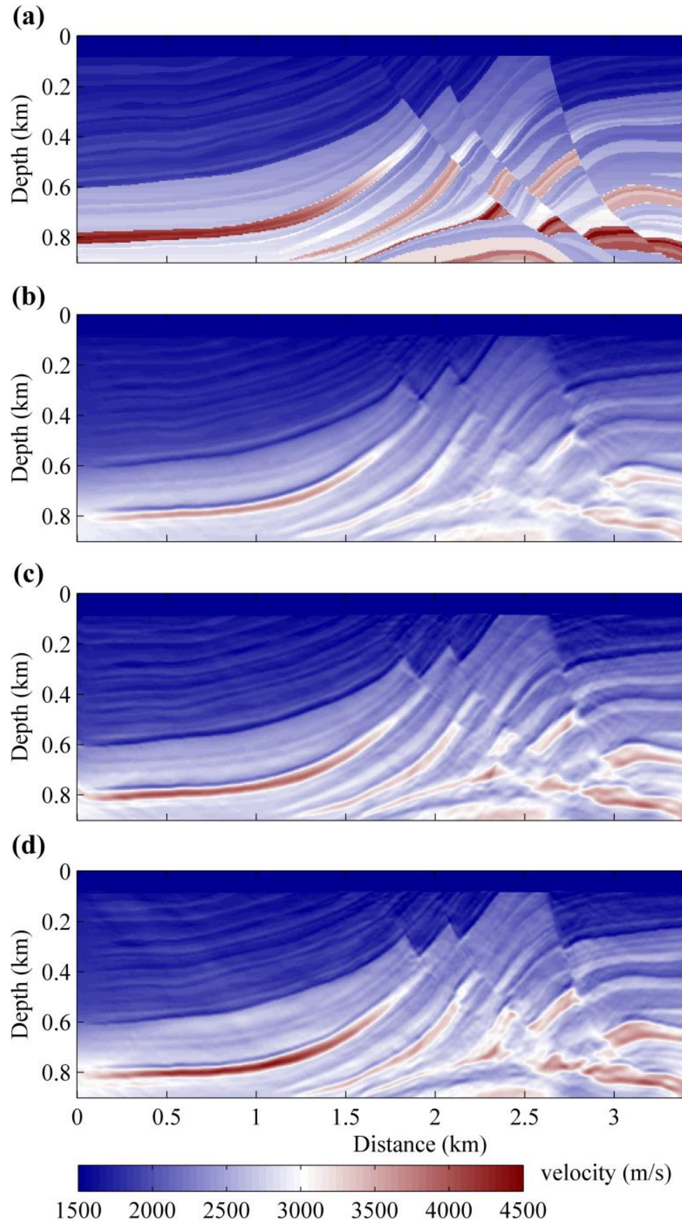


Sensitivity to the Encoded Sources

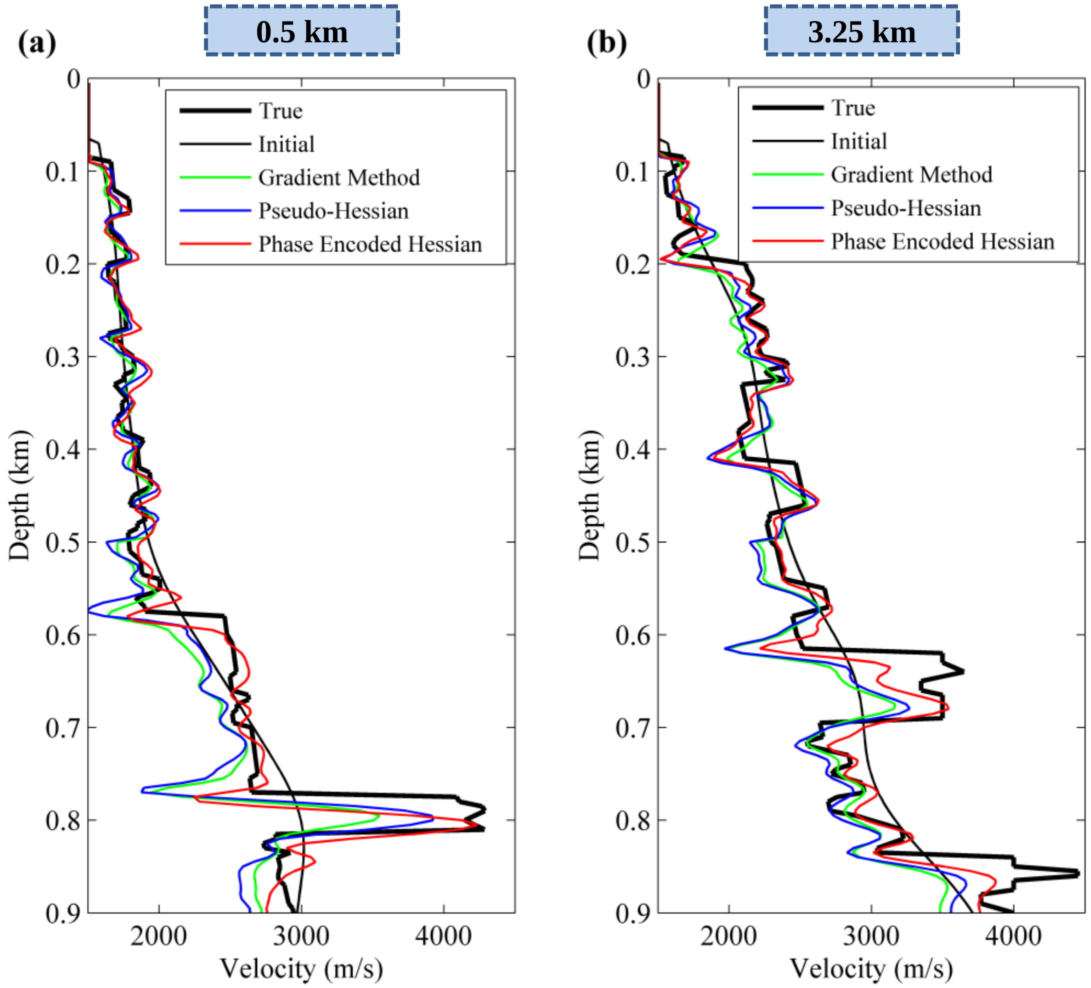
Initial Velocity Model



Different Hessian Approximations



Different Scaling Methods for FWI



Different Scaling Methods for FWI

Reverse Time Migration Image Comparison

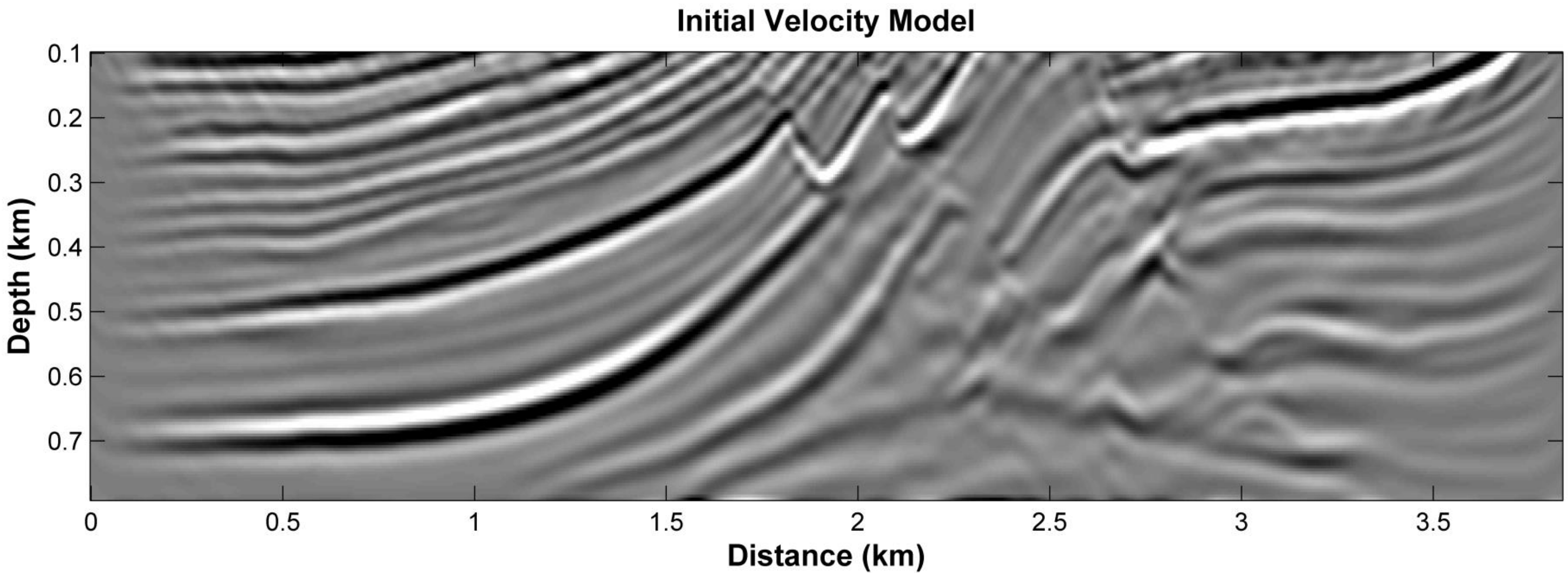


Image By Initial Velocity Model

Reverse Time Migration Image Comparison

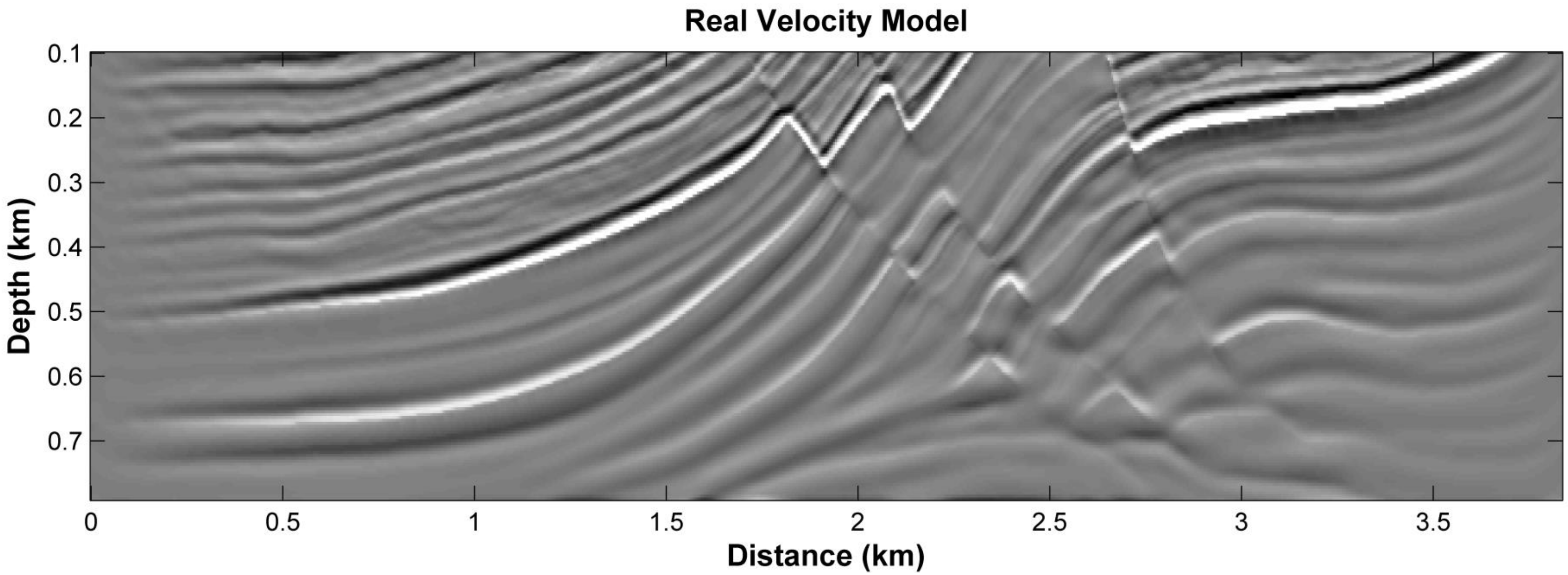


Image By True Velocity Model

Reverse Time Migration Image Comparison

CREWES

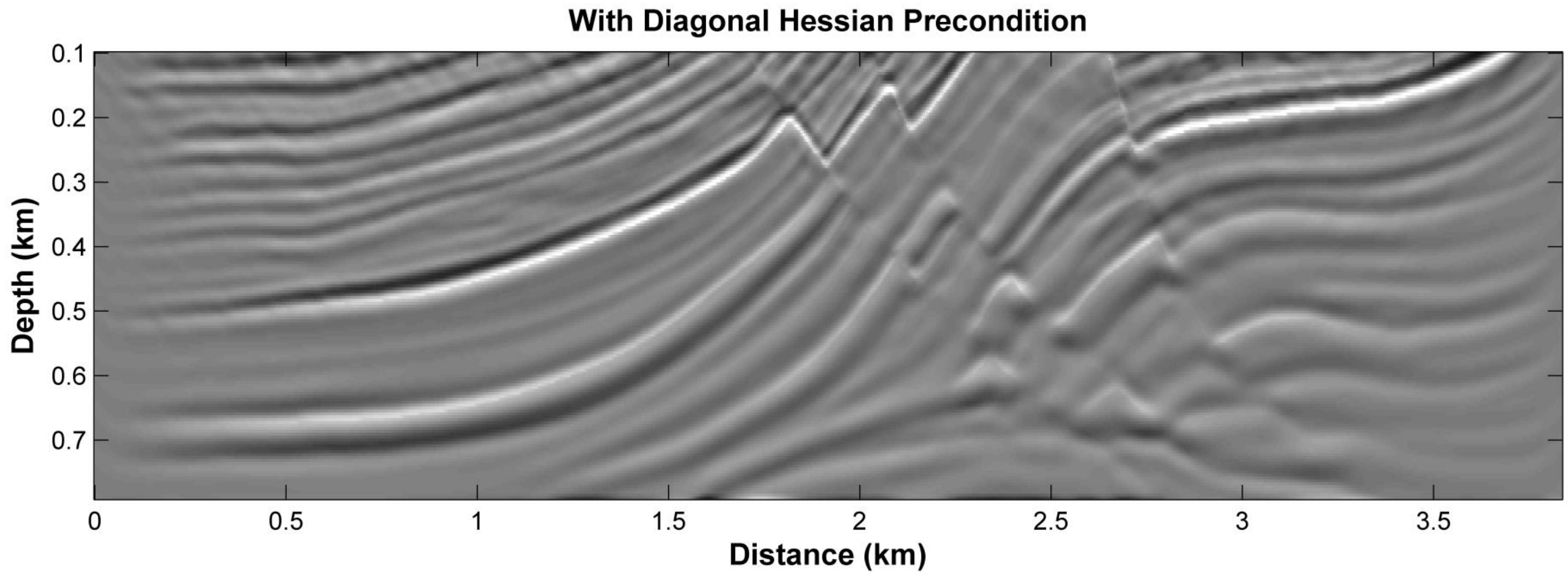


Image By Inverted Velocity Model

UNIVERSITY OF CALGARY

GeoScience

Conclusions

- Hessian matrix server as a nonstationary deconvolution operator to improve the convergence rate of least-squares inverse problem.
- Varying ray-parameter during iterations can reduce the computational cost further and balance the model update.
- If the ray-parameter range is too small, the layers with dip angles cannot be inversed in balance, if the ray-parameter range is too large, the convergence rate will be decreased.
- If the assembled sources are not dense enough, the crosstalk noise will be very obvious, especially for shallow layers.
- Chirp phase encoding strategy can reduce the crosstalk noise better than linear phase encoding strategy with the same number of simulations.
- Diagonal part of the phase encoded Hessian can server as a good approximation of the Hessian to precondition the gradient and increase the convergence rate.
- Full waveform inversion with source encoding is efficient for numerical modeling but will increase difficulties in seismic data acquisition and preprocessing.

Acknowledgements

- All CREWES Sponsors
- Danping, Wenyuan, Vladimir Zubov
- Joe Wong, Raul, Babatunde

Thank You !