Nonlinear FWI:
formulation and examples

Yu Geng, Kris Innanen

Dec 2nd, 2016

Banff, AB

Outline

- Introduction
  - General principle of FWI
  - Nonlinearity of FWI
- Theory
  - Sensitivities: from linear to nonlinear
  - Determining model perturbation before n+1th iteration
  - Two-iteration nonlinear FWI
- Examples
- Conclusions
- Acknowledgements

General principle of FWI

- Forward modeling

$$\left[\omega^2 s(\mathbf{r}) + \nabla^2\right] P(\mathbf{r}, \mathbf{r}_s, \omega) = -\delta(\mathbf{r} - \mathbf{r}_s)$$

- Misfit function

$$\phi(s) = \frac{1}{2} \sum_{\mathbf{r}_s} \sum_{\mathbf{r}_g} \sum_{\omega} \left\| \underbrace{\boxed{P(\mathbf{r}_g, \mathbf{r}_s, \omega)}}_{\text{data}} - \underbrace{\boxed{G(\mathbf{r}_g, \mathbf{r}_s, \omega \mid s_n)}}_{\text{calculated data}} \right\|^2$$

$$= \frac{1}{2} \sum_{\mathbf{r}_s} \sum_{\mathbf{r}_g} \sum_{\omega} \left\| \underbrace{\boxed{\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega \mid s_n)}}_{\text{data residual}} \right\|^2$$

General principle of FWI

- Gradient of misfit

$$g_n(\mathbf{r}) = -\sum_{\mathbf{r}_s}\sum_{\mathbf{r}_g}\sum_{\omega}\mathrm{Re}\left(\boxed{\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s,\omega\,|\,s_n)}{\partial s(\mathbf{r})}}\,\delta P^*(\mathbf{r}_g,\mathbf{r}_s,\omega\,|\,s_n)\right)$$

sensitivity

- Update model iteratively

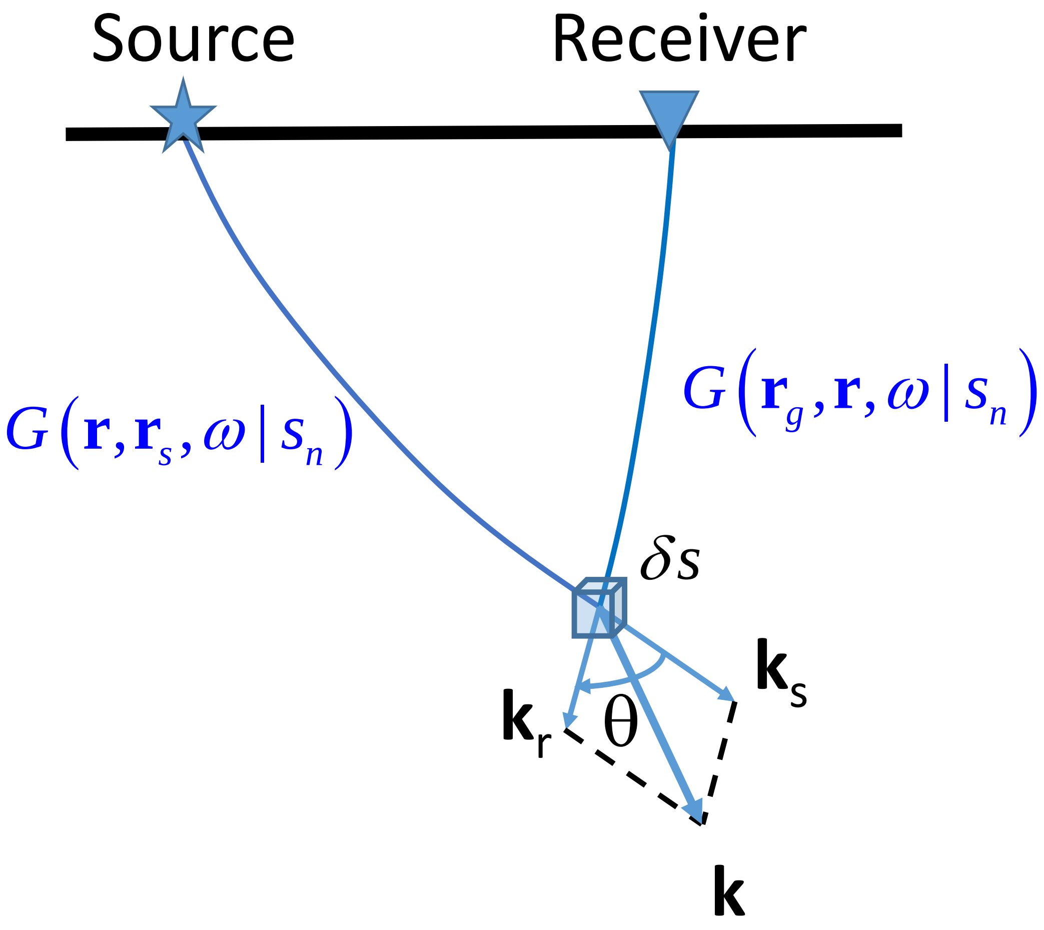
$$s_{n+1}(\mathbf{r}) = s_n(\mathbf{r}) - \mu_n g_n(\mathbf{r})$$

Nonlinearity of FWI

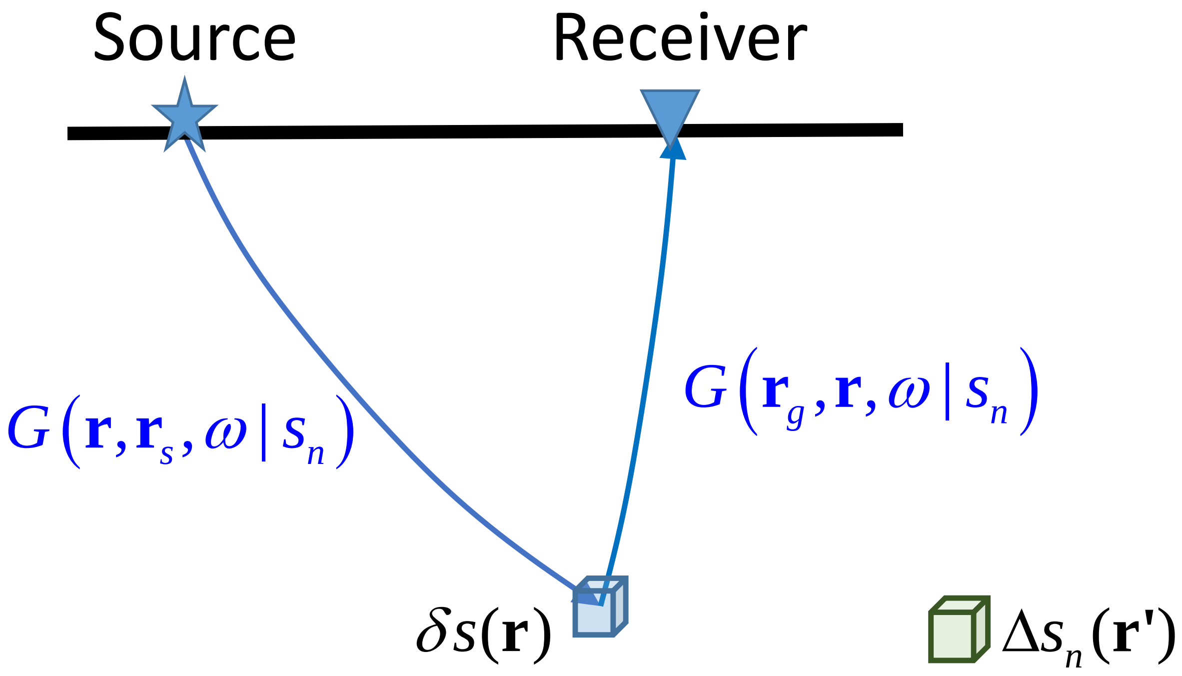
- FWI relies on the diffraction tomography principle

$$\mathbf{k} = \frac{2\omega}{v}\cos\left(\frac{\theta}{2}\right)\mathbf{n}$$

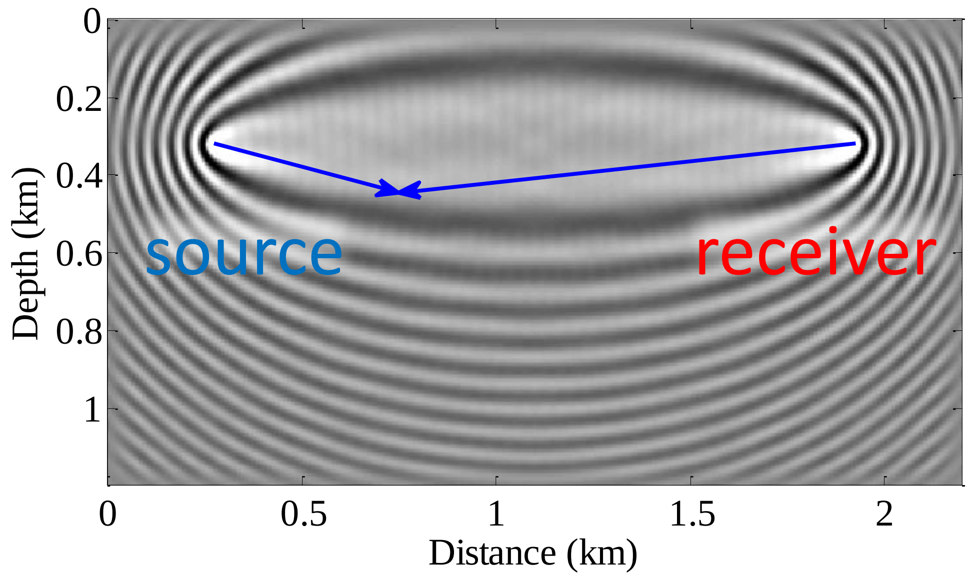
  - One frequency and one aperture map one wavenumber in the model space
  - Low frequencies and wide apertures data help to recover long-to-intermediate wavelengths structures
- Nonlinearity in FWI
  - Multiscale FWI
  - Build good starting model
  - Nonlinear sensitivities



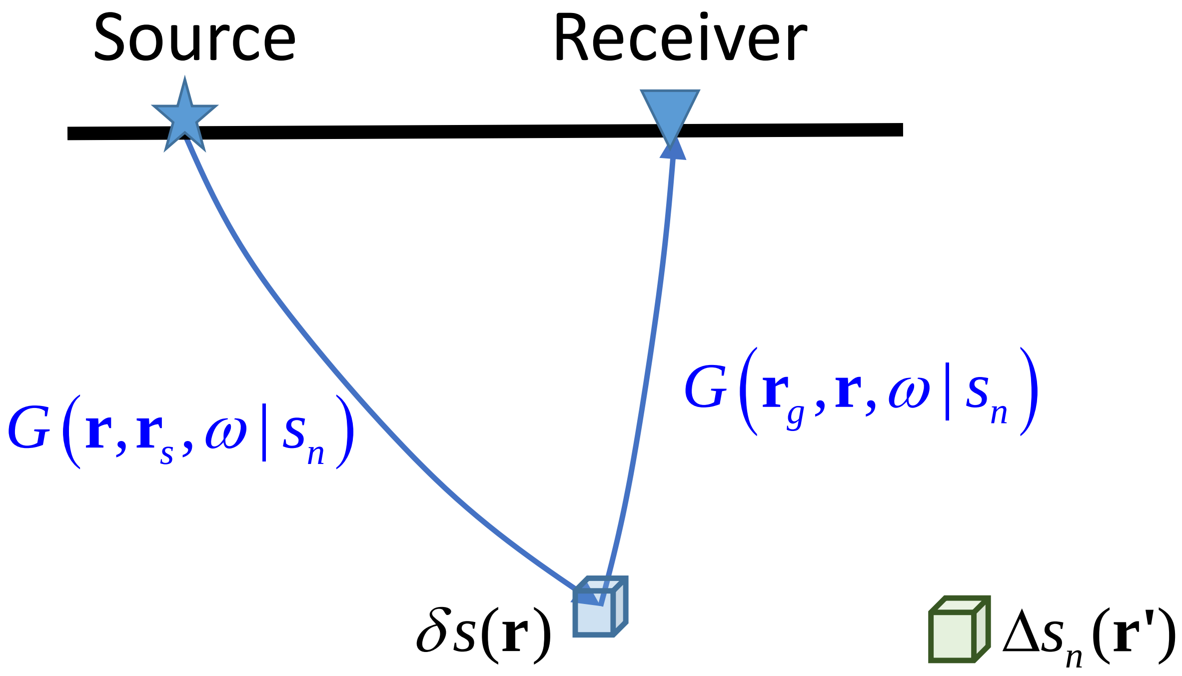
Sensitivities: from linear to nonlinear



$$\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s,\omega|s_n)}{\partial s(\mathbf{r})} = \omega^2 G(\mathbf{r}_g,\mathbf{r},\omega|s_n)\, G(\mathbf{r},\mathbf{r}_s,\omega|s_n)$$



Sensitivities: from linear to nonlinear

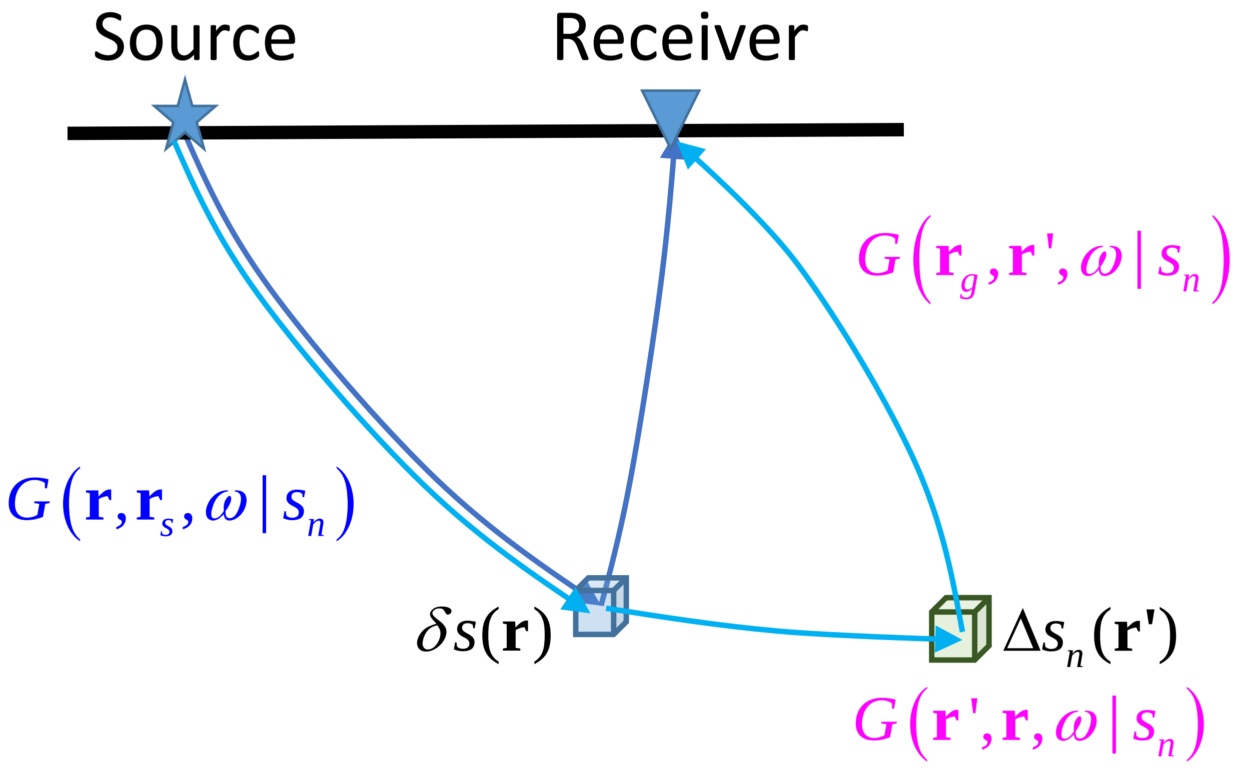


Zero-order sensitivity

$$\left(\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s,\omega\,|\,s_{n+1})}{\partial s(\mathbf{r})}\right)_0 = \frac{\partial G(\mathbf{r}_g,\mathbf{r}_s,\omega\,|\,s_n)}{\partial s(\mathbf{r})}$$

$$= \omega^2 G(\mathbf{r}_g,\mathbf{r},\omega\,|\,s_n)\,G(\mathbf{r},\mathbf{r}_s,\omega\,|\,s_n)$$

Sensitivities: from linear to nonlinear



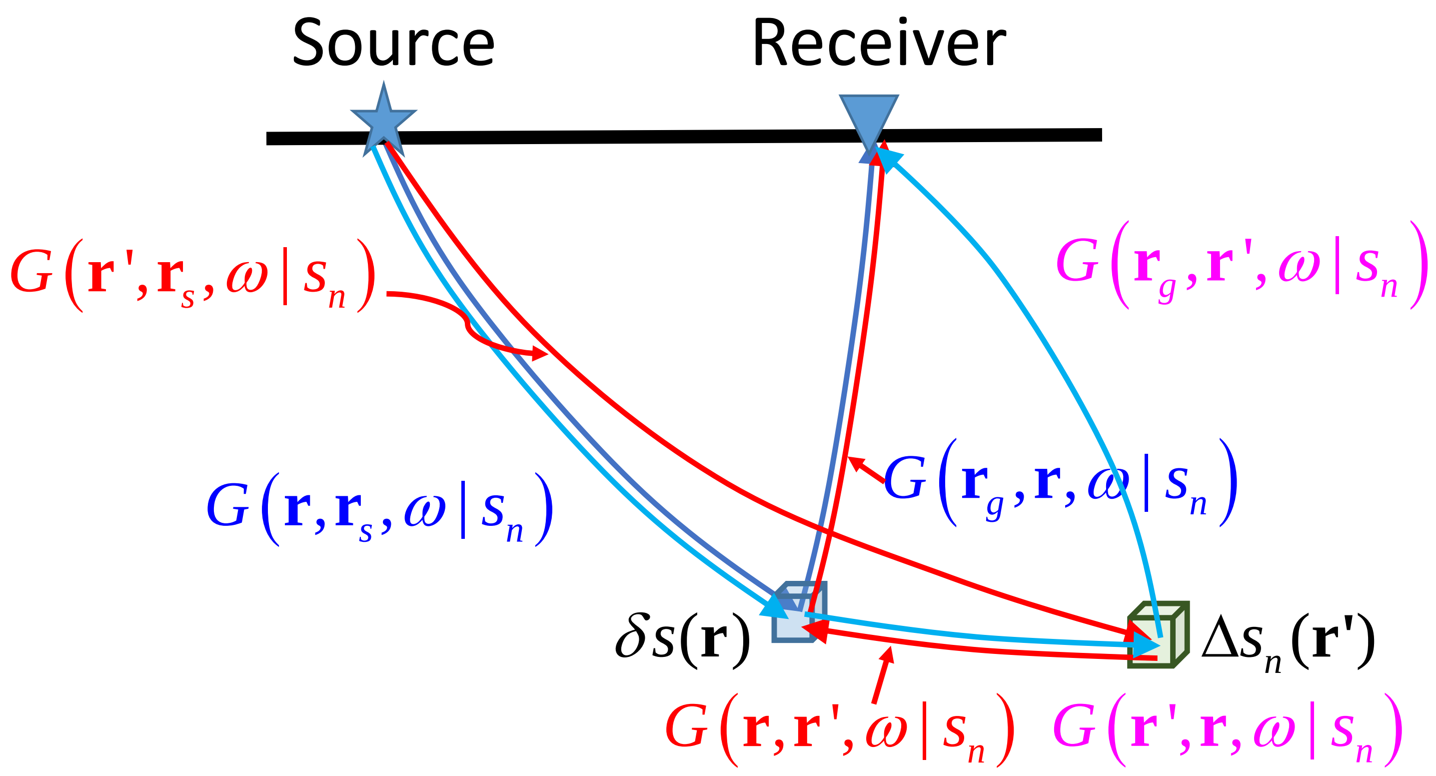
Zero-order sensitivity

$$\left(\frac{\partial G\left(\mathbf{r}_g,\mathbf{r}_s,\omega \mid s_{n+1}\right)}{\partial s\left(\mathbf{r}\right)}\right)_0 = \frac{\partial G\left(\mathbf{r}_g,\mathbf{r}_s,\omega \mid s_n\right)}{\partial s\left(\mathbf{r}\right)}$$

$$= \omega^2 G\left(\mathbf{r}_g,\mathbf{r},\omega \mid s_n\right) G\left(\mathbf{r},\mathbf{r}_s,\omega \mid s_n\right)$$

$$\delta G\left(\mathbf{r}_g,\mathbf{r} \mid s_n,\Delta s_n\right) \approx \omega^2 \int d\mathbf{r}' G\left(\mathbf{r}_g,\mathbf{r}' \mid s_n\right) G\left(\mathbf{r}',\mathbf{r} \mid s_n\right) \Delta s_n\left(\mathbf{r}'\right)$$

Sensitivities: from linear to nonlinear



Zero-order sensitivity

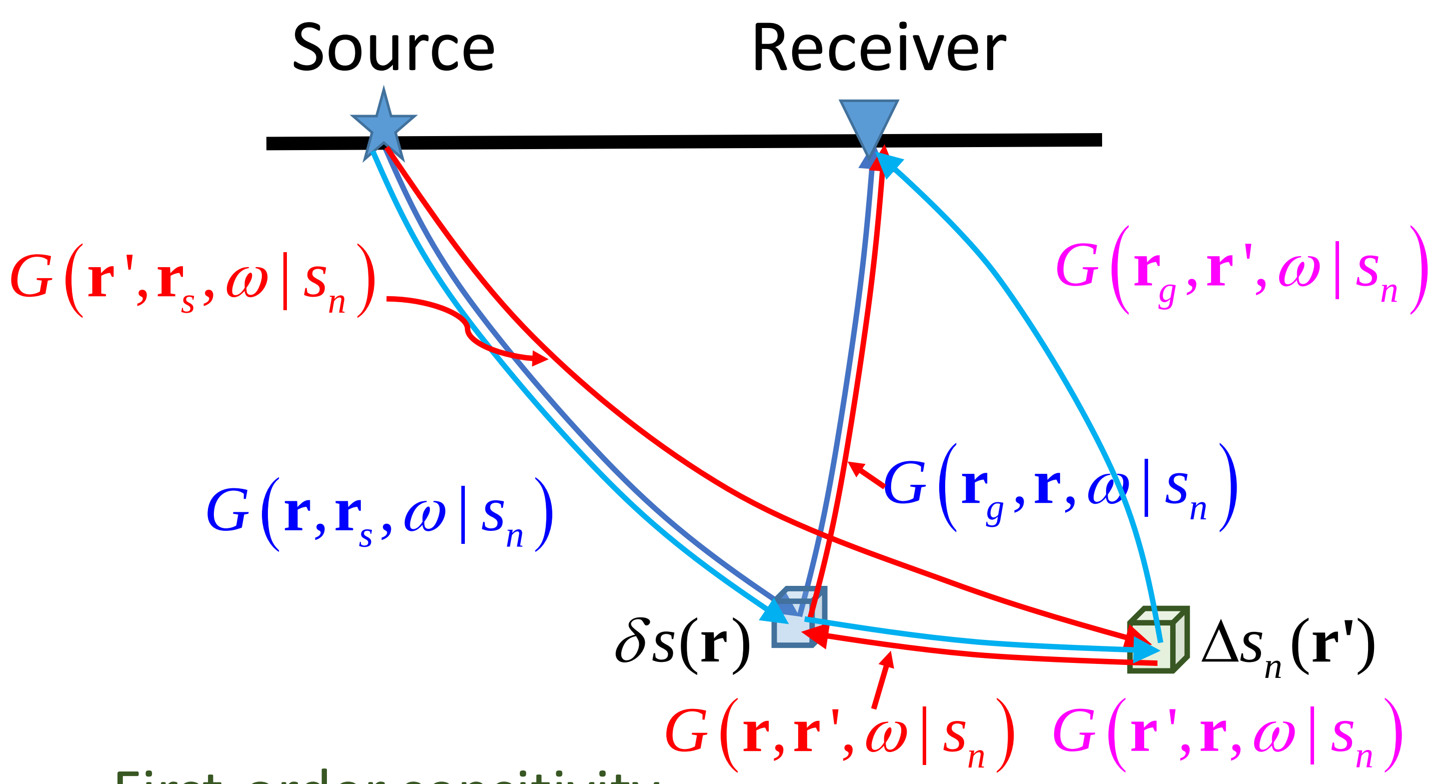
$$\left(\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s,\omega|s_{n+1})}{\partial s(\mathbf{r})}\right)_0 = \frac{\partial G(\mathbf{r}_g,\mathbf{r}_s,\omega|s_n)}{\partial s(\mathbf{r})}$$

$$= \omega^2 G(\mathbf{r}_g,\mathbf{r},\omega|s_n)\,G(\mathbf{r},\mathbf{r}_s,\omega|s_n)$$

$$\delta G(\mathbf{r}_g,\mathbf{r}|s_n,\Delta s_n) \approx \omega^2 \int d\mathbf{r}'\,G(\mathbf{r}_g,\mathbf{r}'|s_n)\,G(\mathbf{r}',\mathbf{r}|s_n)\,\Delta s_n(\mathbf{r}')$$

$$\delta G(\mathbf{r},\mathbf{r}_s|s_n,\Delta s_n) \approx \omega^2 \int d\mathbf{r}'\,G(\mathbf{r},\mathbf{r}'|s_n)\,G(\mathbf{r}',\mathbf{r}_s|s_n)\,\Delta s_n(\mathbf{r}')$$

Sensitivities: from linear to nonlinear



Zero-order sensitivity

$$\left(\frac{\partial G\left(\mathbf{r}_g,\mathbf{r}_s,\omega\,|\,s_{n+1}\right)}{\partial s\left(\mathbf{r}\right)}\right)_0 = \frac{\partial G\left(\mathbf{r}_g,\mathbf{r}_s,\omega\,|\,s_n\right)}{\partial s\left(\mathbf{r}\right)}$$

$$= \omega^2 G\left(\mathbf{r}_g,\mathbf{r},\omega\,|\,s_n\right) G\left(\mathbf{r},\mathbf{r}_s,\omega\,|\,s_n\right)$$

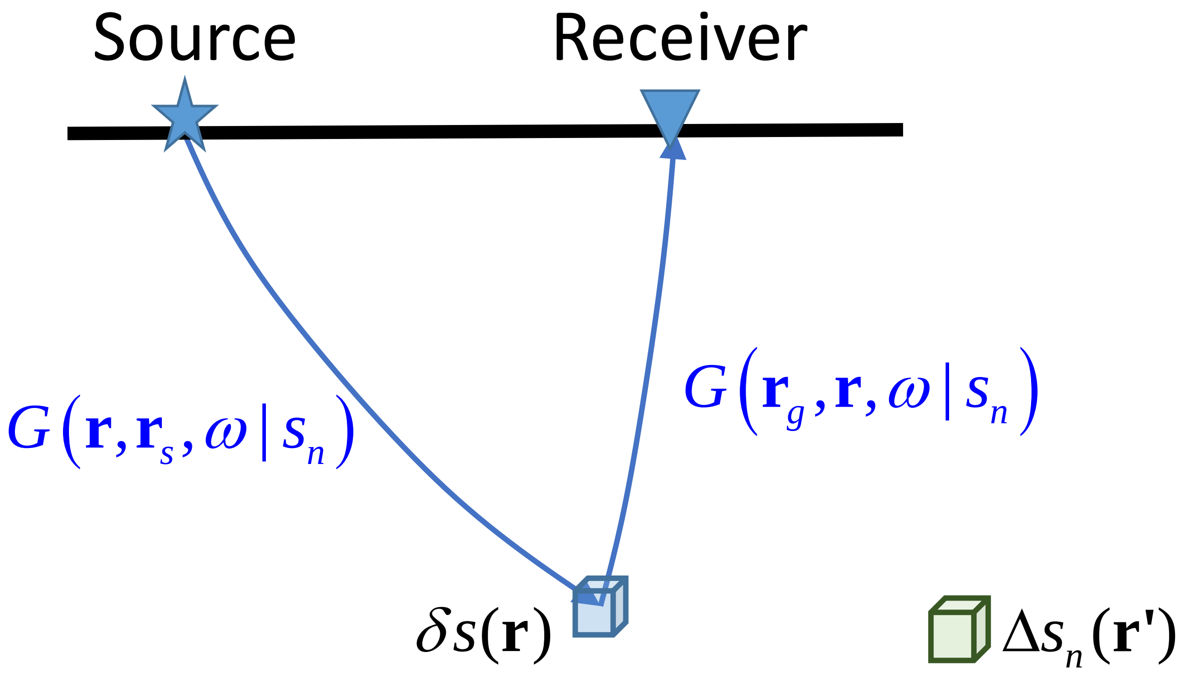
First-order sensitivity

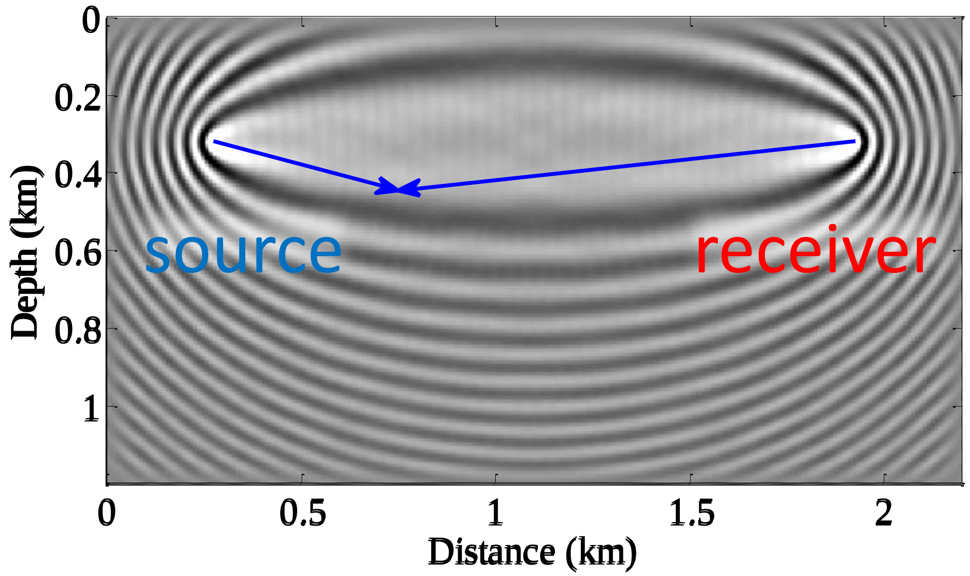
$$\delta G\left(\mathbf{r}_g,\mathbf{r}\,|\,s_n,\Delta s_n\right) \approx \omega^2 \int d\mathbf{r}' G\left(\mathbf{r}_g,\mathbf{r}'\,|\,s_n\right) G\left(\mathbf{r}',\mathbf{r}\,|\,s_n\right) \Delta s_n\left(\mathbf{r}'\right)$$

$$\left(\frac{\partial G\left(\mathbf{r}_g,\mathbf{r}_s\,|\,s_{n+1}\right)}{\partial s\left(\mathbf{r}\right)}\right)_1 = \omega^2 \left[\boxed{\delta G\left(\mathbf{r}_g,\mathbf{r}\,|\,s_n,\Delta s_n\right)} G\left(\mathbf{r},\mathbf{r}_s\,|\,s_n\right) + G\left(\mathbf{r}_g,\mathbf{r}\,|\,s_n\right) \boxed{\delta G\left(\mathbf{r},\mathbf{r}_s\,|\,s_n,\Delta s_n\right)}\right]$$

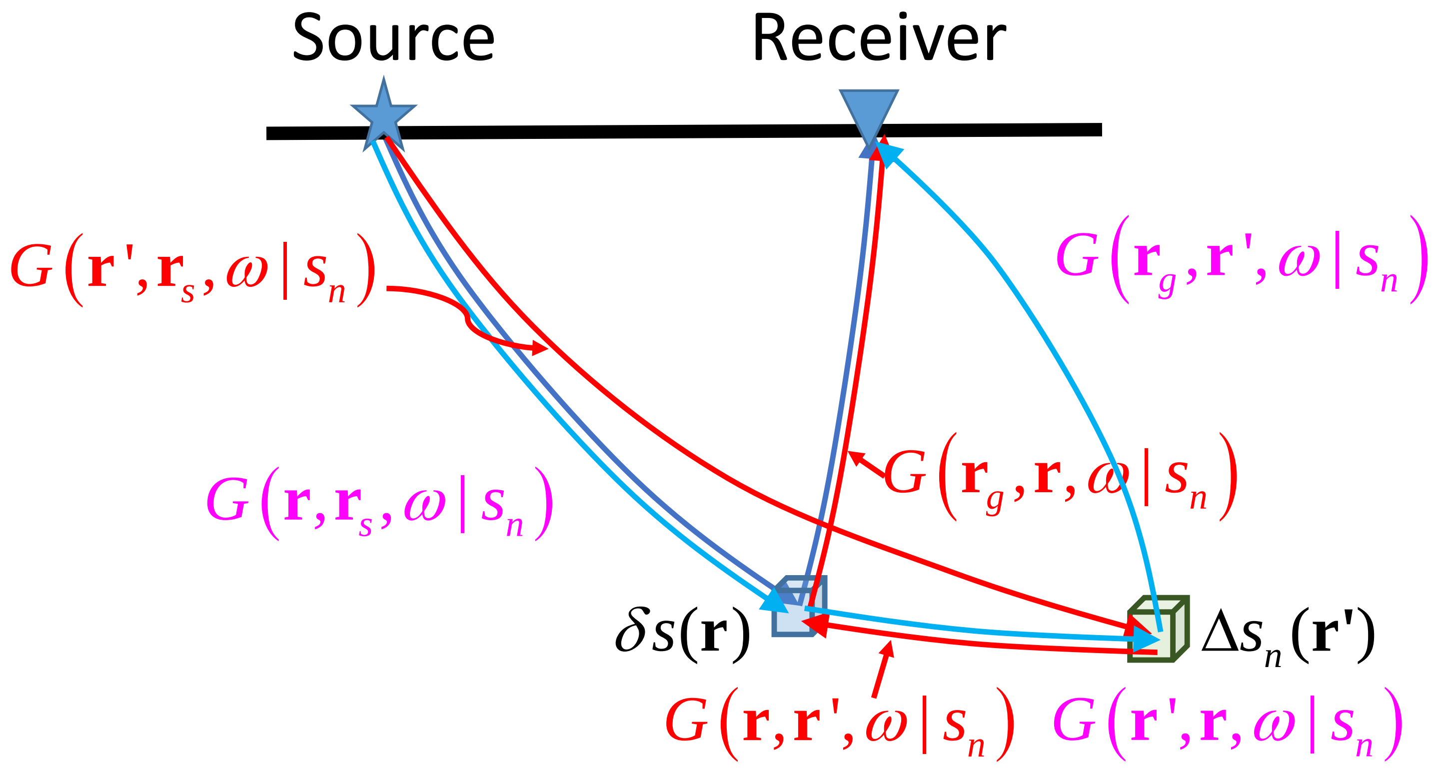
$$\delta G\left(\mathbf{r},\mathbf{r}_s\,|\,s_n,\Delta s_n\right) \approx \omega^2 \int d\mathbf{r}' G\left(\mathbf{r},\mathbf{r}'\,|\,s_n\right) G\left(\mathbf{r}',\mathbf{r}_s\,|\,s_n\right) \Delta s_n\left(\mathbf{r}'\right)$$

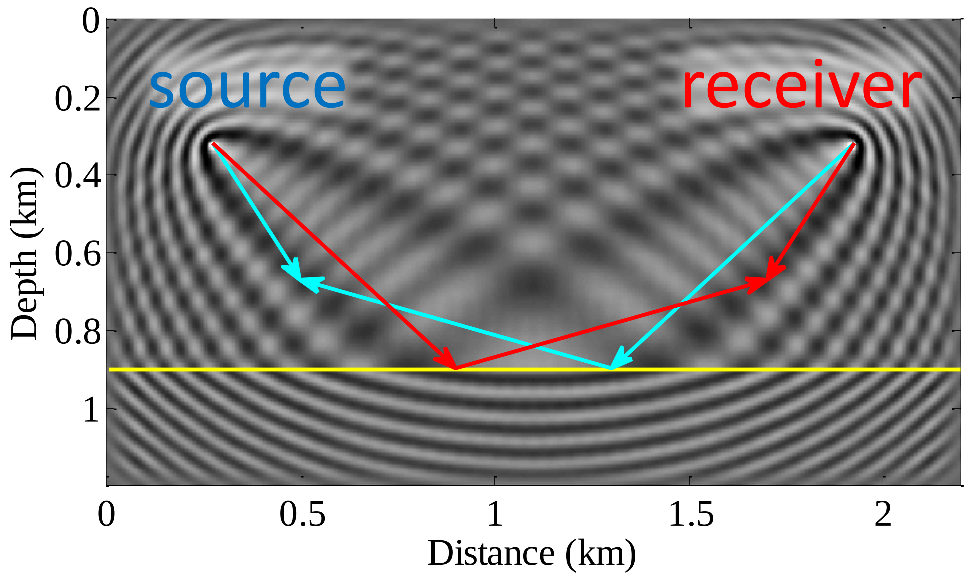
Sensitivities: from linear to nonlinear



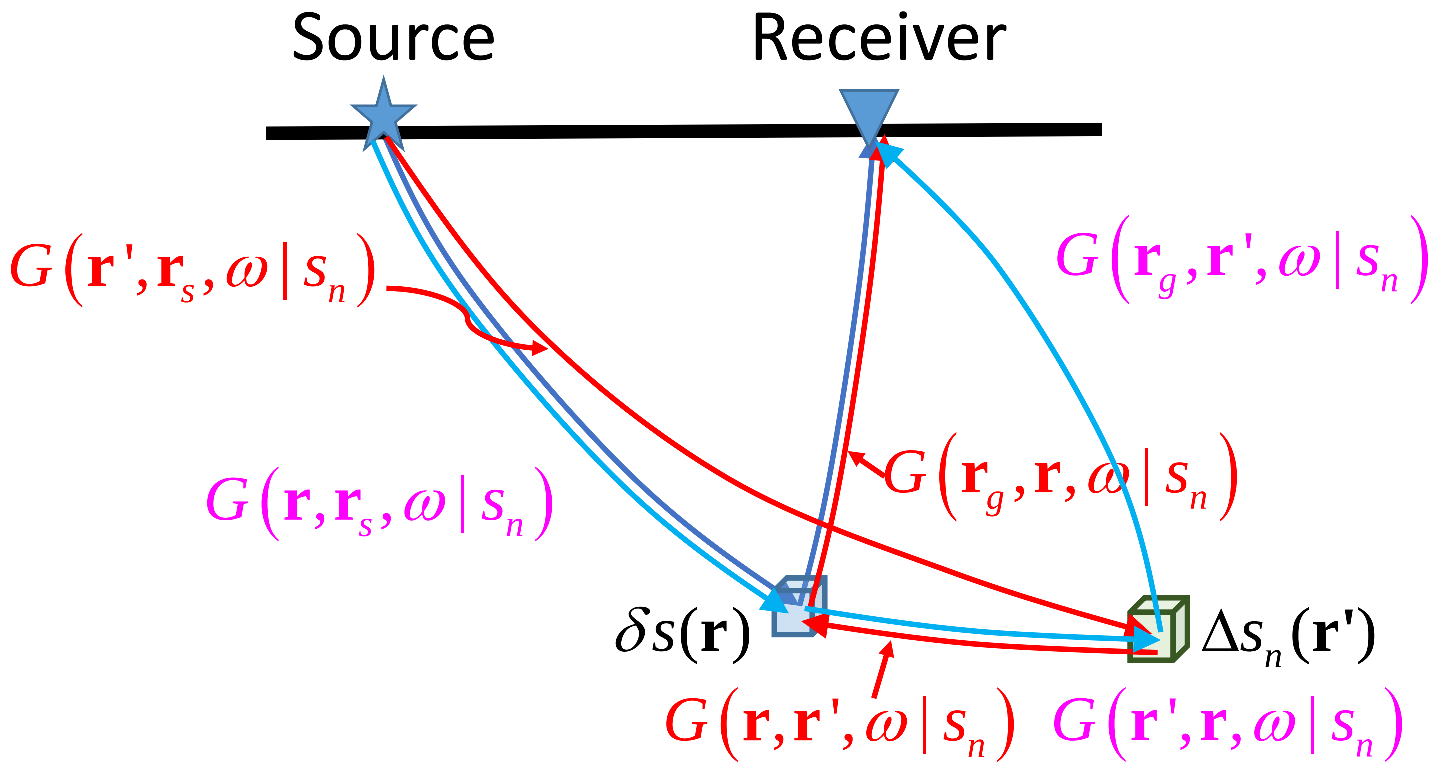


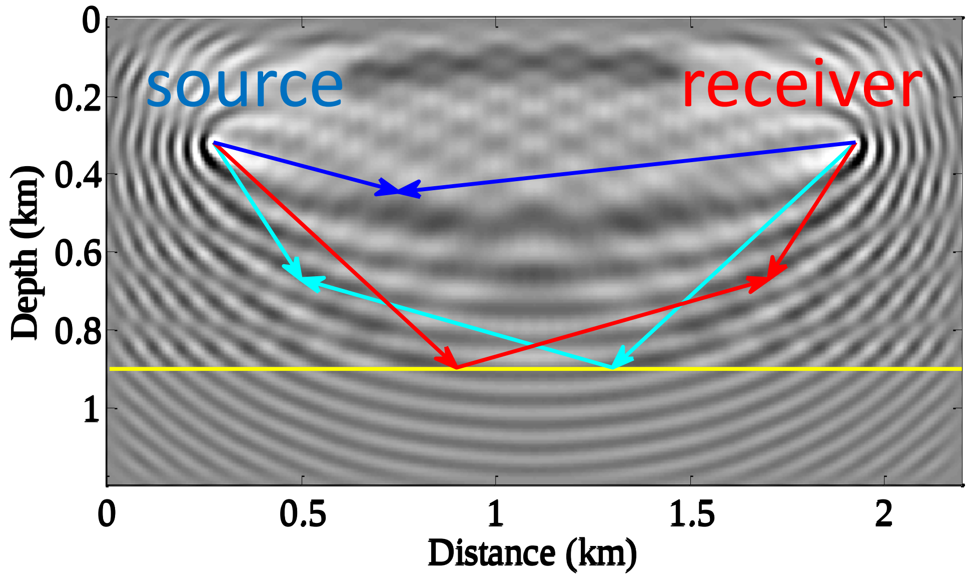
Sensitivities: from linear to nonlinear





Sensitivities: from linear to nonlinear





$$\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s,\omega|s_{n+1})}{\partial s(\mathbf{r})} = \left(\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s,\omega|s_{n+1})}{\partial s(\mathbf{r})}\right)_0 + \left(\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s,\omega|s_{n+1})}{\partial s(\mathbf{r})}\right)_1 + \dots$$

Determining model perturbation before n+1th iteration

- Determining perturbation before n+1th FWI iteration

$$\underbrace{\delta P\left(\mathbf{r}_g,\mathbf{r}_s,\omega \mid s_n\right)=\omega^2\int d\mathbf{r}' G\left(\mathbf{r}_g,\mathbf{r}',\omega \mid s_n\right)\Delta s\left(\mathbf{r}'\right)G\left(\mathbf{r}',\mathbf{r}_s,\omega \mid s_n\right)}^{\text{Linear relationship between } \delta P \text{ and } \Delta s}+\cancel{\ldots}$$

Difference between true model and the current updated model S_n

- Direct nonlinear inverse scattering
- Approach by the solution of a linear inverse problem

Determining model perturbation before n+1th iteration

- Misfit function

data residual

calculated data residual

$$\phi(\delta\tilde{s}) = \frac{1}{2}\sum_{\mathbf{r}_s}\sum_{\mathbf{r}_g}\sum_{\omega}\left\| \boxed{\delta P\left(\mathbf{r}_g,\mathbf{r}_s,\omega \mid s_n\right)} - \boxed{\delta P_{cal}\left(\mathbf{r}_g,\mathbf{r}_s,\omega \mid s_n\right)} \right\|^2$$

- Update $\delta\tilde{s}$ iteratively

$$\tilde{g}_m(\mathbf{r}) = -\sum_{\mathbf{r}_s}\sum_{\mathbf{r}_g}\sum_{\omega}\mathrm{Re}\left(\omega^2 G\left(\mathbf{r}_g,\mathbf{r},\omega \mid s_n\right) G\left(\mathbf{r},\mathbf{r}_s,\omega \mid s_n\right)\delta S_m^*\left(\mathbf{r}_g,\mathbf{r}_s,\omega \mid s_n\right)\right)$$

$$\delta\tilde{s}_{m+1}(\mathbf{r}) = \delta\tilde{s}_m(\mathbf{r}) - \alpha_m \tilde{g}_m(\mathbf{r})$$

Nonlinear FWI

- Direct update of the model (DWI)

$$\tilde{s}_{n+1}(\mathbf{r}) = s_n(\mathbf{r}) + \delta\tilde{s}(\mathbf{r})$$

- Nonlinear FWI with gradient

$$g_n(\mathbf{r}) = -\sum_{\mathbf{r}_s}\sum_{\mathbf{r}_g}\sum_{\omega} \mathrm{Re}\left(\boxed{\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega \mid \tilde{s}_{n+1})}{\partial s(\mathbf{r})}} \delta P^*(\mathbf{r}_g, \mathbf{r}_s, \omega \mid s_n) \right)$$

- FOFWI:

  Data residual in s_n

$$= \omega^2 \left[G(\mathbf{r}, \mathbf{r}_s, \omega \mid s_n) G(\mathbf{r}_g, \mathbf{r}, \omega \mid s_n) + \delta G(\mathbf{r}_g, \mathbf{r}, \omega \mid s_n, \delta\tilde{s}) G(\mathbf{r}, \mathbf{r}_s, \omega \mid s_n) + G(\mathbf{r}_g, \mathbf{r}, \omega \mid s_n) \delta G(\mathbf{r}, \mathbf{r}_s, \omega \mid s_n, \delta\tilde{s}) \right]$$

- NFWI:

  Data residual in $\tilde{s}_{n+1}$

$$= \omega^2 \left[G(\mathbf{r}, \mathbf{r}_s, \omega \mid \tilde{s}_{n+1}) G(\mathbf{r}_g, \mathbf{r}, \omega \mid \tilde{s}_{n+1}) - \delta G(\mathbf{r}_g, \mathbf{r}, \omega \mid s_n, \delta\tilde{s}) \delta G(\mathbf{r}, \mathbf{r}_s, \omega \mid s_n, \delta\tilde{s}) \right]$$

Two-iteration nonlinear FWI

Algorithm 2 Algorithm for frequency domain nonlinear FWI

Input: Recorded seismic data $P(\mathbf{r}_g, \mathbf{r}_s, \omega)$, initial model s_0
Output: Inverted model s_n

```
for ω = ω_min, ..., ω_max do                              ▷ Frequency loop
    for n = 0, ..., n_max do                               ▷ Outer loop for FWI
        Get data residual δP(r_g, r_s, ω|s_n)

        Calculate nonlinear gradient (equation (33) or (39))
        Calculate the step length μ_n using the line search method
        Update to s_{n+1} (equation (8))
    end for
end for
```

No inner iteration=> FWI

Two-iteration nonlinear FWI

Algorithm 2 Algorithm for frequency domain nonlinear FWI

Input: Recorded seismic data $P(\mathbf{r}_g, \mathbf{r}_s, \omega)$, initial model s_0
Output: Inverted model s_n

for $\omega = \omega_{min}, \ldots, \omega_{max}$ **do** ▷ *Frequency loop*
    for $n = 0, \ldots, n_{max}$ **do** ▷ *Outer loop for FWI*
        Get data residual $\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)$
        Initial perturbation $\delta \tilde{s}_0 \leftarrow 0$
        for $m = 0, \ldots, m_{max}$ **do** ▷ *Inner loop for perturbation*
            Calculate $\delta P_{cal}(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)$
            Get data residual $\delta S(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)$
            Calculate the gradient $\tilde{g}_m(\mathbf{r})$ (equation (24))
            Calculate the step length α_m using the line search method
            Update $\tilde{s}_m$ (equation (23))
        end for
        Calculate nonlinear gradient (equation (33) or (39))
        Calculate the step length μ_n using the line search method
        Update to s_{n+1} (equation (8))
    end for
end for

Two-iteration nonlinear FWI

Algorithm 2 Algorithm for frequency domain nonlinear FWI

Input: Recorded seismic data $P(\mathbf{r}_g, \mathbf{r}_s, \omega)$, initial model s_0
Output: Inverted model s_n

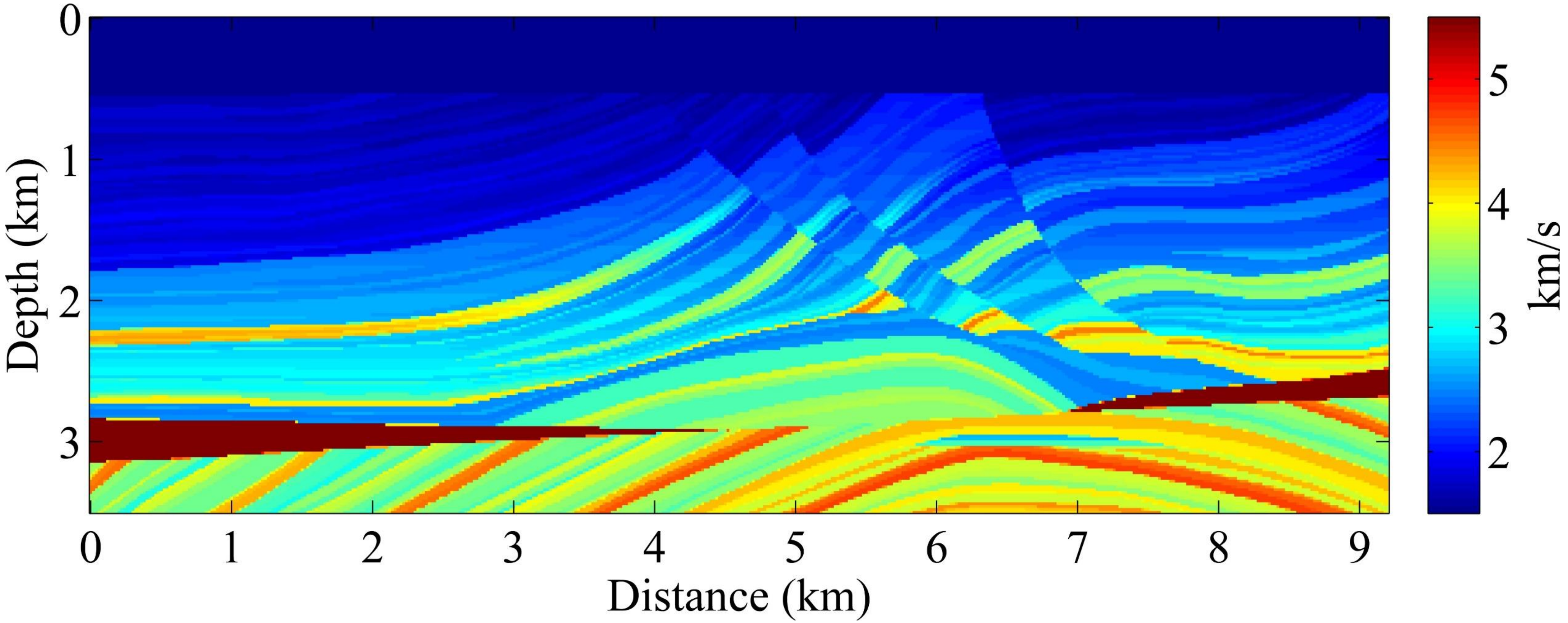
```
for ω = ω_min, ..., ω_max do                                   ▷ Frequency loop
    for n = 0, ..., n_max do                                   ▷ Outer loop for FWI
        Get data residual δP(r_g, r_s, ω|s_n)
        Initial perturbation δs̃_0 ← 0
        for m = 0, ..., m_max do                               ▷ Inner loop for perturbation
            Calculate δP_cal(r_g, r_s, ω|s_n)
            Get data residual δS(r_g, r_s, ω|s_n)
            Calculate the gradient g̃_m(r) (equation (24))
            Calculate the step length α_m using the line search method
            Update s̃_m (equation (23))
        end for
                Direct update

        Update to s_{n+1} (equation (8))
    end for
end for
```

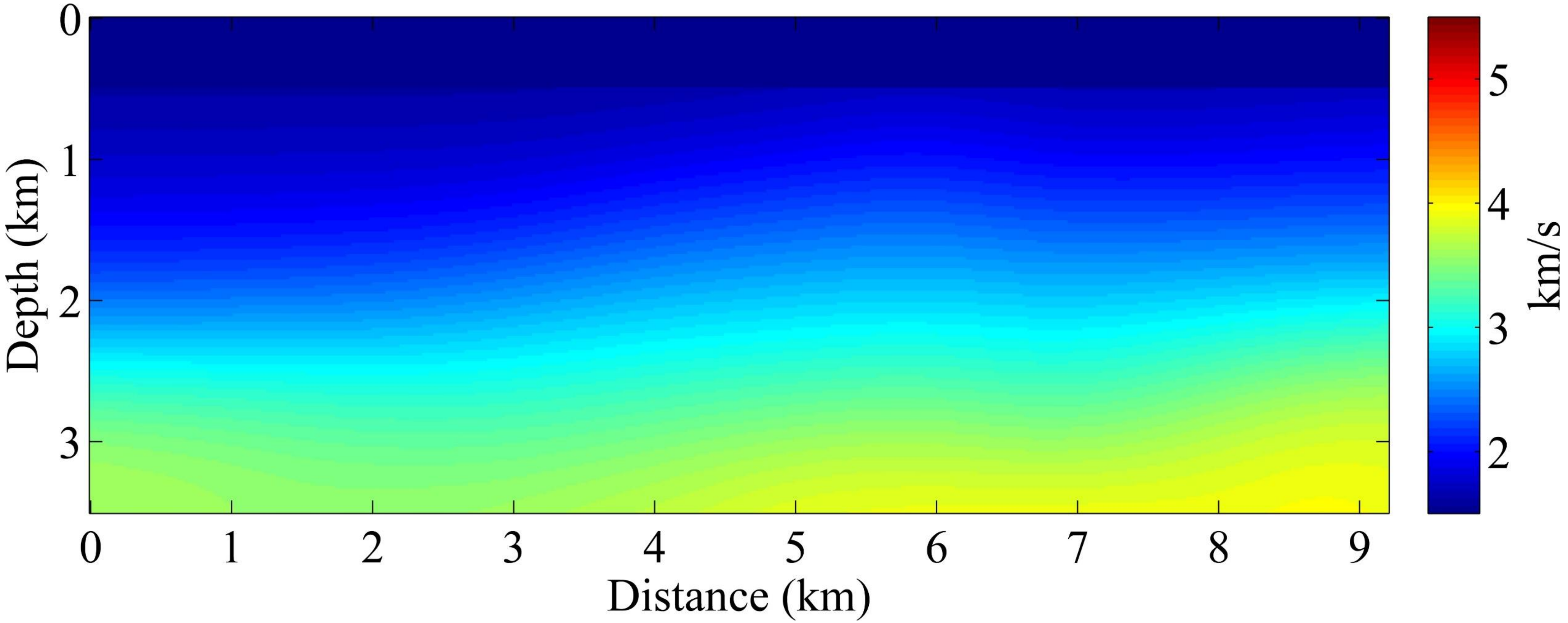
Numerical examples

- 2D Marmousi model with added water layer
  - Model size *3.5kmX9.2Km* with *20m* interval
  - 46 sources with *200m* spacing, starting from *100m*
  - 461 receivers with *20m* spacing, starting from *0m*
- Test with different initial model
  - True model smoothed with Gaussian window
  - Vertically linearly increasing model (*1500-4000m/s*)
- Test with different frequency range
  - Starting from *2Hz* to *15Hz* (*2Hz, 3.3Hz, 5.5Hz, 9Hz, 14.9Hz*)
  - Starting from *4Hz* to *15Hz*
    - *4Hz, 6.6Hz, 14.9Hz*
    - 12 frequencies with *1Hz* interval

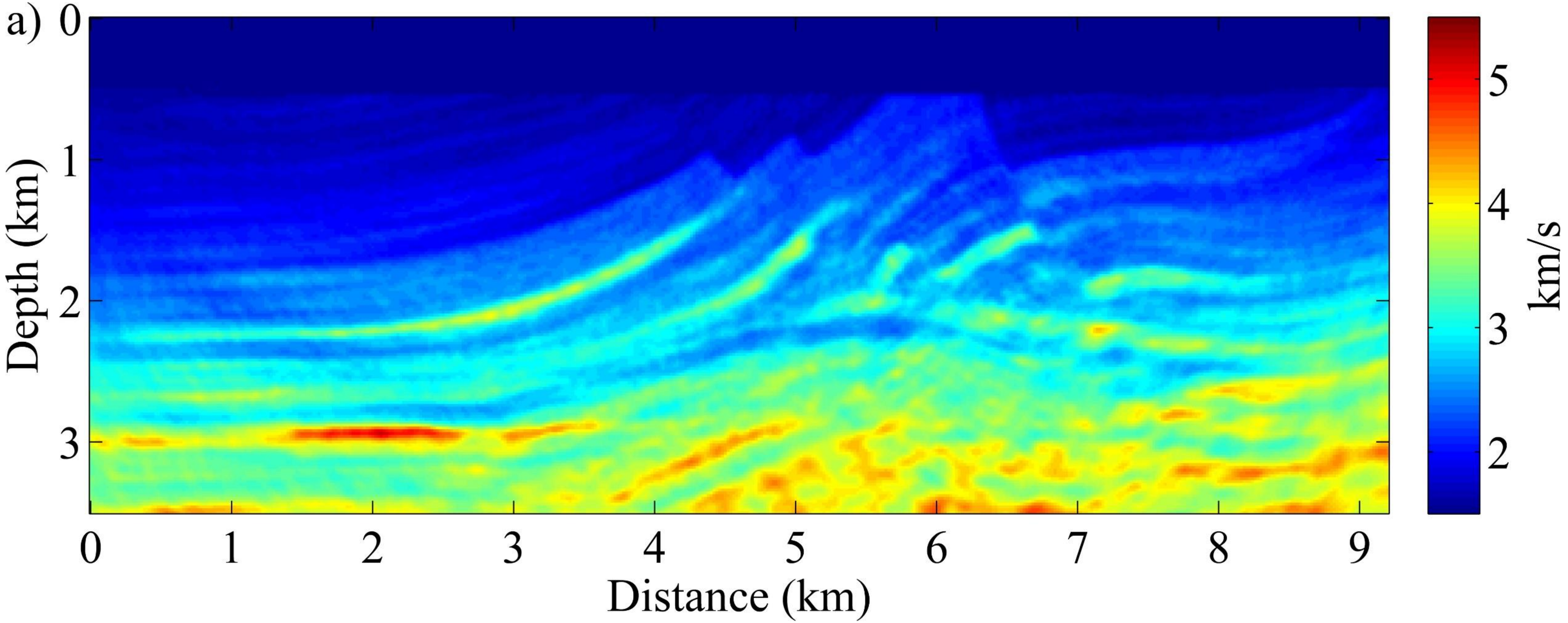
True Marmousi model



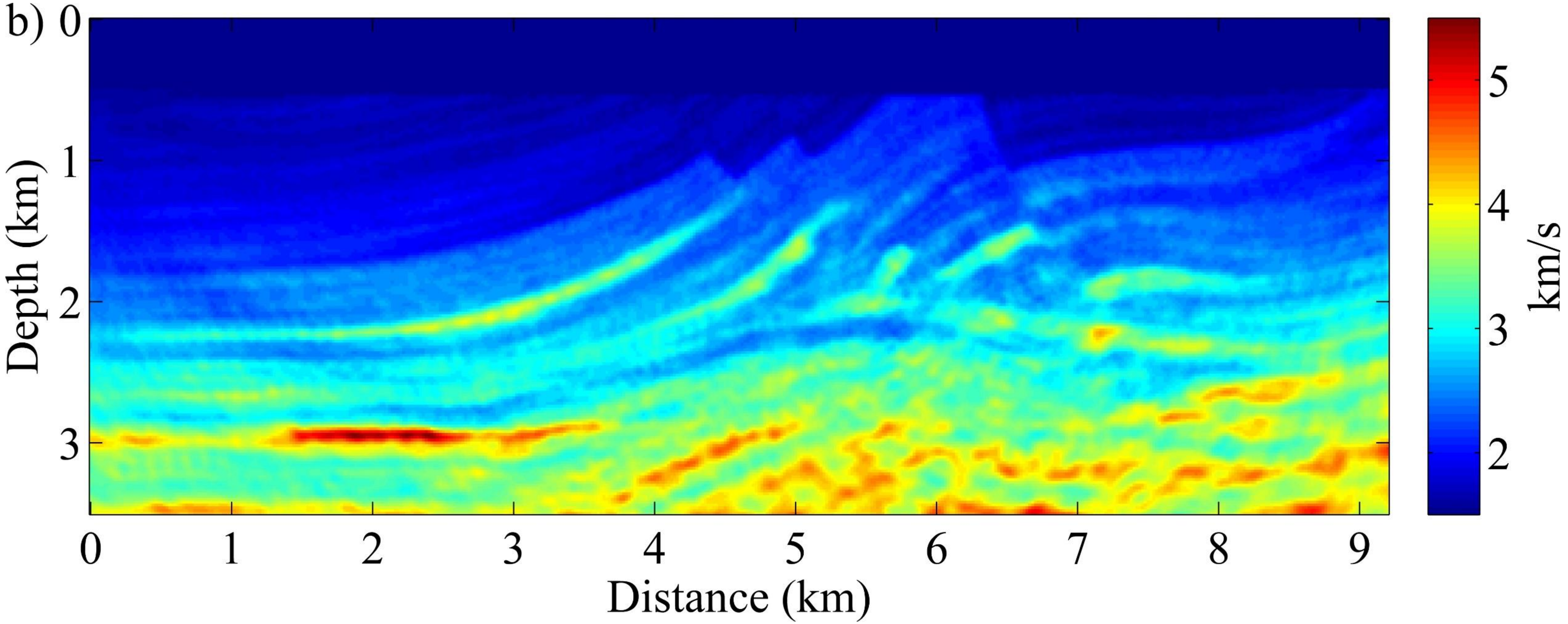
Initial model obtained after smooth the true model



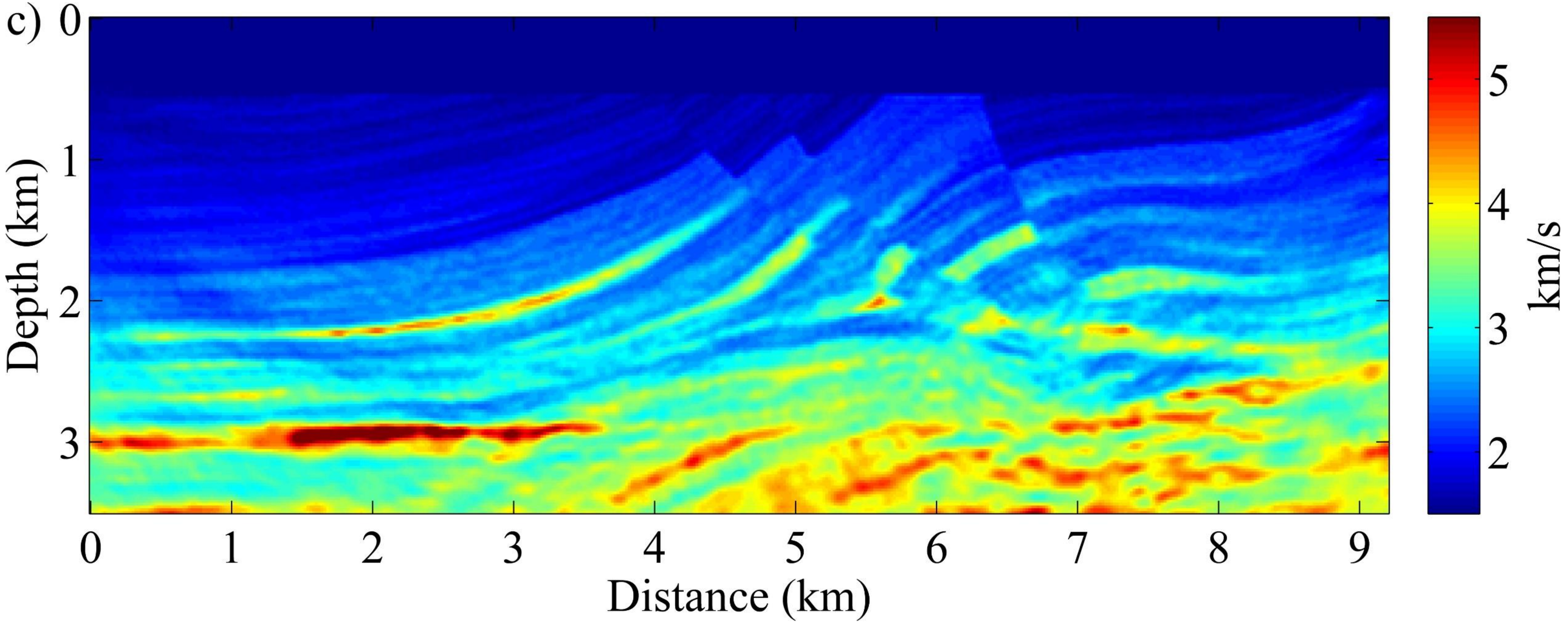
FWI result with 10 iterations for 3 frequencies (4-15Hz)



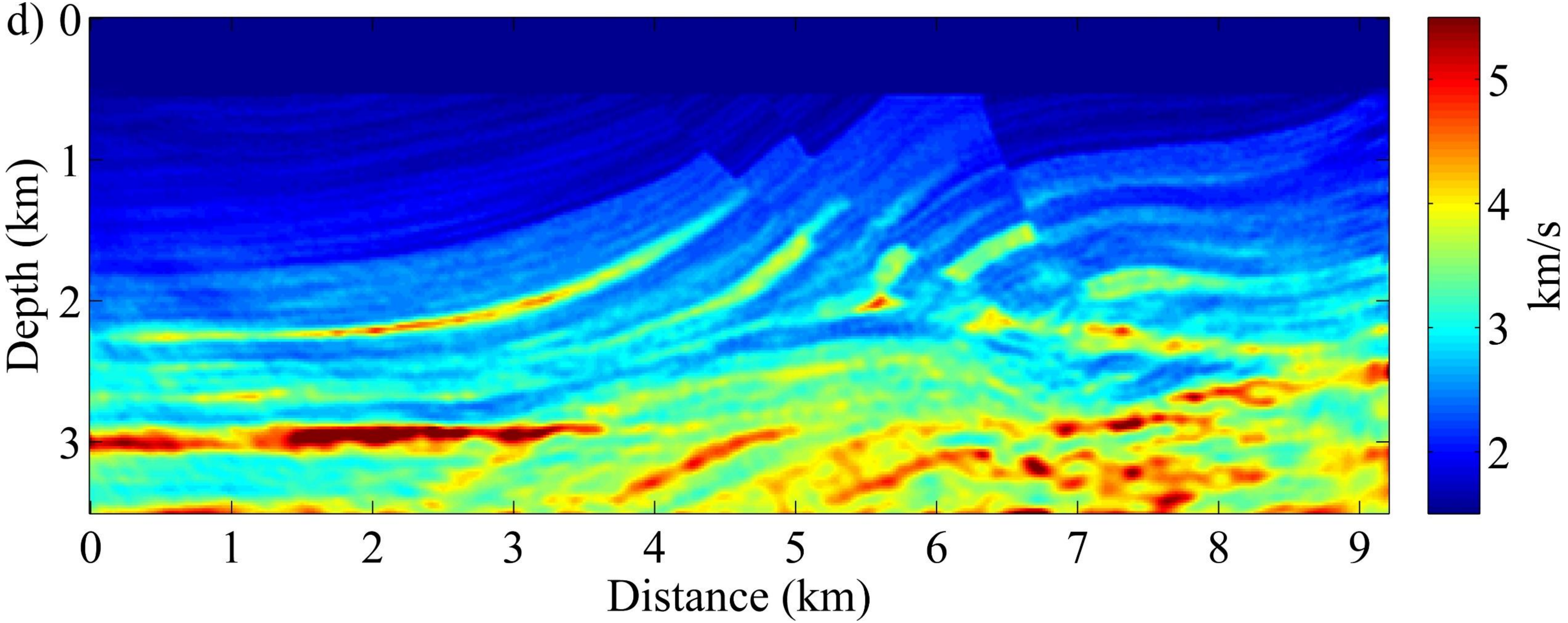
FOFWI result with 10 iterations for 3 frequencies (4-15Hz)



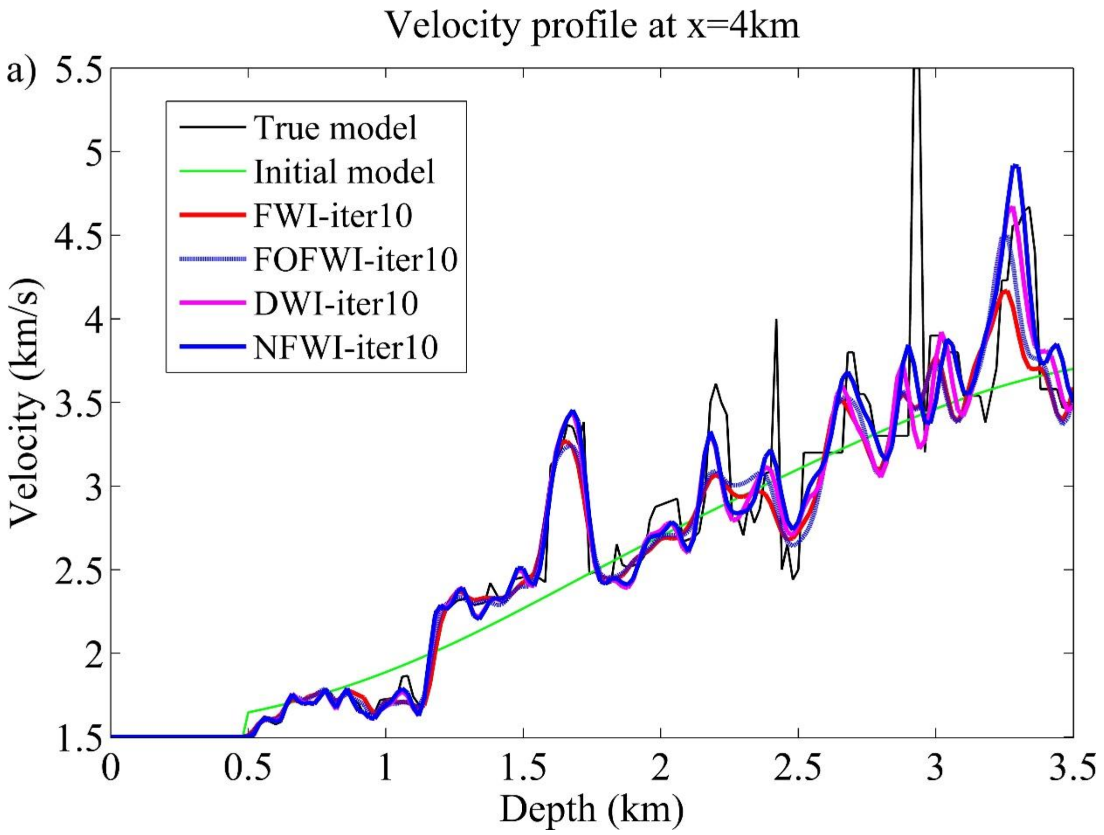
DWI result with 10 iterations for 3 frequencies (4-15Hz)

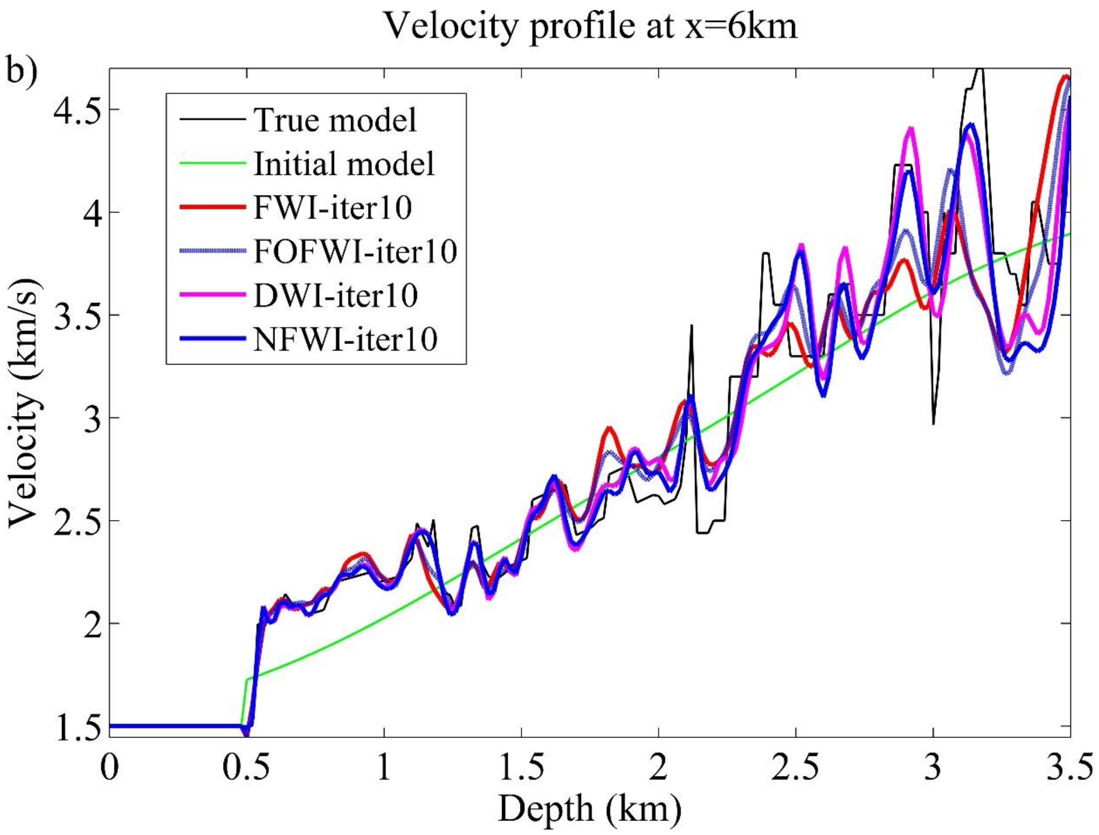


NFWI result with 10 iterations for 3 frequencies (4-15Hz)

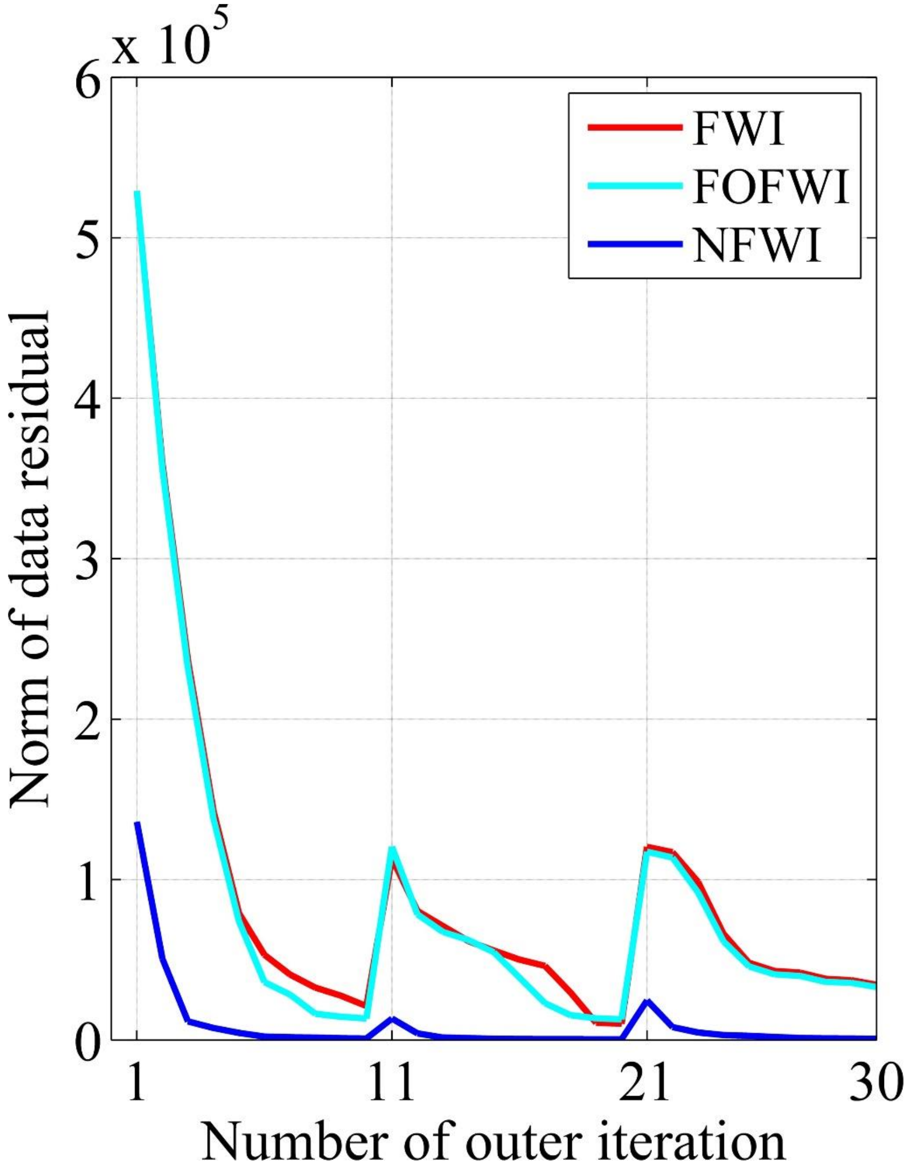


Velocity profiles

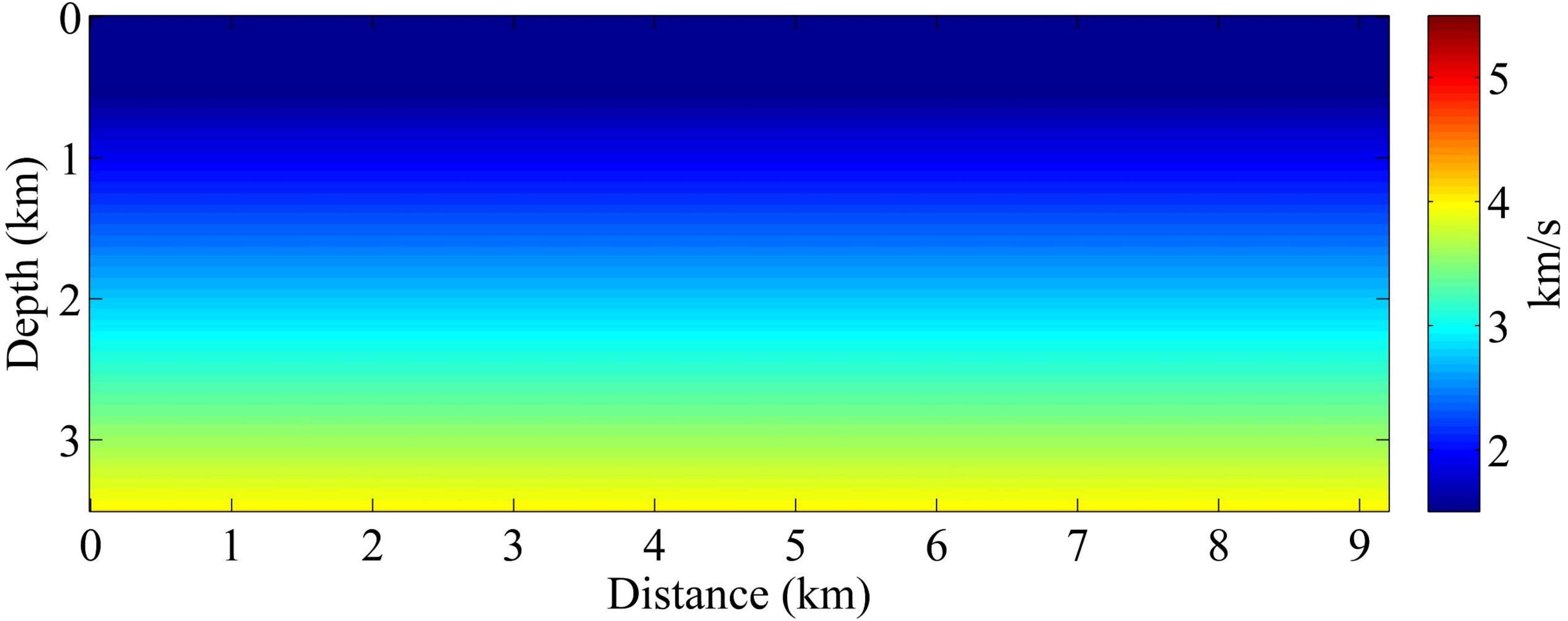




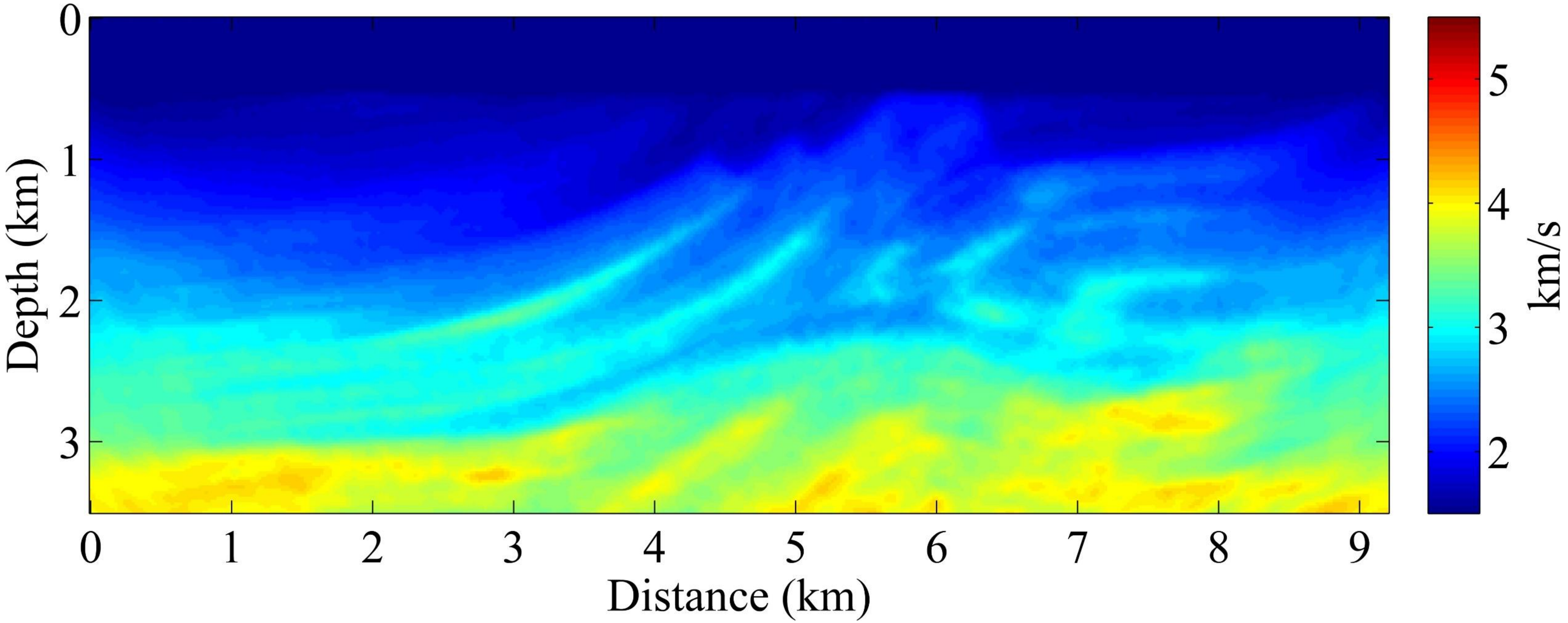
Data residual vs iteration number



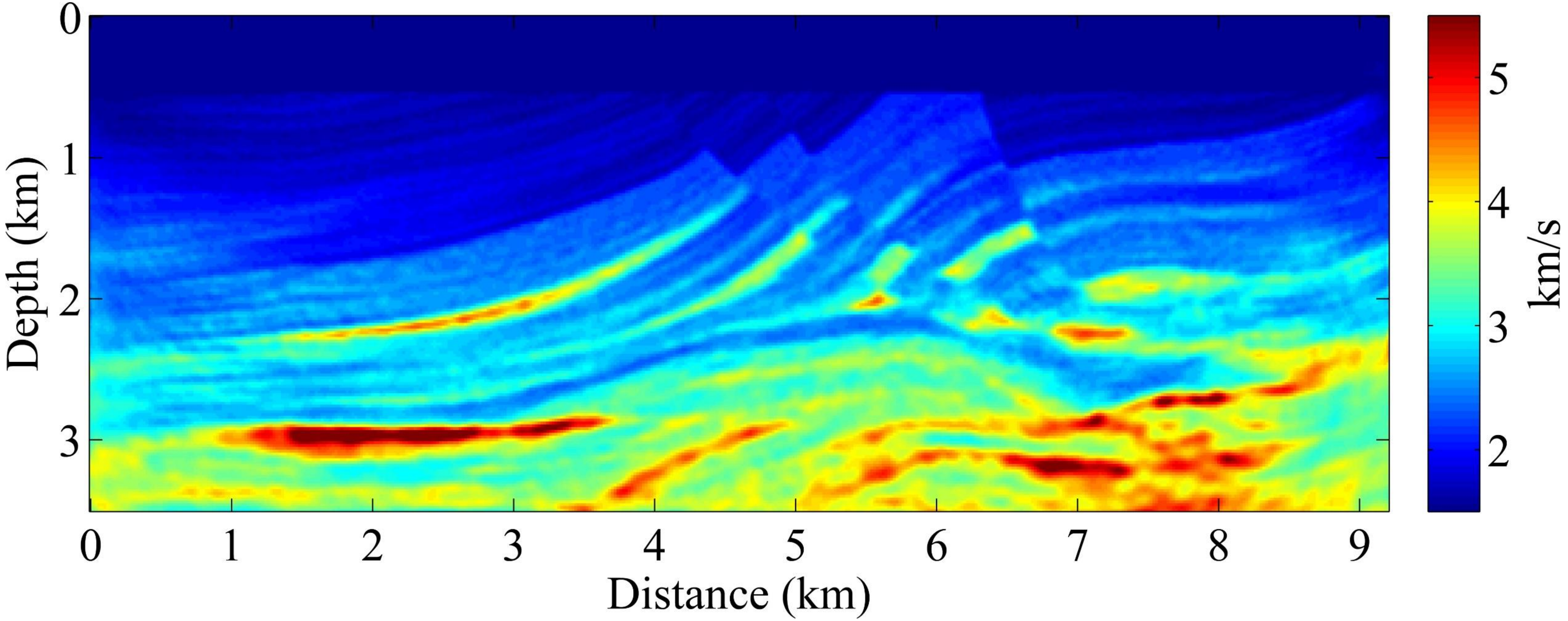
Initial model (linearly changed from 1.5km/s to 4km/s along z)



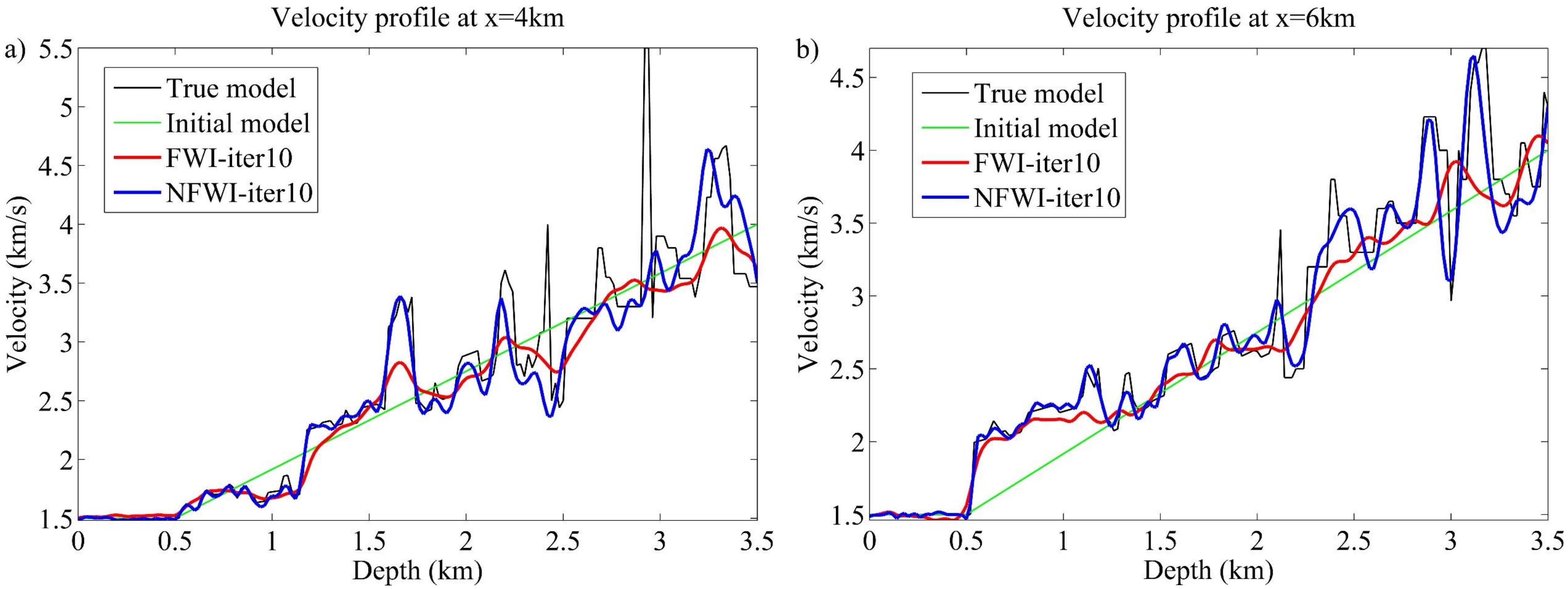
FWI result with 10 iterations for 5 frequencies (2-15Hz)



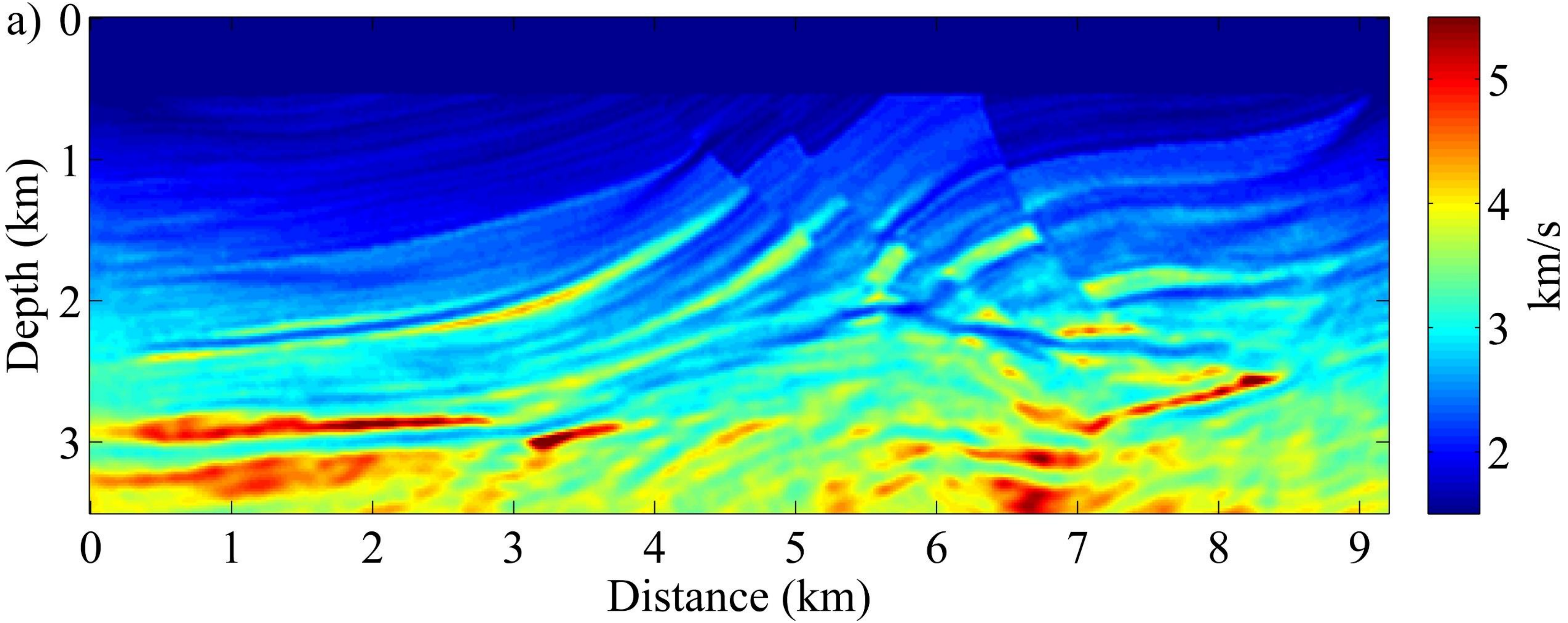
NFWI result with 10 iterations for 5 frequencies (2-15Hz)



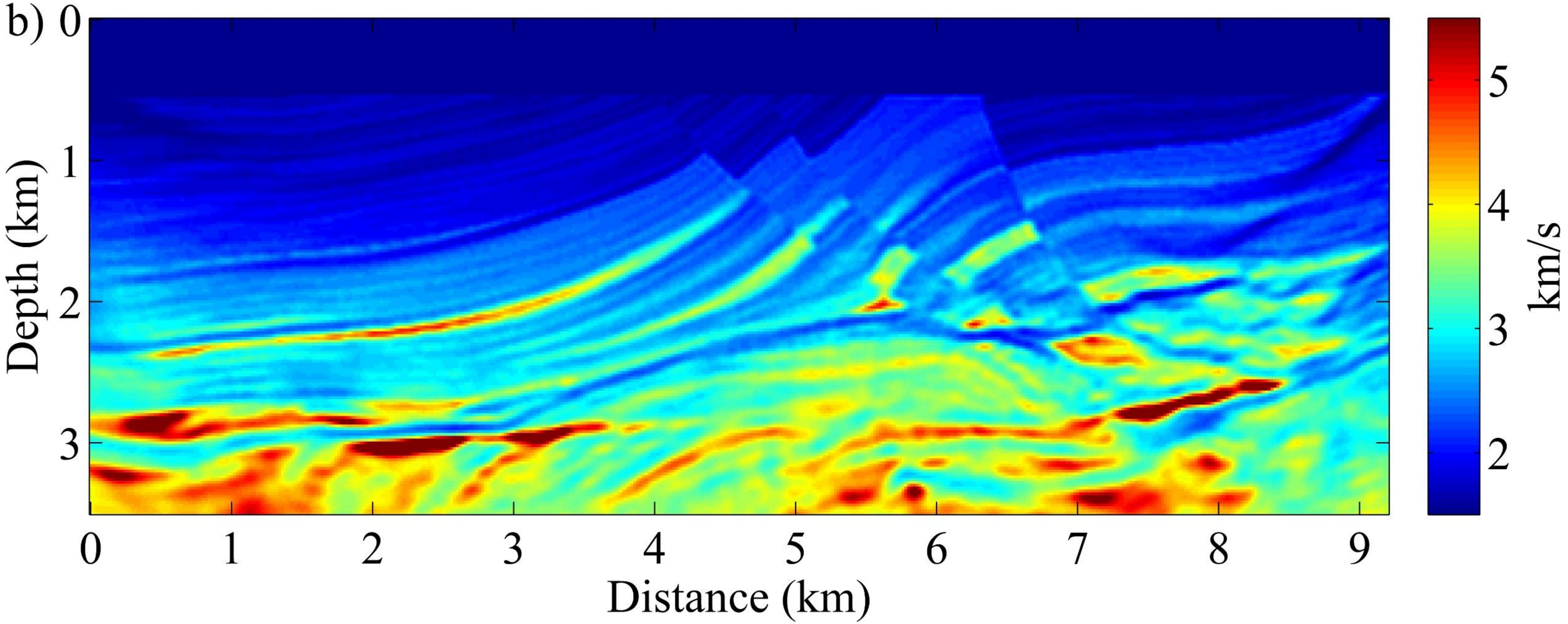
Velocity profiles



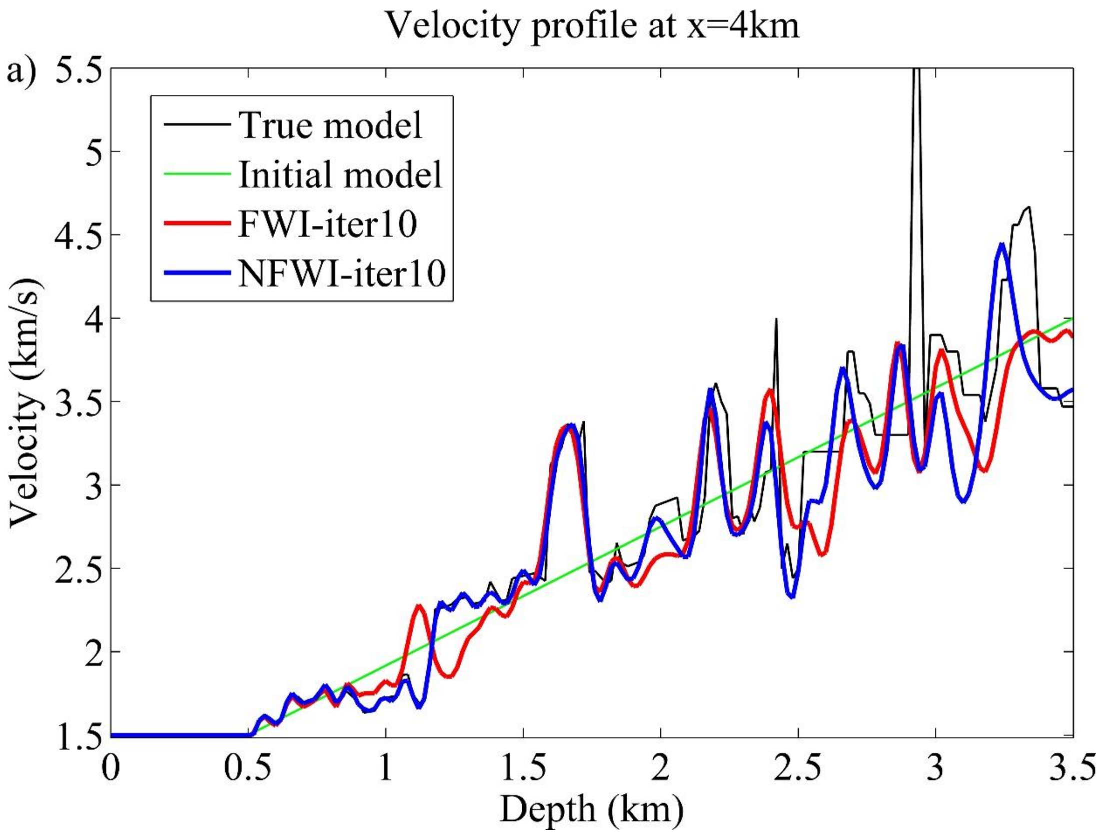
FWI result with 10 iterations for 12 frequencies (4-15Hz)

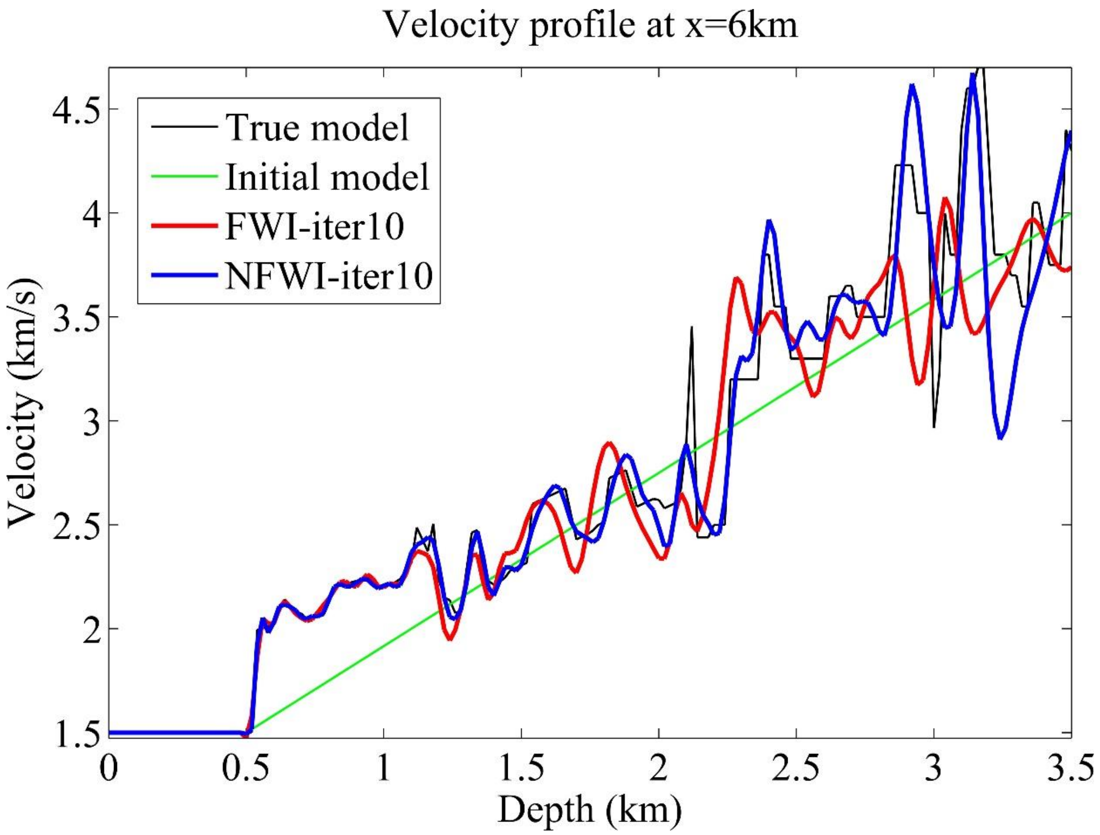


NFWI result with 10 iterations for 12 frequencies (4-15Hz)



Velocity profiles





Conclusions

- Nonlinear sensitivities provide the possibility to better handle the nonlinearity in FWI
  - Update the long wavelength components in the deeper regions as well as the shallow regions
  - Help to converge faster
- A two-iteration nonlinear FWI approach
  - Using data residual from the current iteration to perform a linear inversion to update perturbation
  - Using perturbation to construct gradient

Acknowledgements

- All CREWES Sponsors
- NSERC
- Wenyong Pan, Junxiao Li, Jian Sun, Huaizhen Chen, Khalid Almuteri
- All the other researchers in CREWES

Thank You