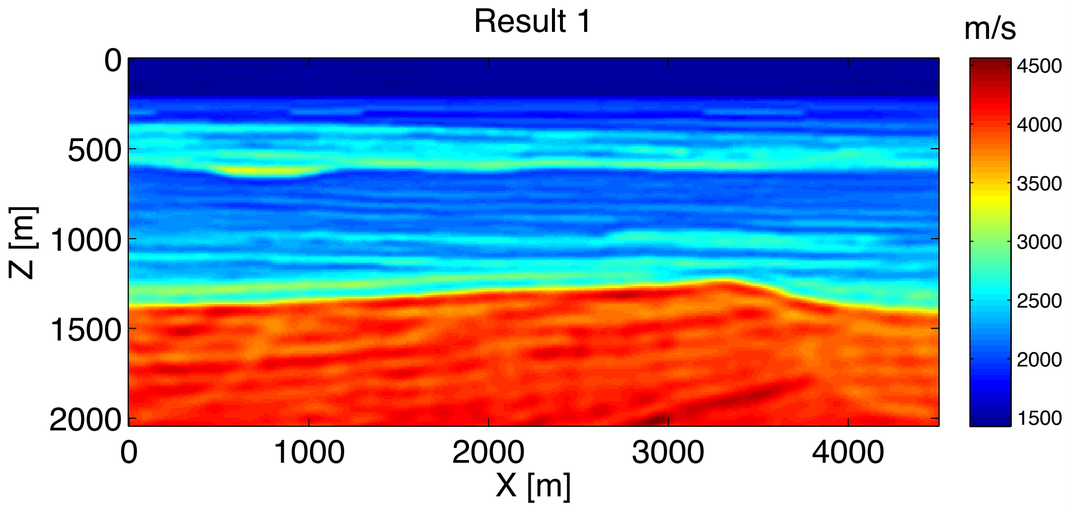
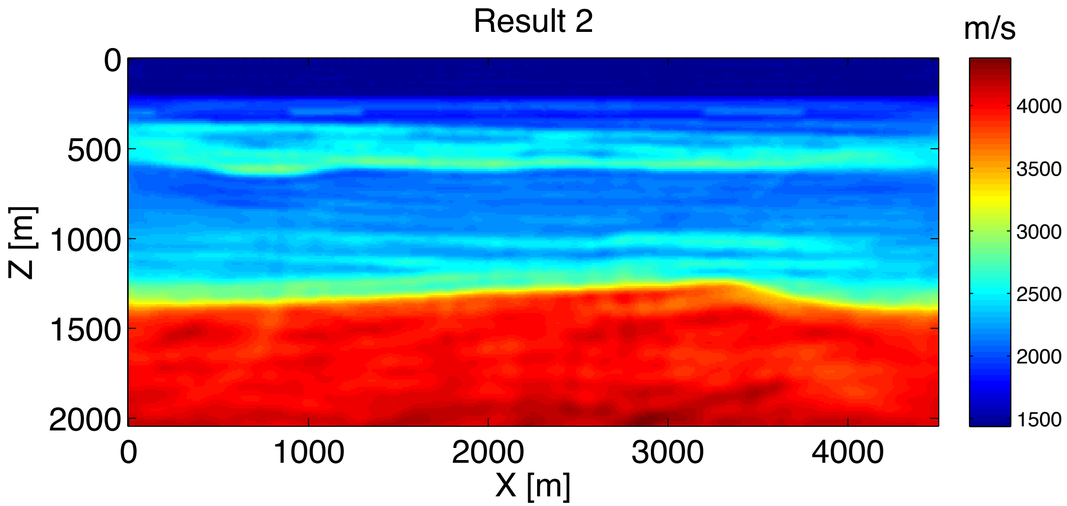
A stochastic quasi-Newton McMC method for uncertainty quantification of full-waveform inversion

Zhilong Fang, Felix J. Herrmann and Chia Ying Lee

SLIM
University of British Columbia

Motivation





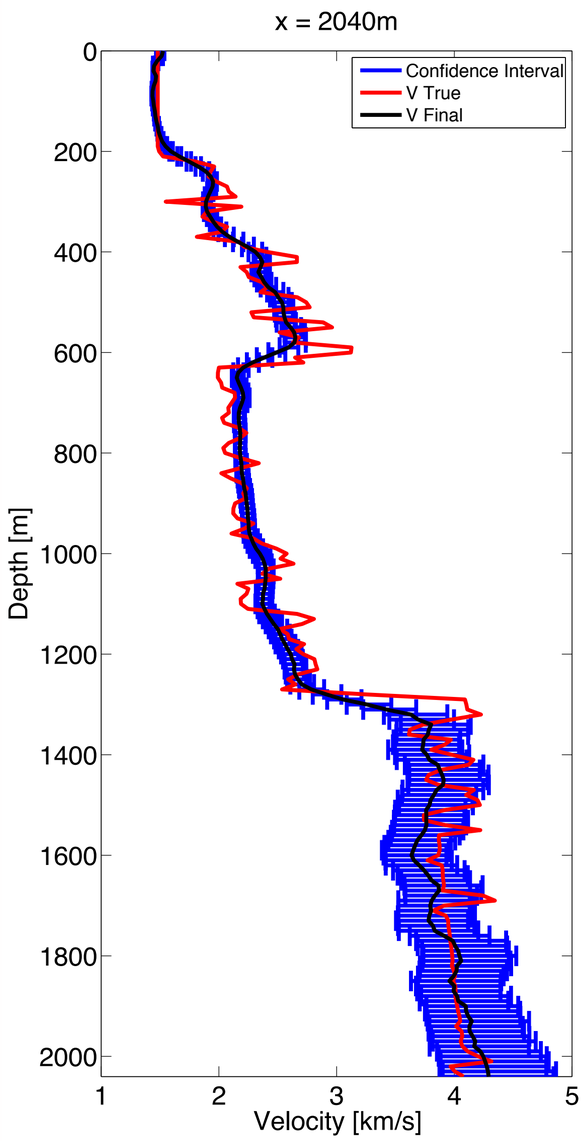
Which one is better?

Misfit?
Eyeball norm?

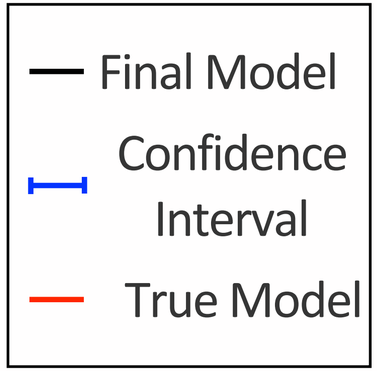
Uncertainty?
Standard Deviation?

Confidence interval

Fast Accurate!!



x = 2040m



Bayesian theory

Deterministic inverse problem:

$$\mathbf{m}^* = \arg\min\left(\frac{1}{2N_s}\sum_{1}^{N_s}\|f_i(\mathbf{m}) - \mathbf{d}_{i\mathrm{obs}}\|_{W_i}^2 + \frac{1}{2}\|\mathbf{m} - \overline{\mathbf{m}}\|_R^2\right)$$

Statistical inverse problem with Bayesian theory:

$$\pi_{post}(\mathbf{m}) := \pi(\mathbf{m}|\mathbf{d}_{obs}) \propto \pi_{prior}(\mathbf{m})\pi(\mathbf{d}_{obs}|\mathbf{m})$$

where $\mathbf{m}$ is the model parameter, and $\mathbf{d}_{obs}$ is the observed data

(James Martin *et al*, 2012)

Bayesian theory

Assume:

$$\text{noise} \sim \mathcal{N}(0, \Gamma_{noise})$$

$$\text{prior model distribution} \sim \mathcal{N}(\mathbf{m}_{prior}, \Gamma_{prior}).$$

Negative log-posterior of the posterior pdf:

$$V(\mathbf{m}) := -\log \pi_{post}(\mathbf{m}) := \frac{1}{2N_s} \sum_{1}^{N_s} \|f_i(\mathbf{m}) - \mathbf{d}_{i\text{obs}}\|^2_{\Gamma^{-1}_{i\text{noise}}} + \frac{1}{2}\|\mathbf{m} - \mathbf{m}_{prior}\|^2_{\Gamma^{-1}_{prior}}$$

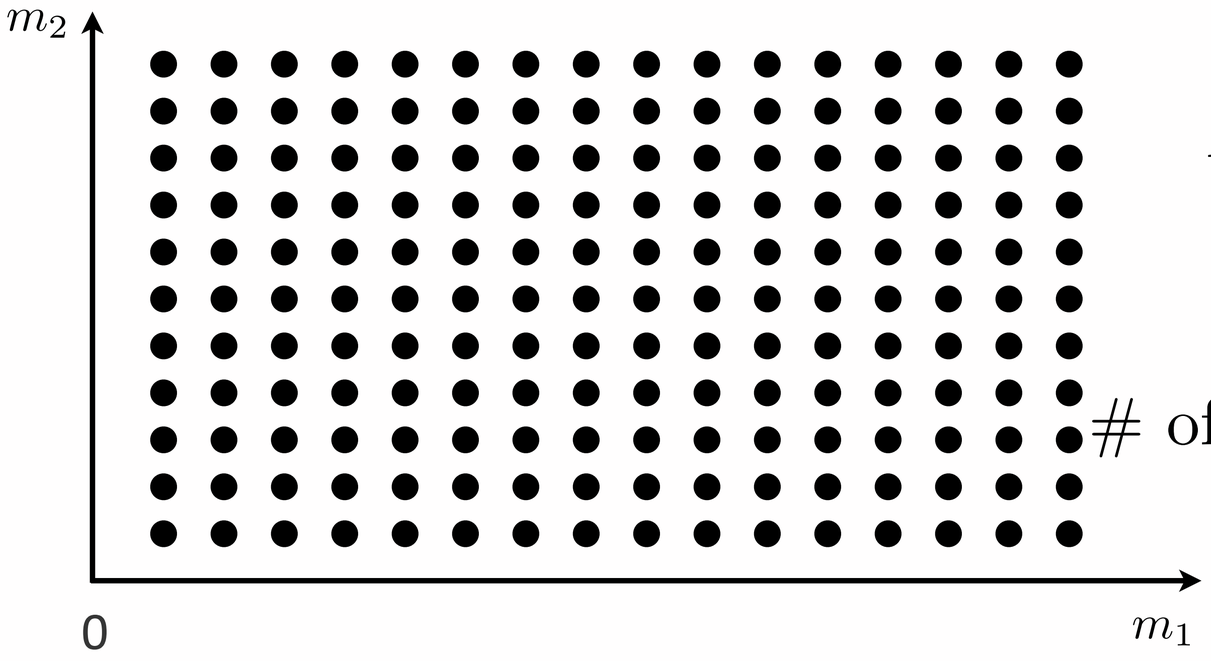
$$\mathbf{m}_{MAP} = \mathbf{m}^*$$

$$\|\mathbf{x}\|^2_{\Gamma^{-1}_{noise}} := \mathbf{x}^T \Gamma^{-1}_{noise} \mathbf{x}$$

Sampling the posterior pdf

How to obtain the posterior probability density function?

- Compute $\pi_{post}(\mathbf{m})$ by discretization? $N = N_{sub\_sample} * N_{para}$



$$N_{para} >> 10^6$$
$$N_{sub\_sample} >> 10^2$$
$$N >> 10^8$$
$$N_s >> 10^2$$

$$\#\text{ of PDE solvers } >> N * N_s$$

Approximate the pdf

How to obtain the posterior probability density function?

- Markov chain Monte Carlo method ? Metropolis - Hasting method

At sample $\mathbf{m}_k$
Draw sample $\mathbf{y}$ from the proposal distribution $\tilde{\pi}_k(\mathbf{m})$

if $\min(1, \dfrac{\pi_{post}(\mathbf{y})\tilde{\pi}_y(\mathbf{m}_k)}{\pi_{post}(\mathbf{m}_k)\tilde{\pi}_k(\mathbf{y})}) > \alpha$

    set $\mathbf{m}_{k+1} = \mathbf{y}$

else

    regenerate $\mathbf{y}$

end

Approximate the pdf

How to obtain the posterior probability density function?

- Stochastic Newton MCMC ?

$$V(\mathbf{m}) := -\log \pi_{post}(\mathbf{m}) := \frac{1}{2N_s}\sum^{N_s} \|f_i(\mathbf{m}) - \mathbf{d}_{i\text{obs}}\|^2_{\Gamma^{-1}_{i\text{noise}}} + \frac{1}{2}\|\mathbf{m} - \mathbf{m}_{prior}\|^2_{\Gamma^{-1}_{prior}}$$

$$\tilde{\pi}_k(\mathbf{m}) \backsim \mathcal{N}(\mathbf{m}_k - \mathbf{H}_k^{-1}\mathbf{g}_k, \mathbf{H}_k^{-1})$$

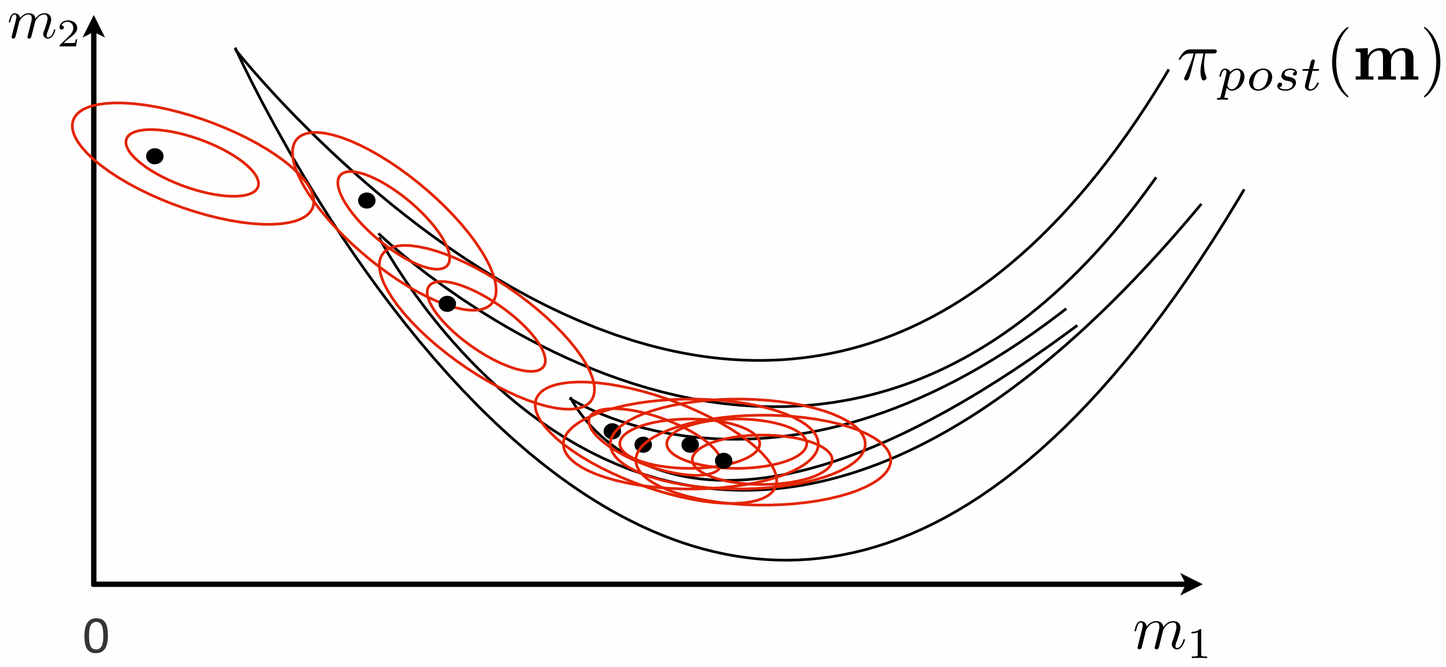
$$\underset{n\times n}{\begin{bmatrix} \mathbf{L}^T \end{bmatrix}} \underset{n\times n}{\begin{bmatrix} \mathbf{H}_{\text{misfit}} \end{bmatrix}} \underset{n\times n}{\begin{bmatrix} \mathbf{L} \end{bmatrix}} \approx \underset{n\times r}{\begin{bmatrix} \mathbf{V}_r \end{bmatrix}} \underset{r\times r}{\begin{bmatrix} \mathbf{D}_r \end{bmatrix}} \underset{r\times n}{\begin{bmatrix} \mathbf{V}_r^T \end{bmatrix}}$$

$$r \ll n$$

(James Martin *et al*, 2012)

Approximate the pdf

$$\tilde{\pi}_k(\mathbf{m}) \backsim \mathcal{N}(\mathbf{m}_k - \mathbf{H}_k^{-1}\mathbf{g}_k, \mathbf{H}_k^{-1})$$



(James Martin *et al*, 2012)

Approximate the pdf

Challenges:

- Low - rank approximation may not be correct.
- The computational cost of estimating Hessian is huge.
- The computational cost of calculating the posterior distribution probability density function is huge.

$$\text{Computational Cost} \sim \mathcal{O}(N_{\text{sample}} * N_s)$$

BFGS Hessian

At the stage *k*: approximated Hessian $\mathbf{B}_k$ and gradient $\mathbf{g}_k$,

Update $\mathbf{B}_k$ by adding two rank one matries:

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \mathbf{U}_k + \mathbf{V}_k$$

where $\mathbf{B}_{k+1}$ should satisfy

$$\mathbf{B}_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k) = \mathbf{g}_{k+1} - \mathbf{g}_k$$

$$\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$$
$$\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k,$$

No additional computational cost!!!

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{\mathbf{B}_k \mathbf{s}_k \mathbf{s}_k^T \mathbf{B}_k}{\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k}$$

$$\mathbf{B}_{k+1}^{-1} = \left(\mathbf{I} - \frac{\mathbf{s}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}\right) \mathbf{B}_k^{-1} \left(\mathbf{I} - \frac{\mathbf{y}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}\right) + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}$$

l-BFGS Hessian with pseudo GN Hessian as a starter

Limited memory BFGS Hessian:

$$\{\mathbf{s}_i\}_{k-M+1\leq i\leq k}$$
$$\{\mathbf{y}_i\}_{k-M+1\leq i\leq k}$$
$$\mathbf{B}_0^{(k)}$$

$$\longrightarrow$$

$$\mathbf{B}_k$$
$$\mathbf{B}_k^{-1}$$

Choice of $\mathbf{B}_0^{(k)}$: pseudo GN Hessian

(Choi, Y, *et al*, 2012)

Randomized source subsampling

To speed up the computation of posterior pdf:

$$V(\mathbf{m}) := -\log \pi_{post}(\mathbf{m}) := \frac{1}{2N_s}\sum_{1}^{N_s} \|f_i(\mathbf{m}) - \mathbf{d}_{i\mathrm{obs}}\|^2_{\Gamma^{-1}_{i\mathrm{noise}}} + \frac{1}{2}\|\mathbf{m} - \mathbf{m}_{prior}\|^2_{\Gamma^{-1}_{prior}}$$

Randomized source subsampling:

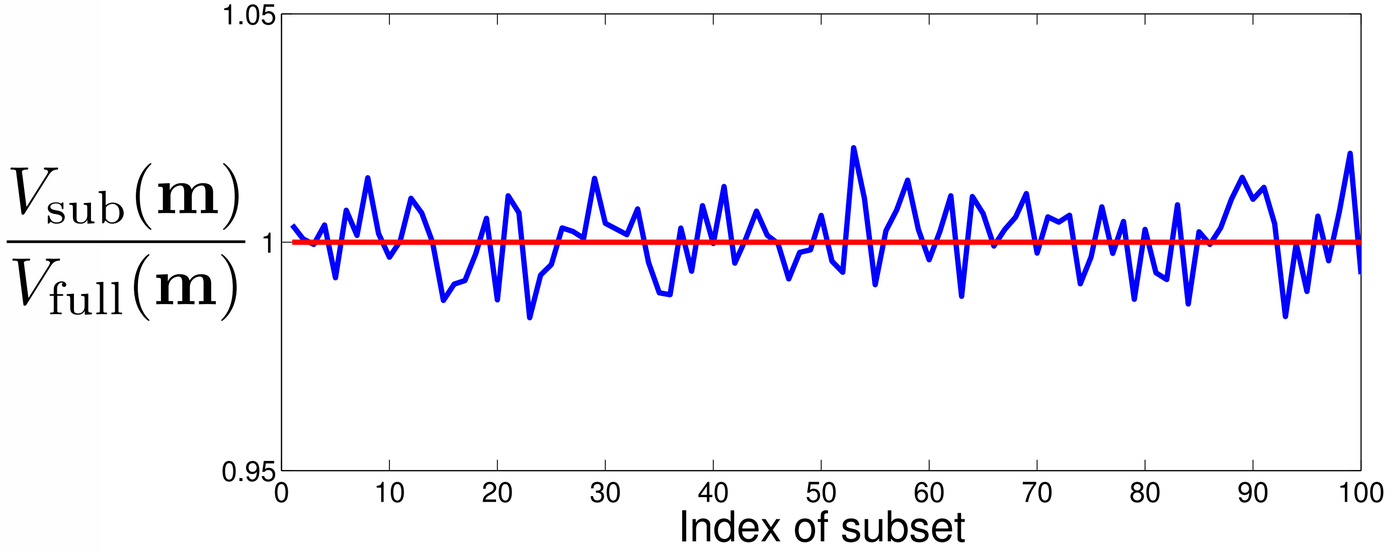
$$\frac{1}{N_s}\sum_{1}^{N_s} \|f_i(\mathbf{m}) - \mathbf{d}_{i\mathrm{obs}}\|^2_{\Gamma^{-1}_{i\mathrm{noise}}} = \frac{1}{\|\mathcal{I}_s\|}\sum_{i\in\mathcal{I}_s} \|f_i(\mathbf{m}) - \mathbf{d}_{i\mathrm{obs}}\|^2_{\Gamma^{-1}_{i\mathrm{noise}}} + \epsilon$$

(Farbod Roosta-Khorasani *et al*, 2013; Eldad *et al,* 2012; Michael P. Friedlander *et al*, 2013, van Leeuwen *et al*, 2013)

Randomized source subsampling

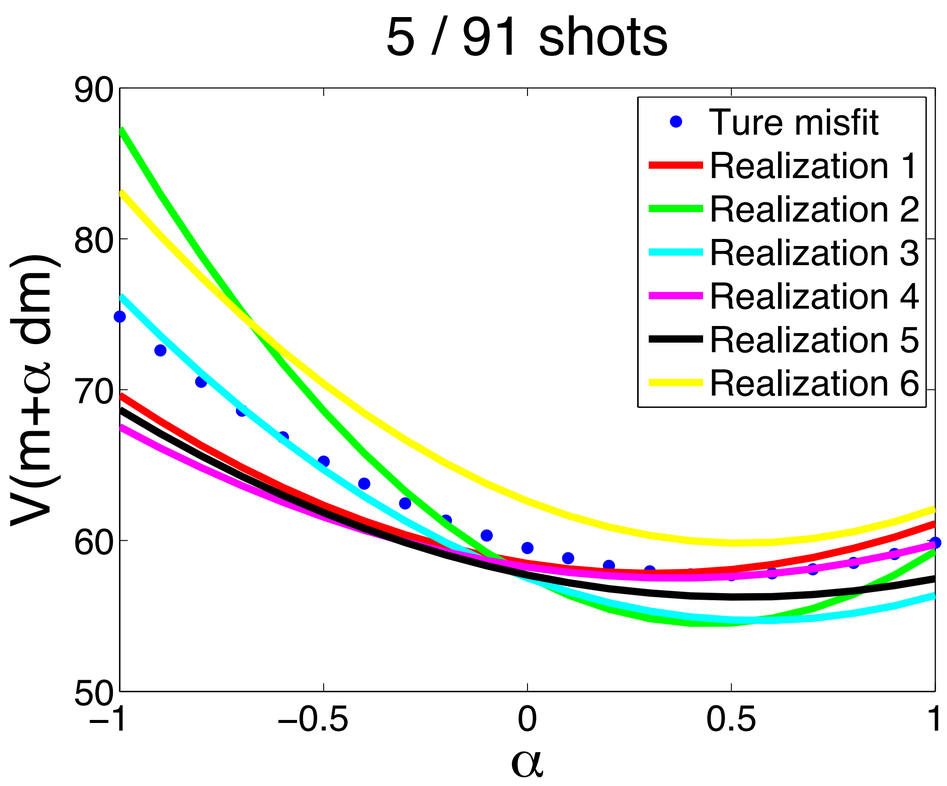
Computational Cost: $\mathcal{O}(N_s) \rightarrow \mathcal{O}(N_{rs}),\ N_{rs} << N_s$

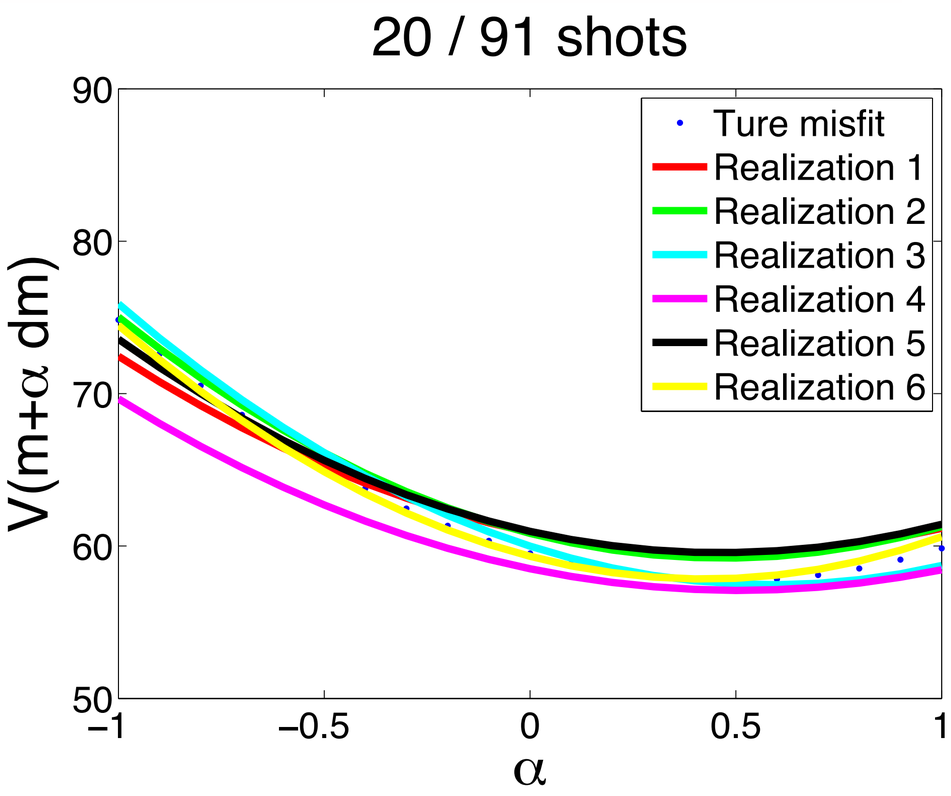
Subset misfit *vs* Full misfit



Randomized source subsampling

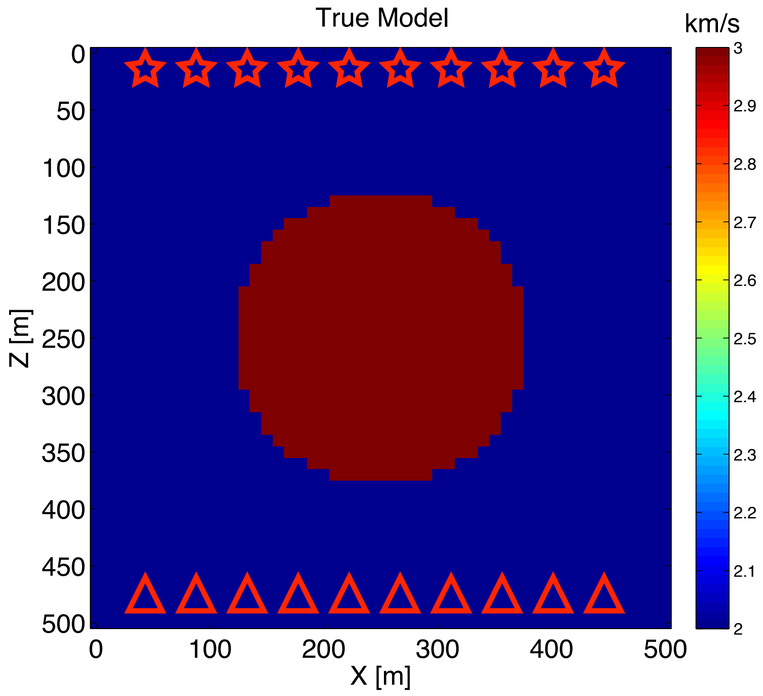
Computational Cost: $\mathcal{O}(N_s) \rightarrow \mathcal{O}(N_{rs})$





Numerical Experiments

Camambert model



☆ - source

△ - receiver

Statistical parameter to be inverted:

Standard deviation : σ

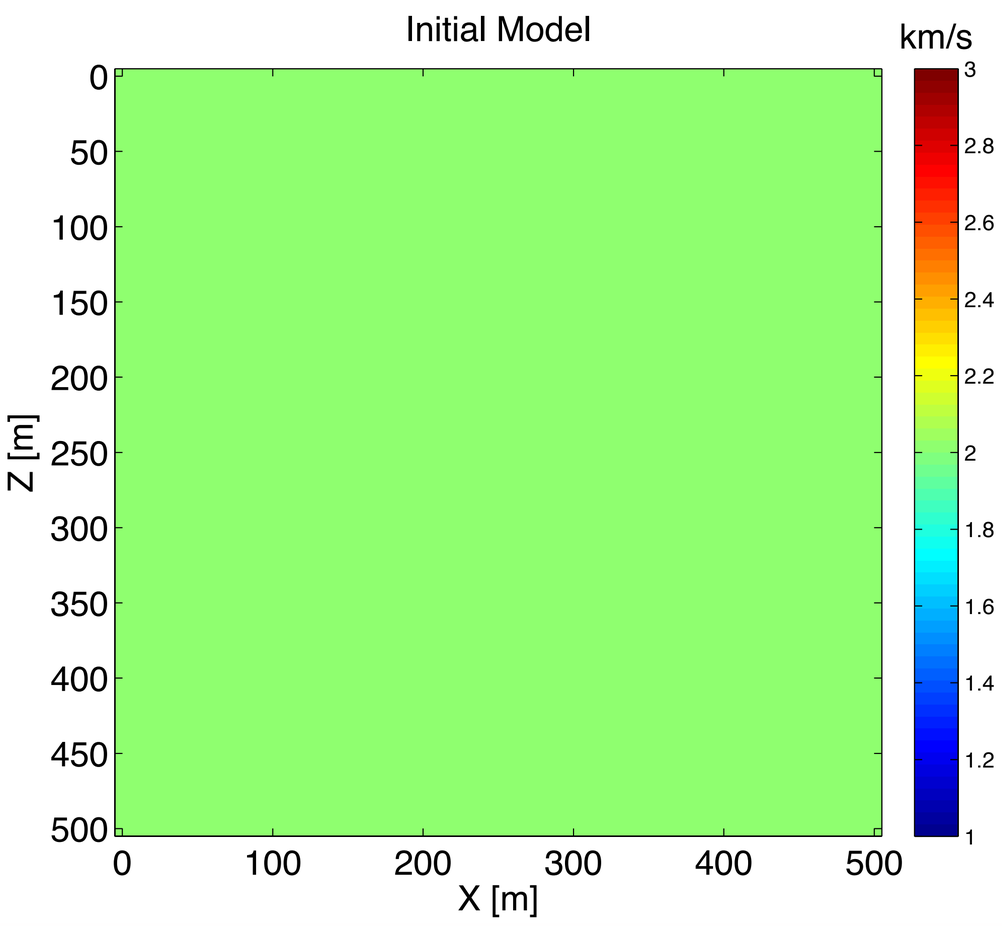
Confidence interval: $P(\mathbf{m} \in \mathrm{I}_{ci}) \geq \alpha$

Acquisition Geometry:

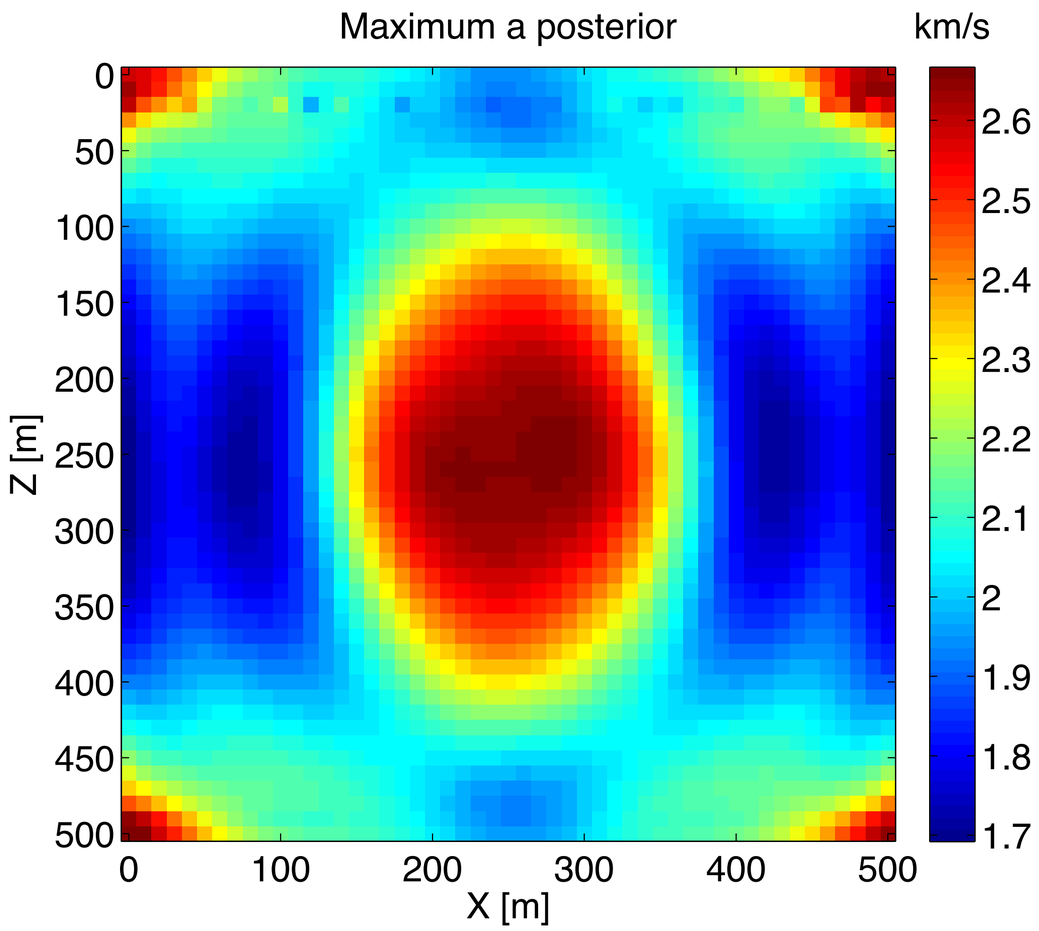
26 shots

51 receivers

10 frequencies

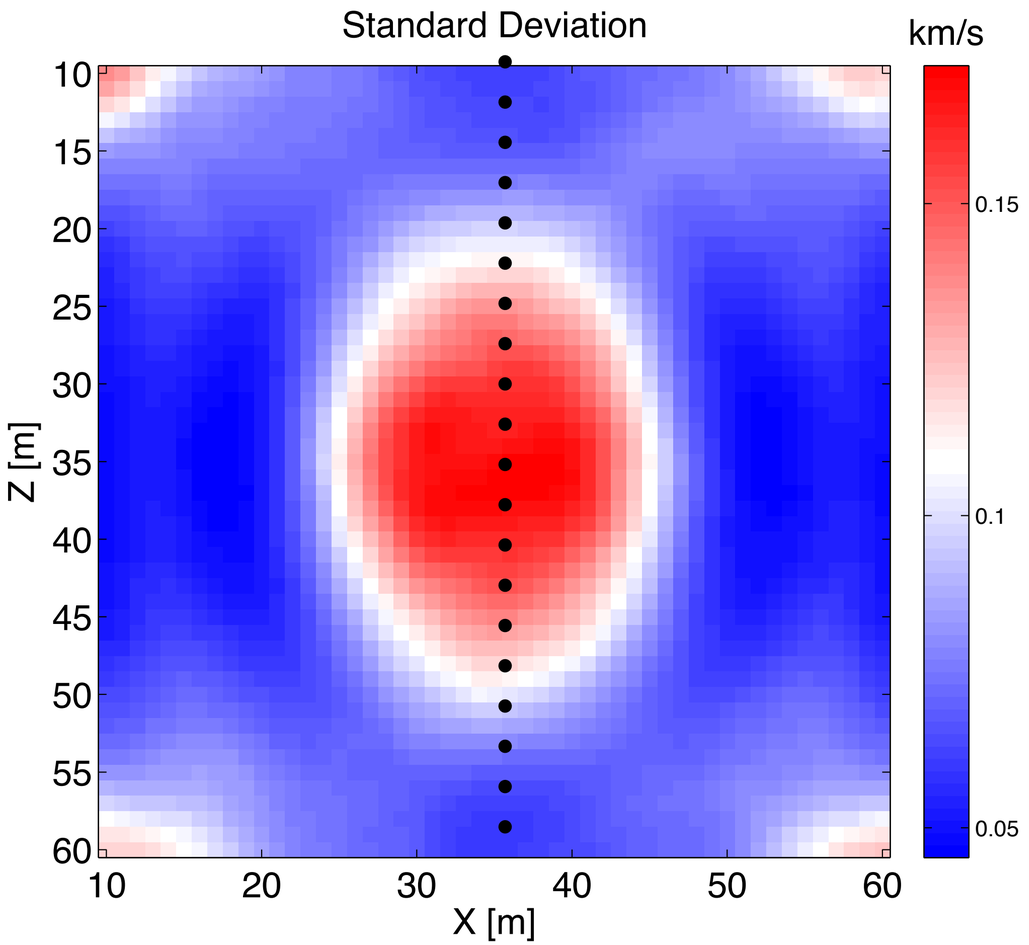


Initial model

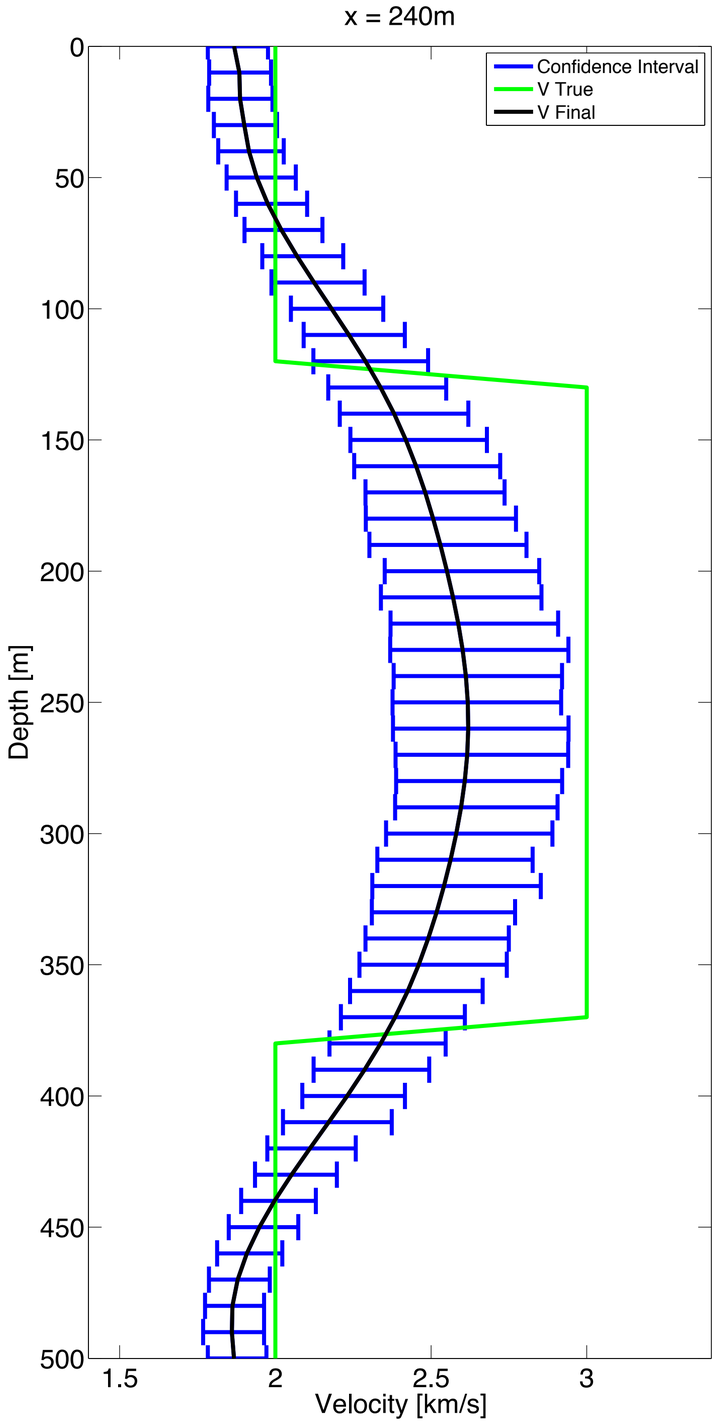


MAP

Nrs = 5



Standard Deviation



$\alpha = 0.9$

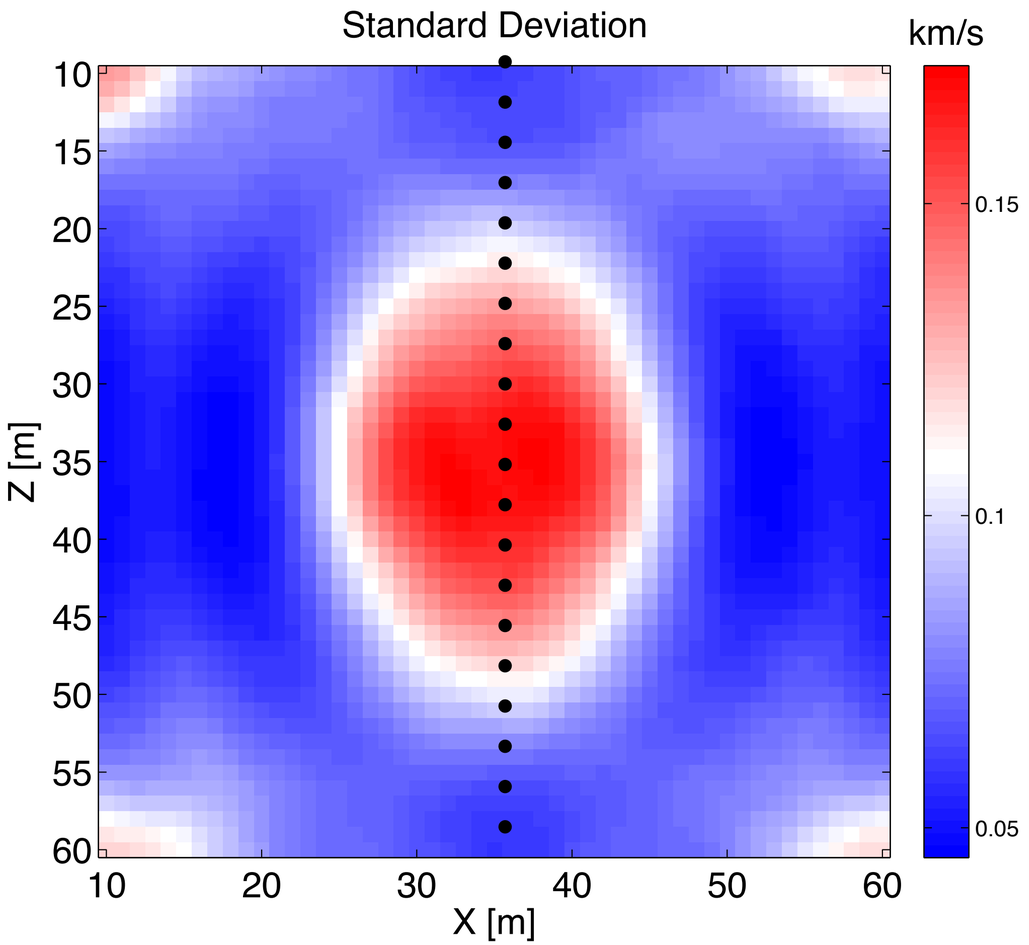
Final Model

Confidence Interval

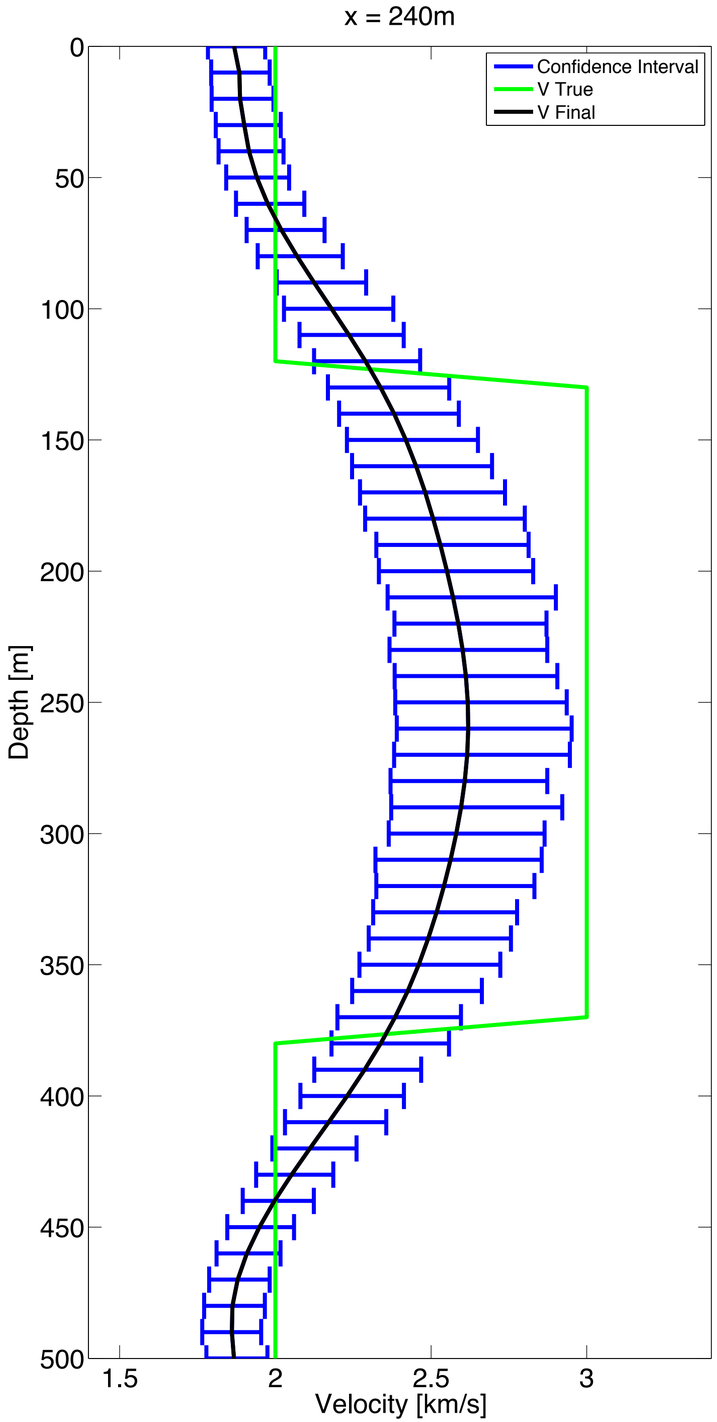
True Model

Confidence Interval

Nrs = 10



Standard Deviation



$\alpha = 0.9$

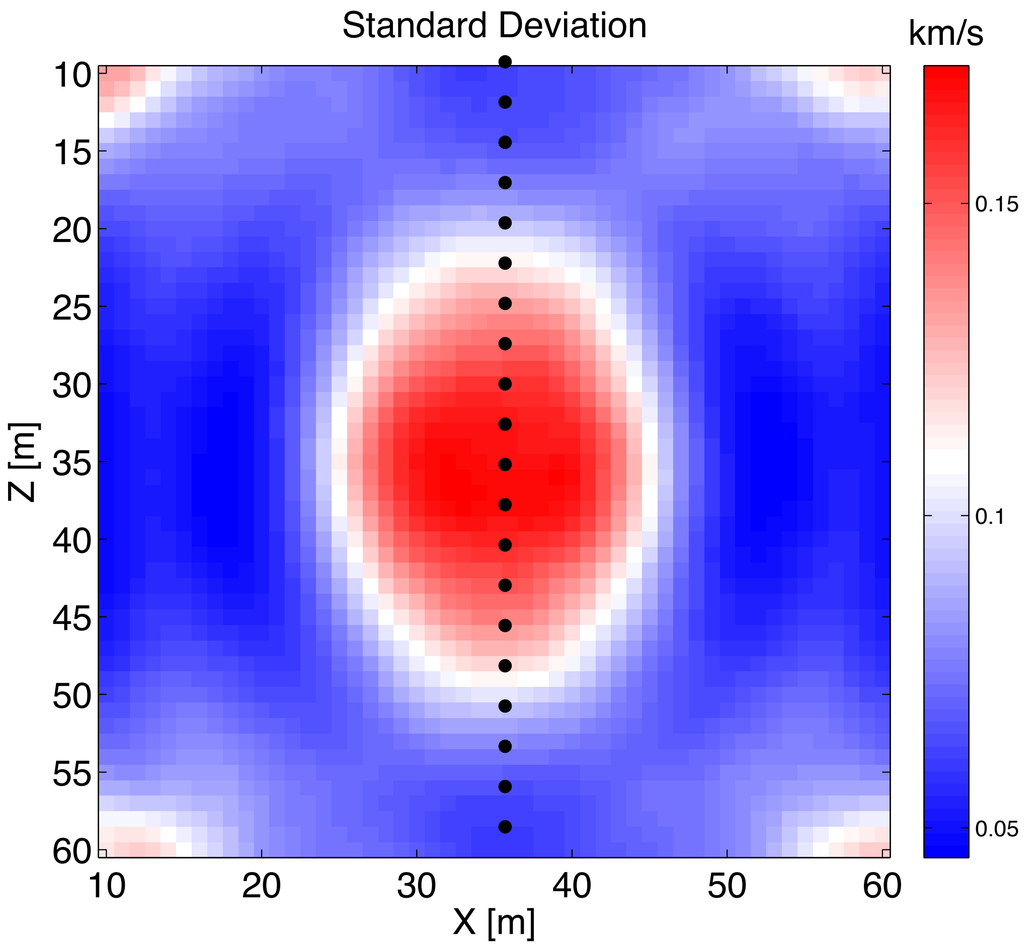
Final Model

Confidence Interval

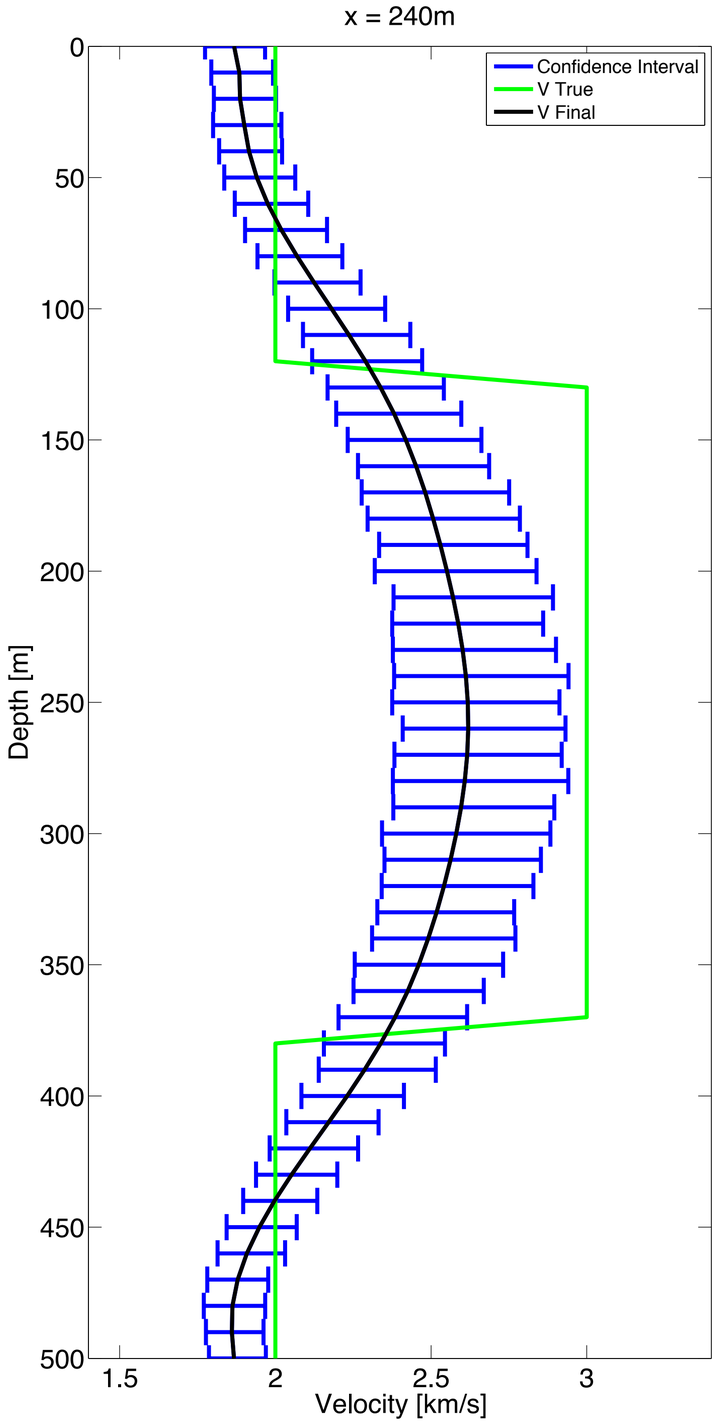
True Model

Confidence Interval

Nrs = 26



Standard Deviation



$\alpha = 0.9$

Final Model

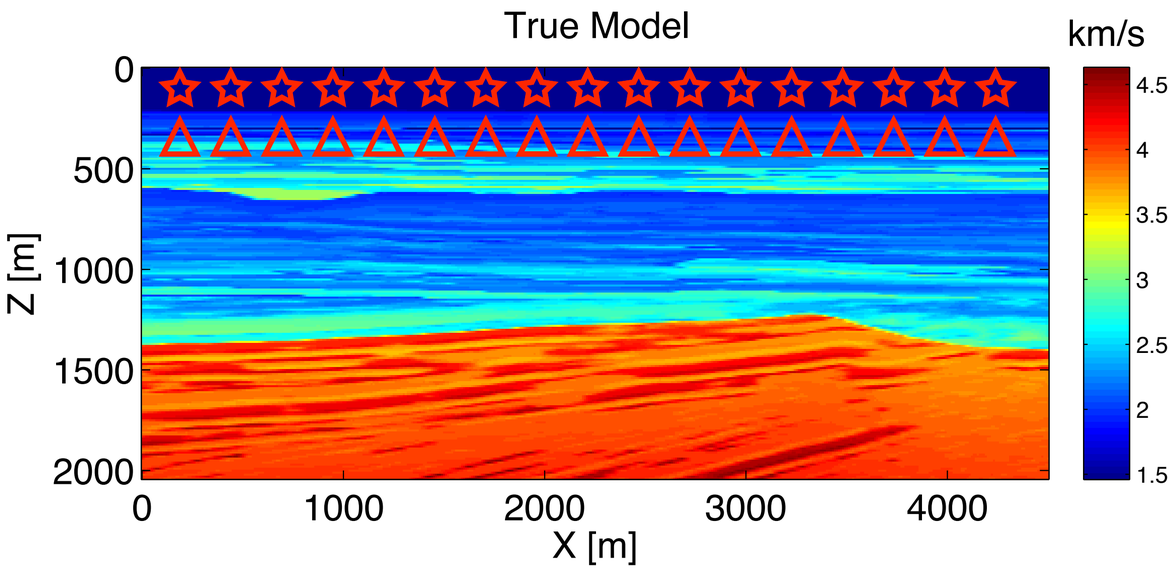
Confidence Interval

True Model

Confidence Interval

Numerical Experiments

BG model



☆ - source

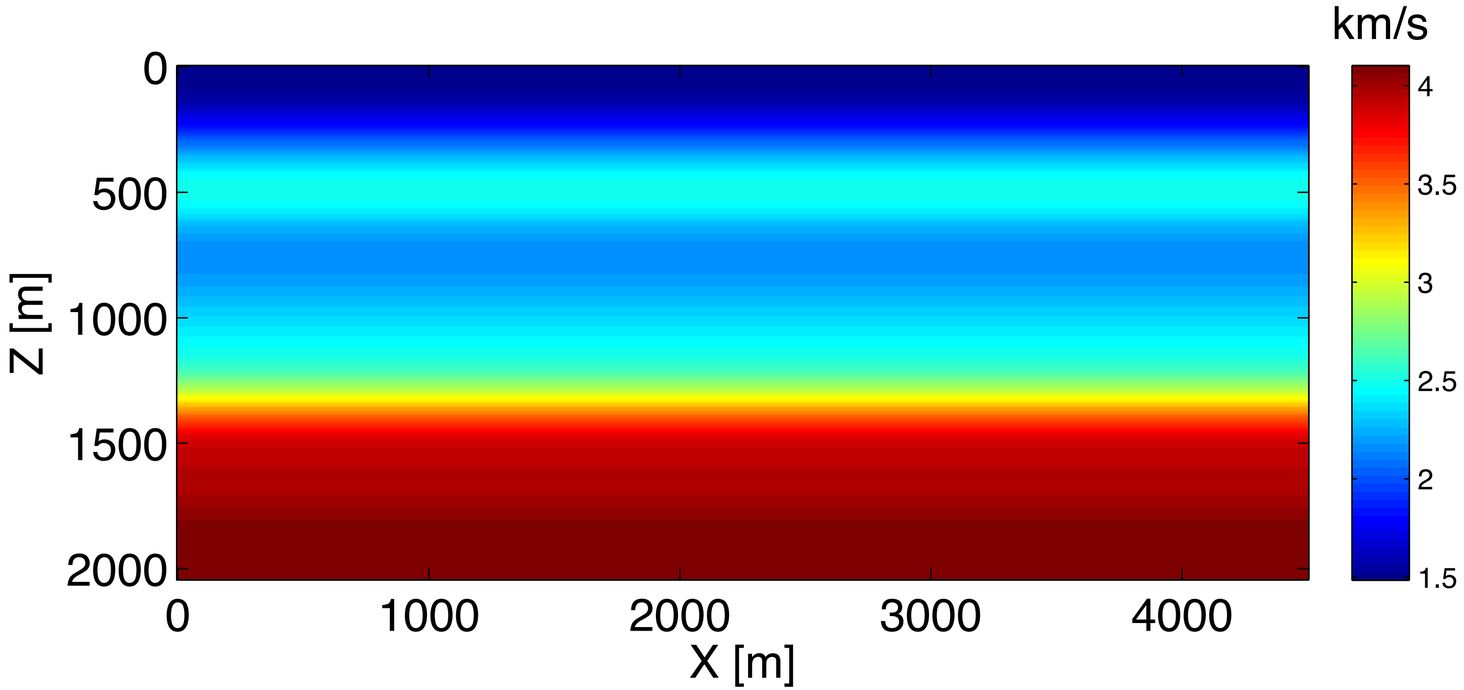
△ - receiver

Acquisition Geometry:

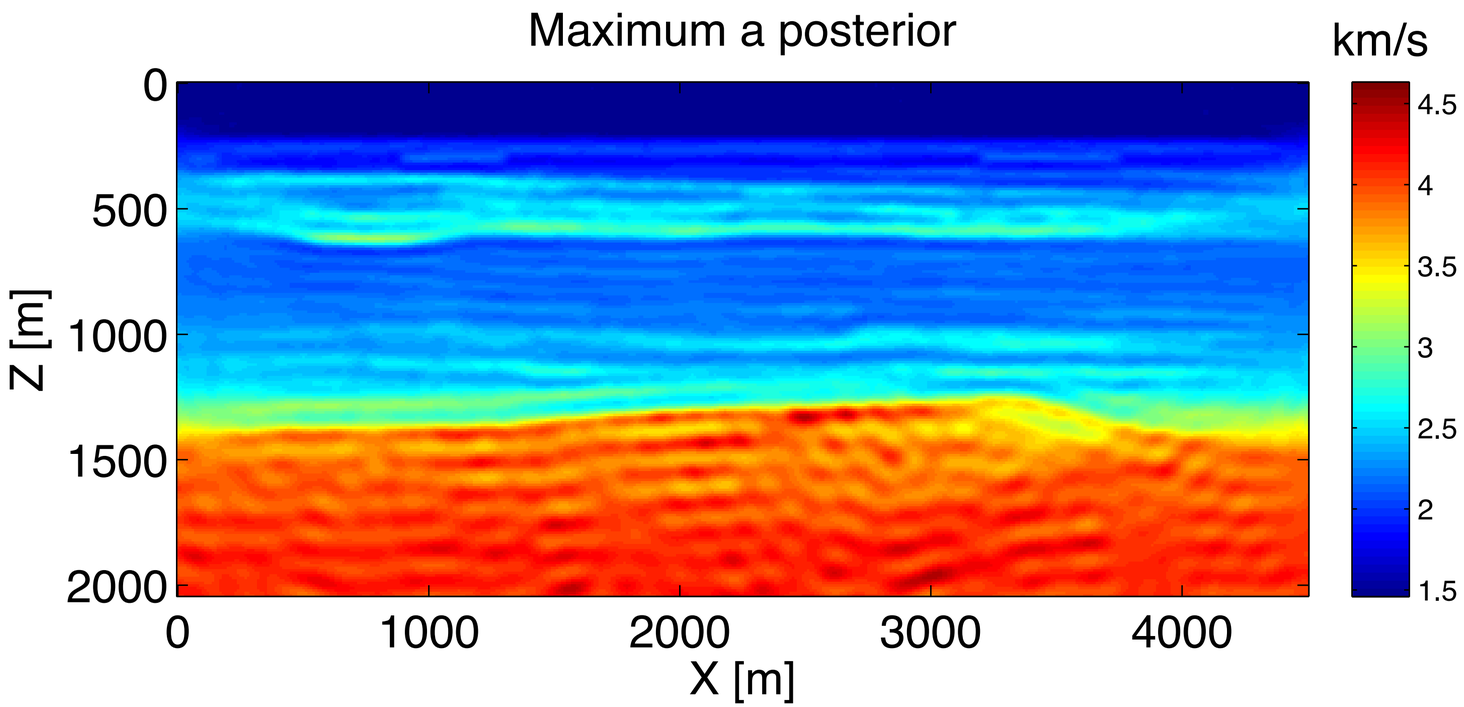
91 shots

451 receivers

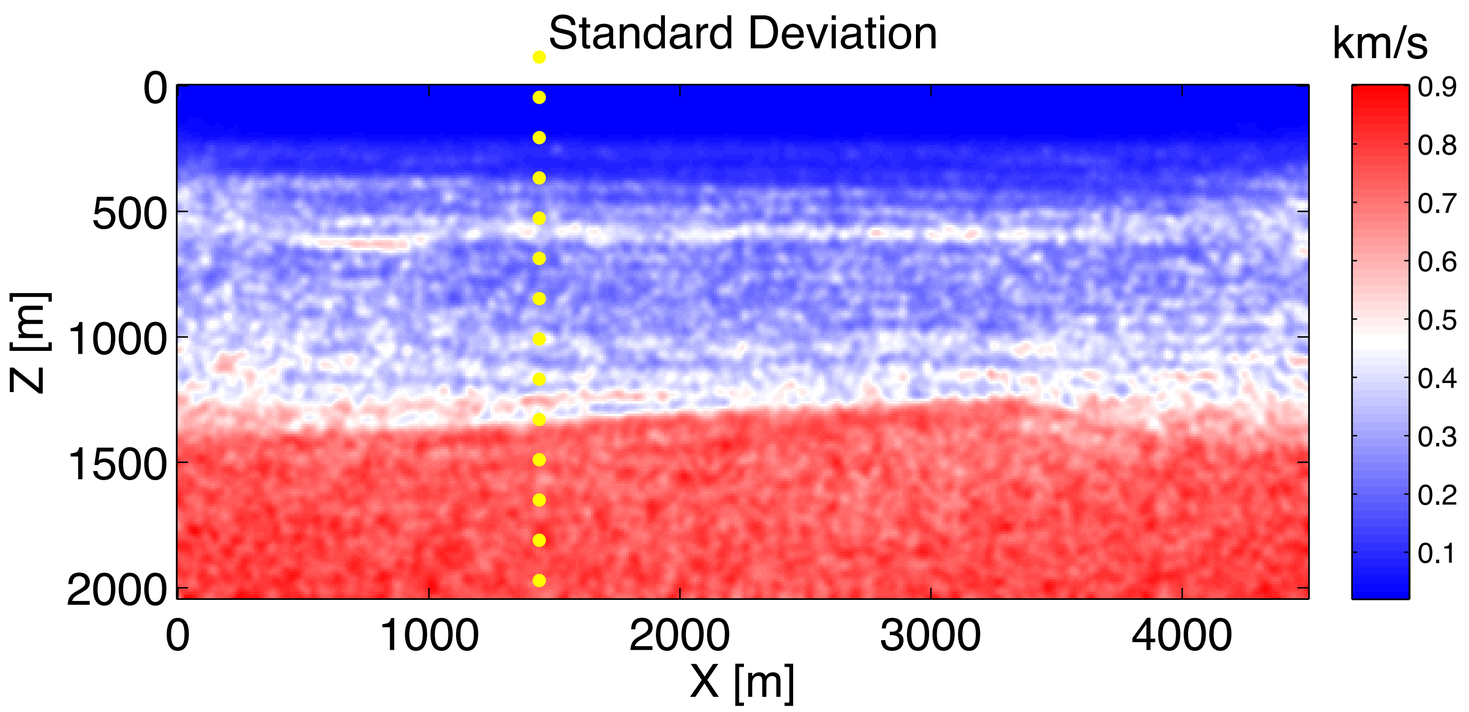
15 frequencies from 3Hz to 17Hz



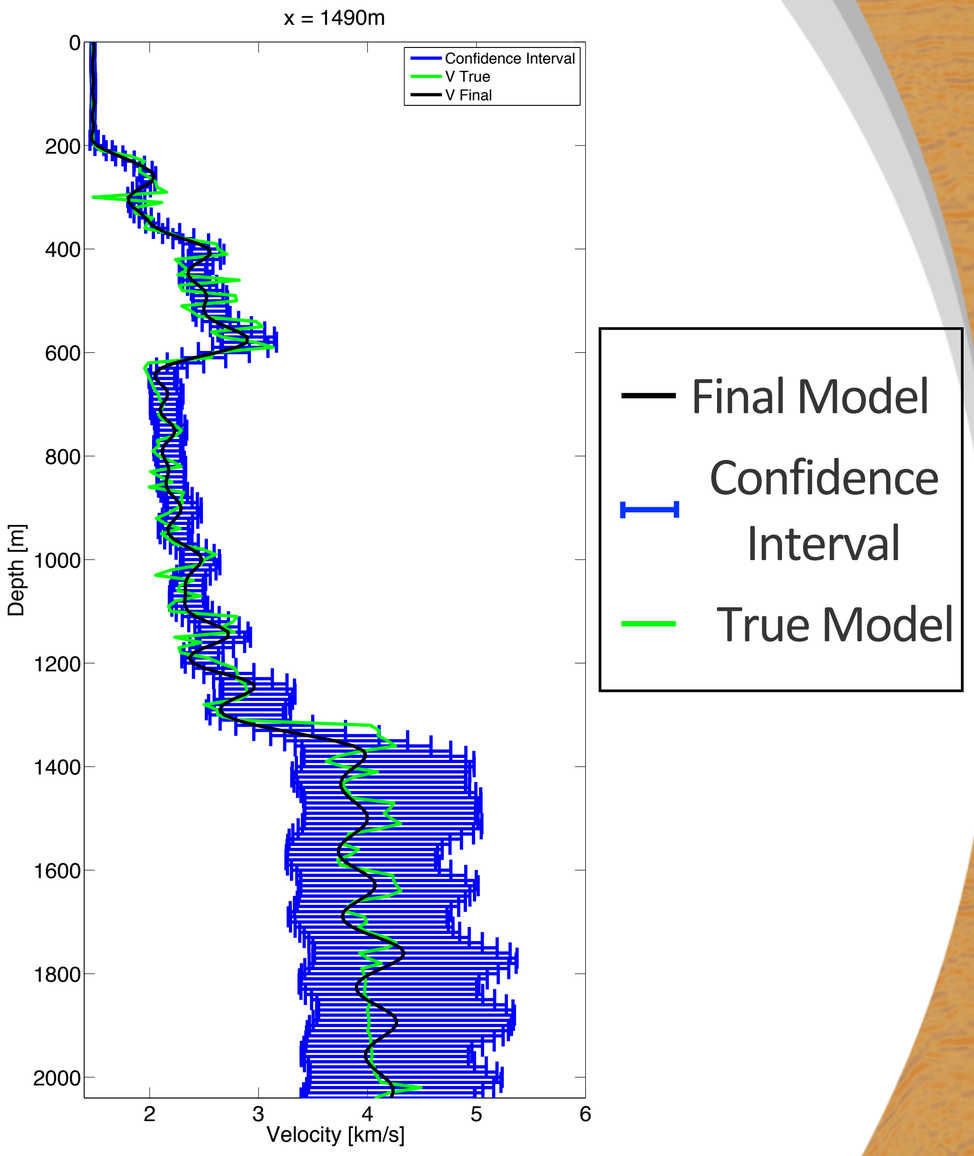
Initial model



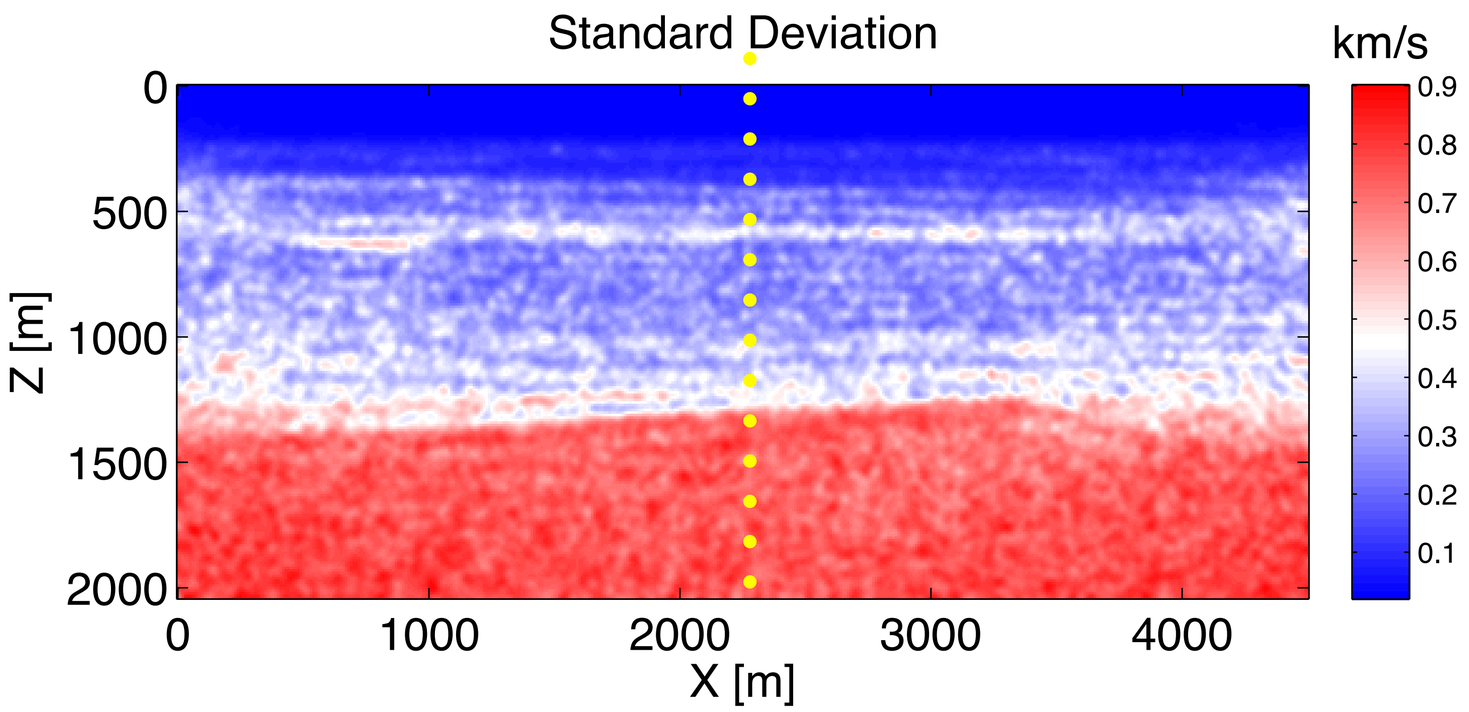
MAP



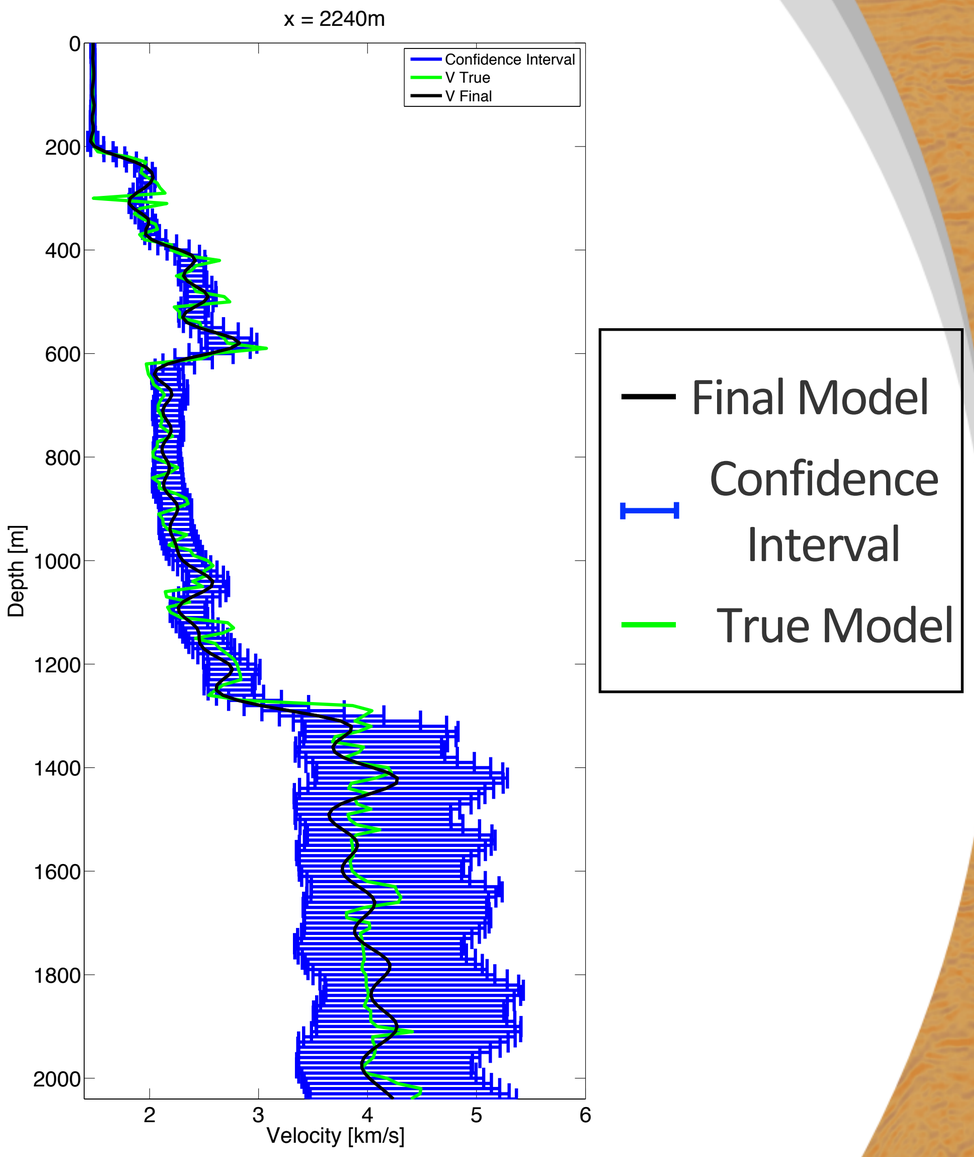
Standard deviation



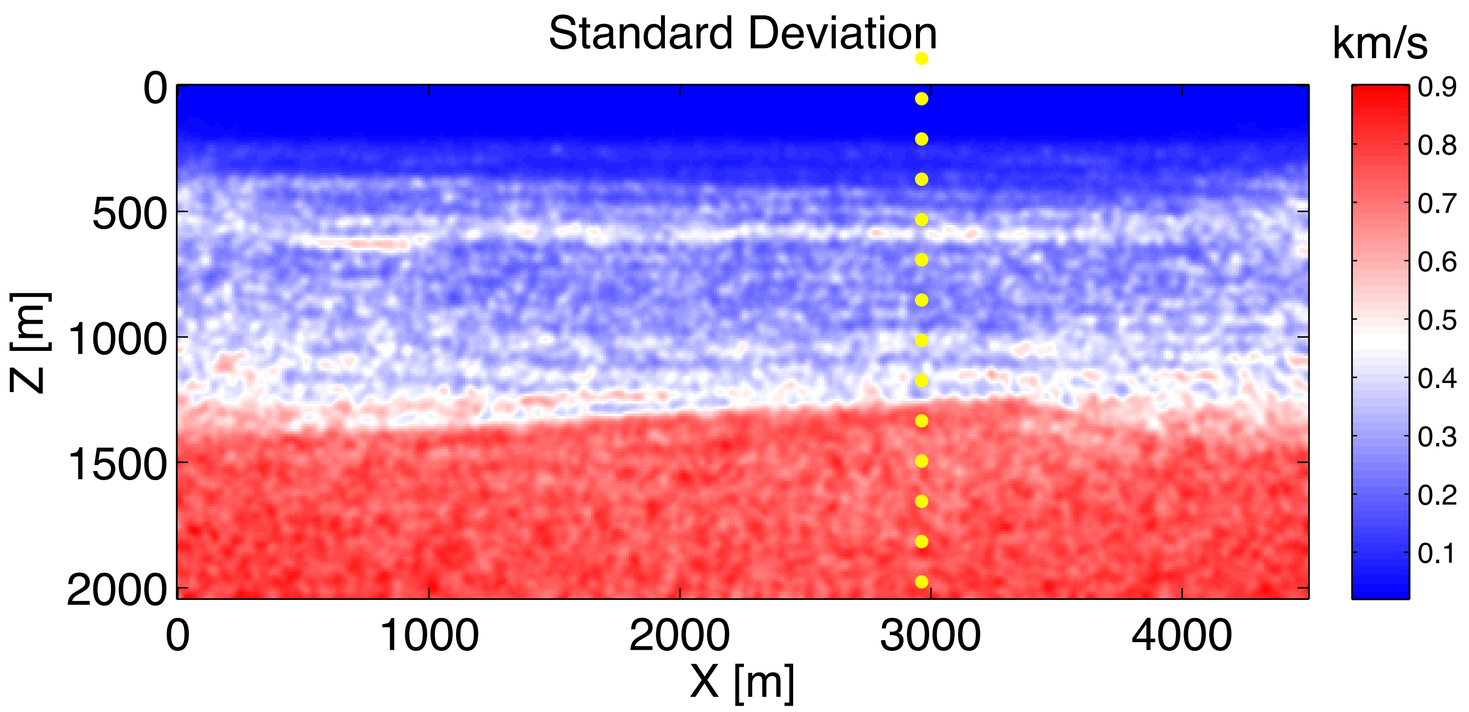
x = 1490m



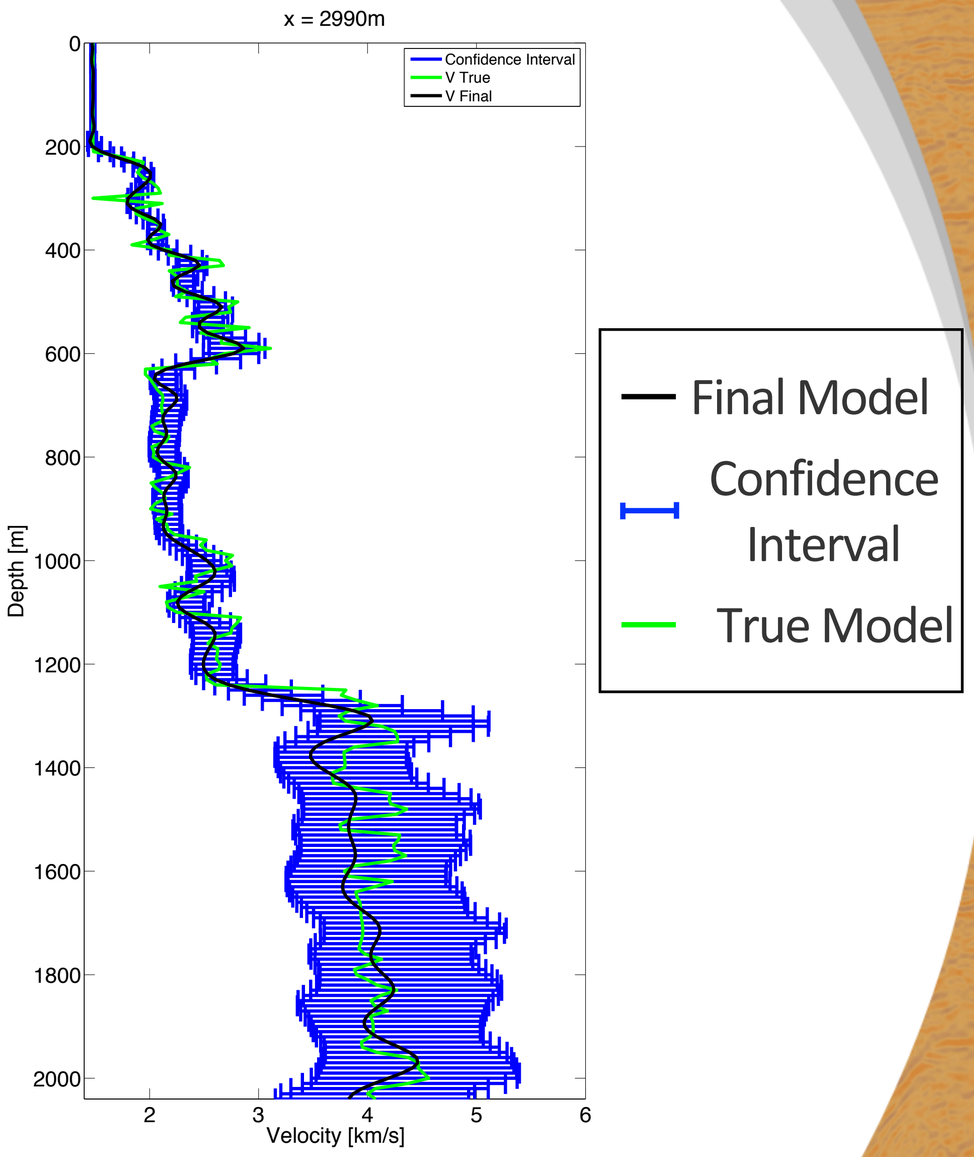
Standard deviation



x = 2240m



Standard deviation



x = 2990m

Conclusions

- Using the L-BFGS Hessian, we reduce the computational cost of estimating the Hessian.
- Using the randomized source sub-sampling method, we decrease the computational cost of the McMC.

Future Work

- Use penalty method frame work to reset up the probability distribution function.
- Use different sampling method to analyze the posterior distribution.
- Uncertainty quantification for 3D FWI.

Acknowledgements

Thank you for your attention !

https://www.slim.eos.ubc.ca/

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Statoil, Total SA, WesternGeco, and Woodside.