A scaled gradient projection method for total variation regularized full waveform inversion

Ernie Esser, Tristan van Leeuwen, Aleksandr Y. Aravkin and Felix J. Herrmann

SINBAD Consortium Meeting, June 20, 2014

Acoustic Full Waveform Inversion in Frequency Domain

$$\min_{m,u} \sum_{sv} \frac{1}{2} \|Pu_{sv} - d_{sv}\|^2 \quad \text{s.t.} \quad A_v(m)u_{sv} = q_{sv}$$ [Tarantola 1984, Virieux and Operto 2009]

where $A_v(m)u_{sv} = q_{sv}$ is the discrete Helmholtz equation

$$A_v(m) = \omega_v^2 \,\mathrm{diag}(m) + L \ .$$

- ω_v is angular frequency and L is a discrete Laplacian
- $s = 1, ..., N_s$ is the source index and $v = 1, ..., N_v$ is the frequency index
- m is the model, the reciprocal of velocity squared
- N is the number of points in the spatial discretization
- $u_{sv} \in \mathbb{C}^N$ denotes the wavefield for source s and frequency v
- $q_{sv} \in \mathbb{C}^N$ denotes the sources
- $d_{sv} \in \mathbb{C}^{N_r}$ denotes the observed data
- P projects the wavefields onto the N_r receiver locations

Wavefield Reconstruction Inversion

Relax the PDE constraint to a quadratic penalty and solve

$$\min_{m,u} \sum_{sv} \frac{1}{2}\|Pu_{sv} - d_{sv}\|^2 + \frac{\lambda^2}{2}\|A_v(m)u_{sv} - q_{sv}\|^2$$ [van Leeuwen and Herrmann 2013]

A natural alternating minimization strategy is to iterate

$$\bar{u}_{sv}(m^n) = \arg\min_{u_{sv}} \frac{1}{2}\|Pu_{sv} - d_{sv}\|^2 + \frac{\lambda^2}{2}\|A_v(m^n)u_{sv} - q_{sv}\|^2 \quad \text{for all } s, v$$

$$m^{n+1} = \arg\min_{m} \sum_{sv} \frac{\lambda^2}{2}\|L\bar{u}_{sv}(m^n) + \omega_v^2 \operatorname{diag}(\bar{u}_{sv}(m^n))m - q_{sv}\|^2$$

Gauss Newton Interpretation

Let $G(m) = \sum_{sv} G_{sv}(m)$, where

$$G_{sv}(m) = \frac{1}{2}\|P\bar{u}_{sv}(m) - d_{sv}\|^2 + \frac{\lambda^2}{2}\|A_v(m)\bar{u}_{sv}(m) - q_{sv}\|^2 \; .$$

Then alternating minimization can be interpreted (formally) as a Gauss Newton method for minimizing $G(m)$.

Computing the Gradient and Gauss Newton Hessian

Using a variable projection argument [Aravkin and van Leeuwen 2012],

$$\nabla G(m^n) = \sum_{sv} \mathrm{Re}\left(\lambda^2 \omega_v^2 \operatorname{diag}(\bar{u}_{sv}(m^n))^* (\omega_v^2 \operatorname{diag}(\bar{u}_{sv}(m^n)) m^n + L\bar{u}_{sv}(m^n) - q_{sv})\right)$$

The Gauss Newton approximation to the Hessian of G at m^n is **diagonal**, given by

$$H^n = \sum_{sv} \mathrm{Re}(\lambda^2 \omega_v^4 \operatorname{diag}(\bar{u}_{sv}(m^n))^* \operatorname{diag}(\bar{u}_{sv}(m^n)))$$ [van Leeuwen and Herrmann 2013]

Scaled Gradient Descent Framework

A scaled gradient descent approach [Bertsekas 1999] for minimizing G can be written as

$$\Delta m = \arg \min_{\Delta m \in R^N} \sum_{sv} \Delta m^T \nabla G_{sv}(m^n) + \frac{1}{2} \Delta m^T H_{sv}^n \Delta m + c_n \Delta m^T \Delta m$$

$$m^{n+1} = m^n + \Delta m \ .$$

In addition to Gauss Newton, this general framework includes gradient descent and Newton's method.

Including Convex Constraints

Add the constraint $m \in C$, where C is a convex set and iterate

$$\Delta m = \arg \min_{\Delta m \in R^N} \sum_{sv} \Delta m^T \nabla G_{sv}(m^n) + \frac{1}{2} \Delta m^T H_{sv}^n \Delta m + c_n \Delta m^T \Delta m$$
$$\text{s.t. } m^n + \Delta m \in C$$

$$m^{n+1} = m^n + \Delta m \ .$$

Bound Constraints without Increasing Computational Cost

Bound constraint example: $C = \{m \in \mathbb{R}^N : m_j \in [B_1, B_2]\}$.

When H is diagonal and positive, the update for Δm is simple.

$$\Delta m = \arg\min_{\Delta m \in R^N} \Delta m^T \nabla G(m^n) + \frac{1}{2} \Delta m^T H \Delta m$$
$$\text{s.t. } m^n + \Delta m \in C$$
$$= H^{-1} \max\left(H(B_1 - m^n), \min\left(H(B_2 - m^n), -\nabla G(m^n) +\right)\right)$$

Total Variation Regularization

If we represent m as a N_1 by N_2 image, we can define

$$\begin{aligned}\|m\|_{TV} &= \frac{1}{h}\sum_{ij}\sqrt{(m_{i+1,j}-m_{i,j})^2+(m_{i,j+1}-m_{i,j})^2}\\ &= \sum_{ij}\frac{1}{h}\left\|\begin{bmatrix}(m_{i,j+1}-m_{i,j})\\(m_{i+1,j}-m_{i,j})\end{bmatrix}\right\|\\ &= \|Dm\|_{1,2}\ ,\end{aligned}$$

where D is a discrete gradient operator applied to a vectorized m.

Proposed Model and Algorithm

Solve

$$\min_m G(m) \qquad \text{s.t.} \qquad m \in [B_1, B_2] \text{ and } \|m\|_{TV} \leq \tau$$

by iterating

$$\Delta m = \arg\min_{\Delta m} \Delta m^T \nabla G(m^n) + \frac{1}{2}\Delta m^T H^n \Delta m + c_n \Delta m^T \Delta m$$
$$\text{s.t.} \qquad m^n + \Delta m \in [B_1, B_2] \text{ and } \|m^n + \Delta m\|_{TV} \leq \tau$$

$$m^{n+1} = m^n + \Delta m \ .$$

Solving the Convex Subproblem

There are many effective primal dual methods for solving the convex subproblem for Δm based on finding a saddle point of the Lagrangian

$$\begin{aligned}\mathcal{L}(\Delta m, p) =& \Delta m^T \nabla G(m^n) + \frac{1}{2}\Delta m^T (H^n + 2c_n \mathrm{I})\Delta m + p^T D(m^n + \Delta m) - \tau \|p\|_{\infty,2} \\ &\text{for} \quad m^n + \Delta m \in [B_1, B_2] \ ,\end{aligned}$$

which can be related to the primal problem by noting that

$$\sup_p p^T Dm - \tau\|p\|_{\infty,2} = \begin{cases} 0 & \text{if } \|Dm\|_{1,2} \leq \tau \\ \infty & \text{otherwise.} \end{cases}$$

Modified PDHG Iterations

The modified PDHG method [Zhu and Chan 2008, Chambolle and Pock 2011, Esser, Zhang and Chan 2010, He and Yuan 2012] finds a saddle point by iterating

$$
\begin{aligned}
p^{k+1} &= \arg\min_{p} \tau \|p\|_{\infty,2} - p^T D(m^n + \Delta m^k) + \frac{1}{2\delta}\|p - p^k\|^2 \\
\Delta m^{k+1} &= \arg\min_{\Delta m} \Delta m^T \nabla G(m^n) + \frac{1}{2}\Delta m^T (H^n + 2c_n \mathrm{I})\Delta m \\
&\qquad + \Delta m^T D^T(2p^{k+1} - p^k) + \frac{1}{2\alpha}\|\Delta m - \Delta m^k\|^2 \\
&\text{s.t.} \qquad m^n + \Delta m \in [B_1, B_2]
\end{aligned}
$$

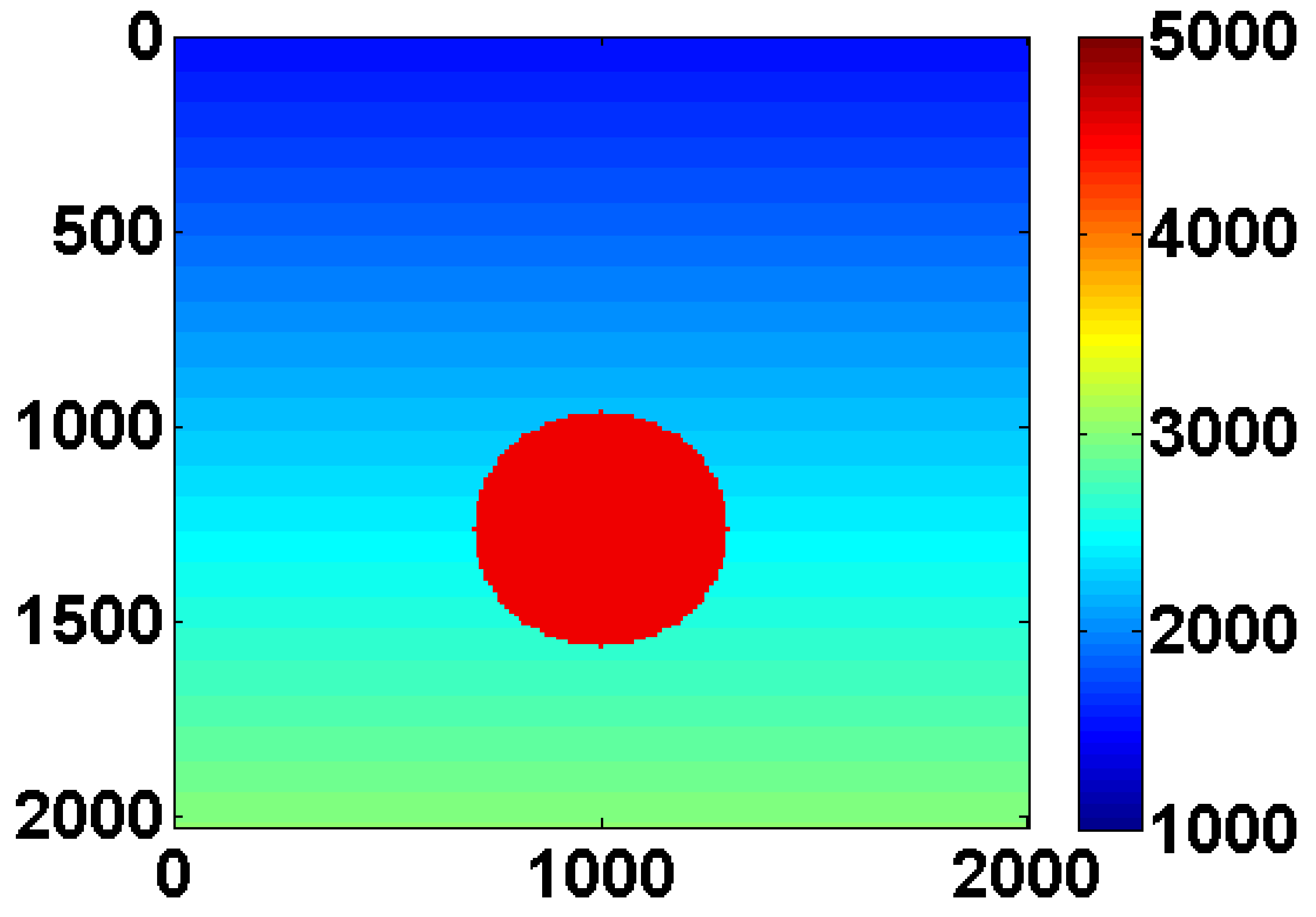
The p^{k+1} update involves a projection that can be efficiently computed, and

$$
\begin{aligned}
\Delta m^{k+1} = (H^n + \xi_n \mathrm{I})^{-1} \max\big((H^n + \xi_n \mathrm{I})(B_1 - m^n), \\
\min\Big((H^n + \xi_n \mathrm{I})(B_2 - m^n), -\nabla G(m^n) + \frac{\Delta m^k}{\alpha} - D^T(2p^{k+1} - p^k)\Big)\big)
\end{aligned}
$$

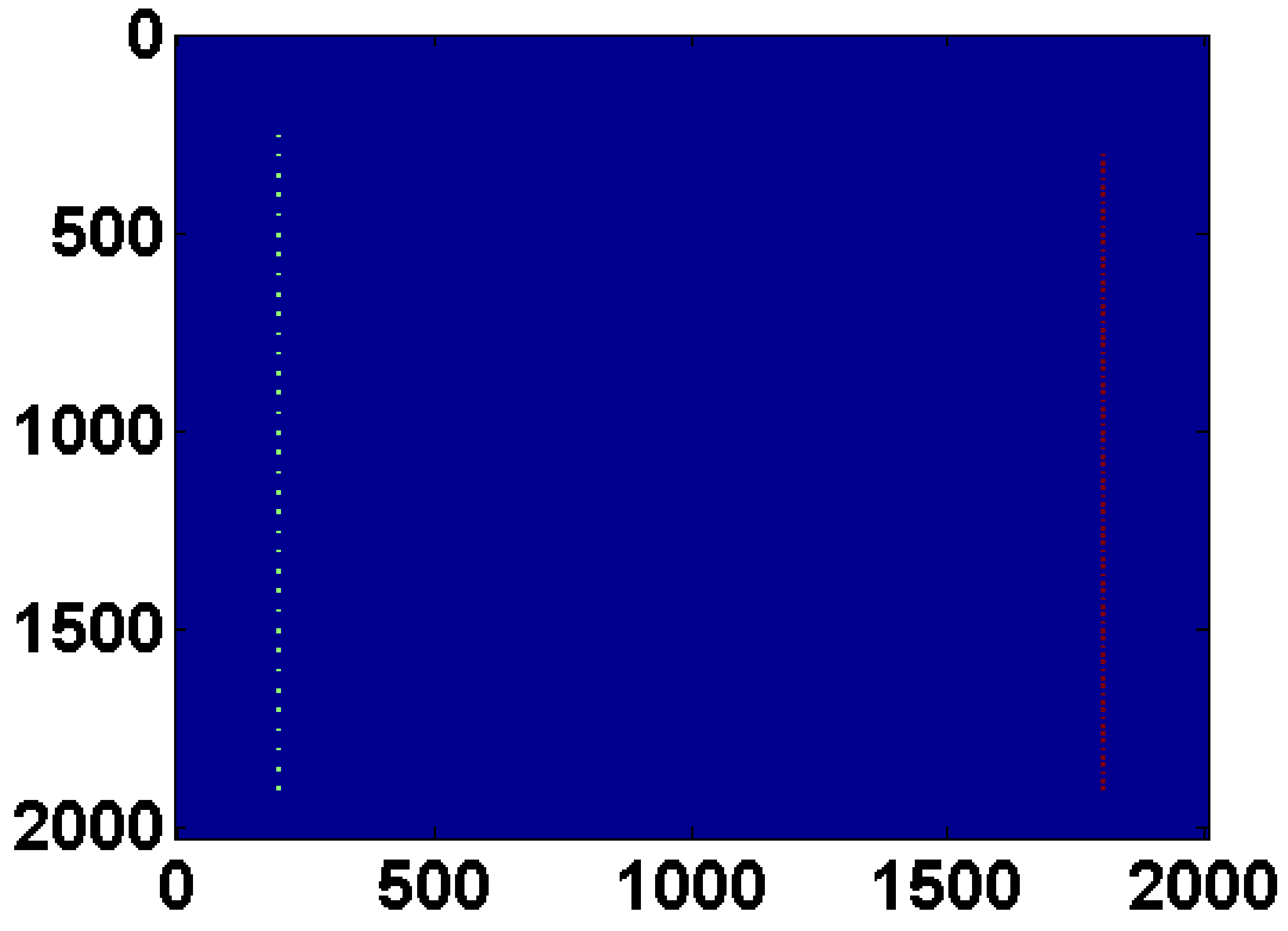
where $\xi_n = 2c_n + \frac{1}{\alpha}$

Numerical Experiment

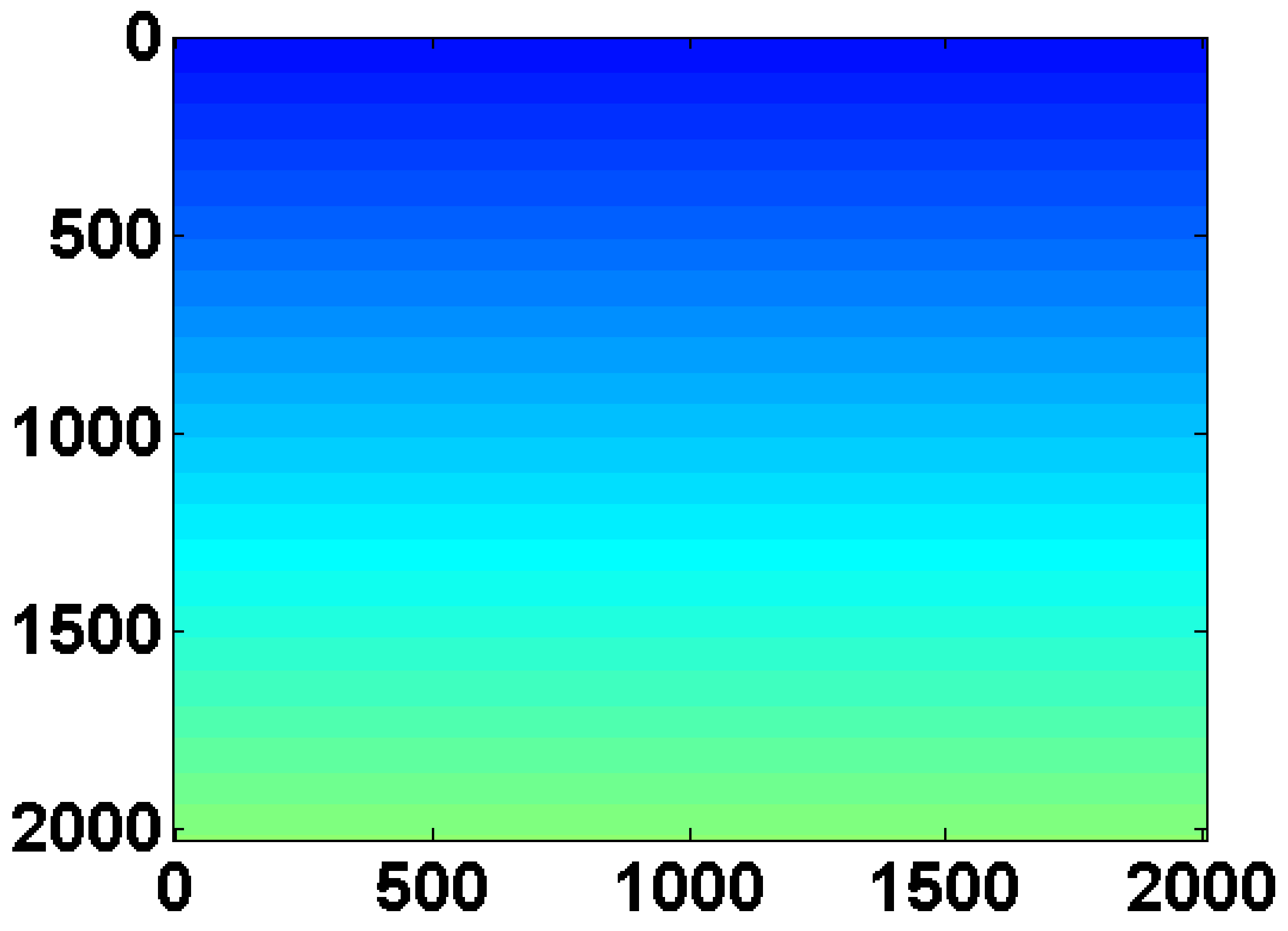
True velocity



Source and receiver locations



Initial velocity

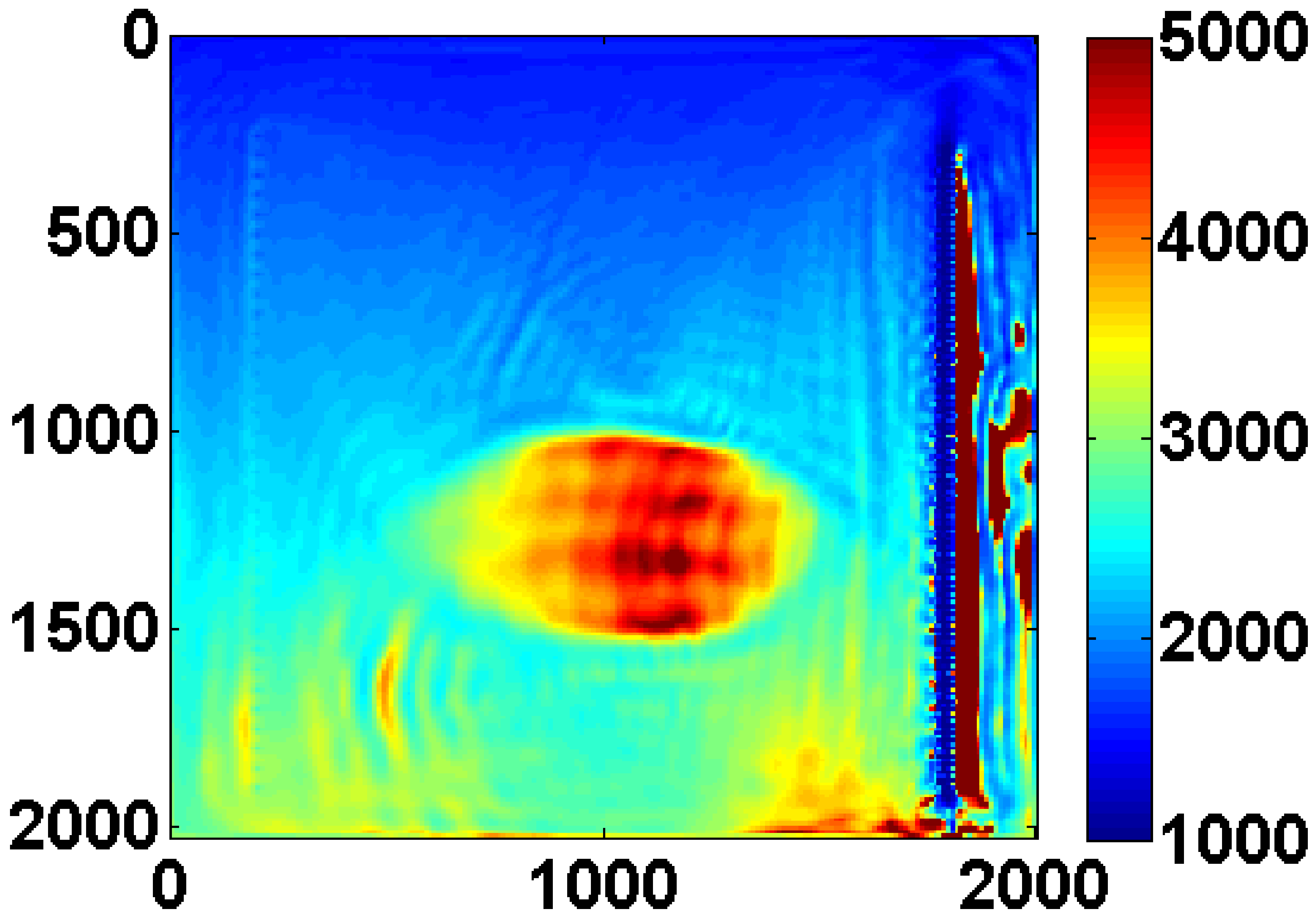


Modeling Details

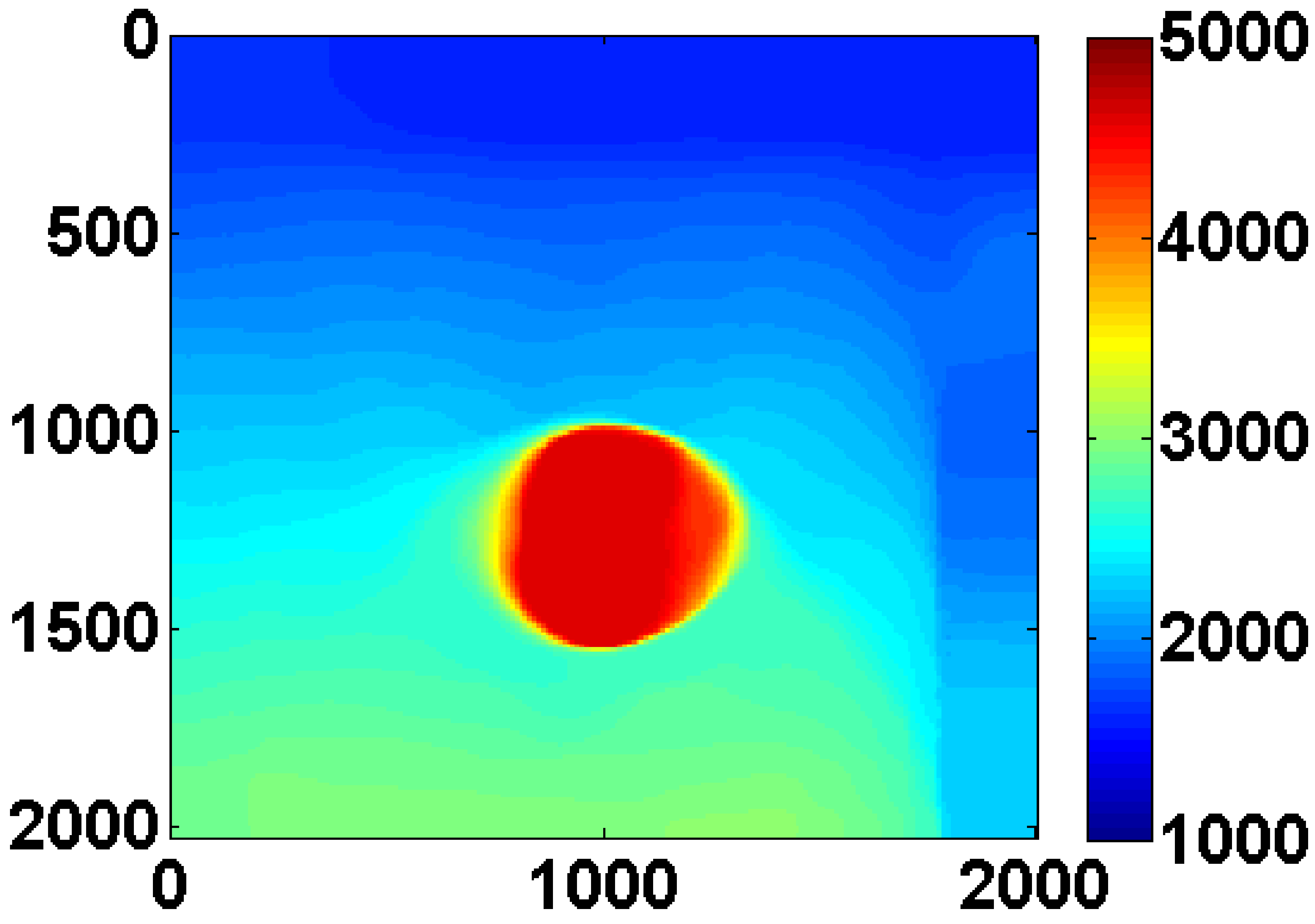
- model size: 203 by 201
- mesh size: 10m
- number of sources: 39
- number of receivers: 96
- frequency range: 3-20Hz in overlapping batches of 2
- maximum number of outer iterations per frequency batch: 25
- maximum number of inner iterations for convex subproblems: 2000
- known Ricker wavelet sources with 30Hz peak frequency
- sequential shots
- added frequency dependent Gaussian noise proportional to magnitude of data

Effect of TV Regularization on Noisy Data

Results from synthetic noisy data using frequencies 3-20Hz in overlapping batches of 2



Without TV constraint

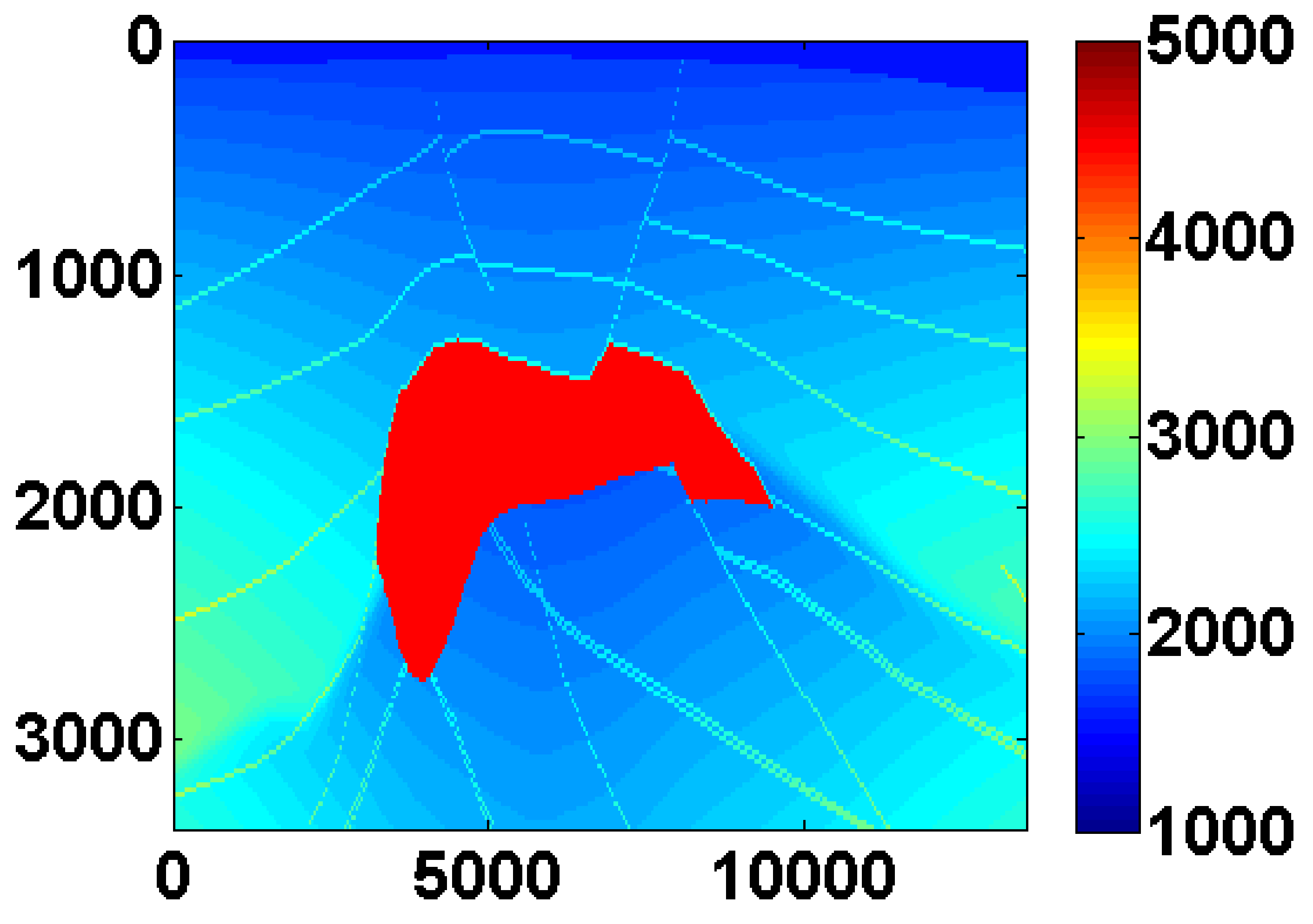


With TV constraint

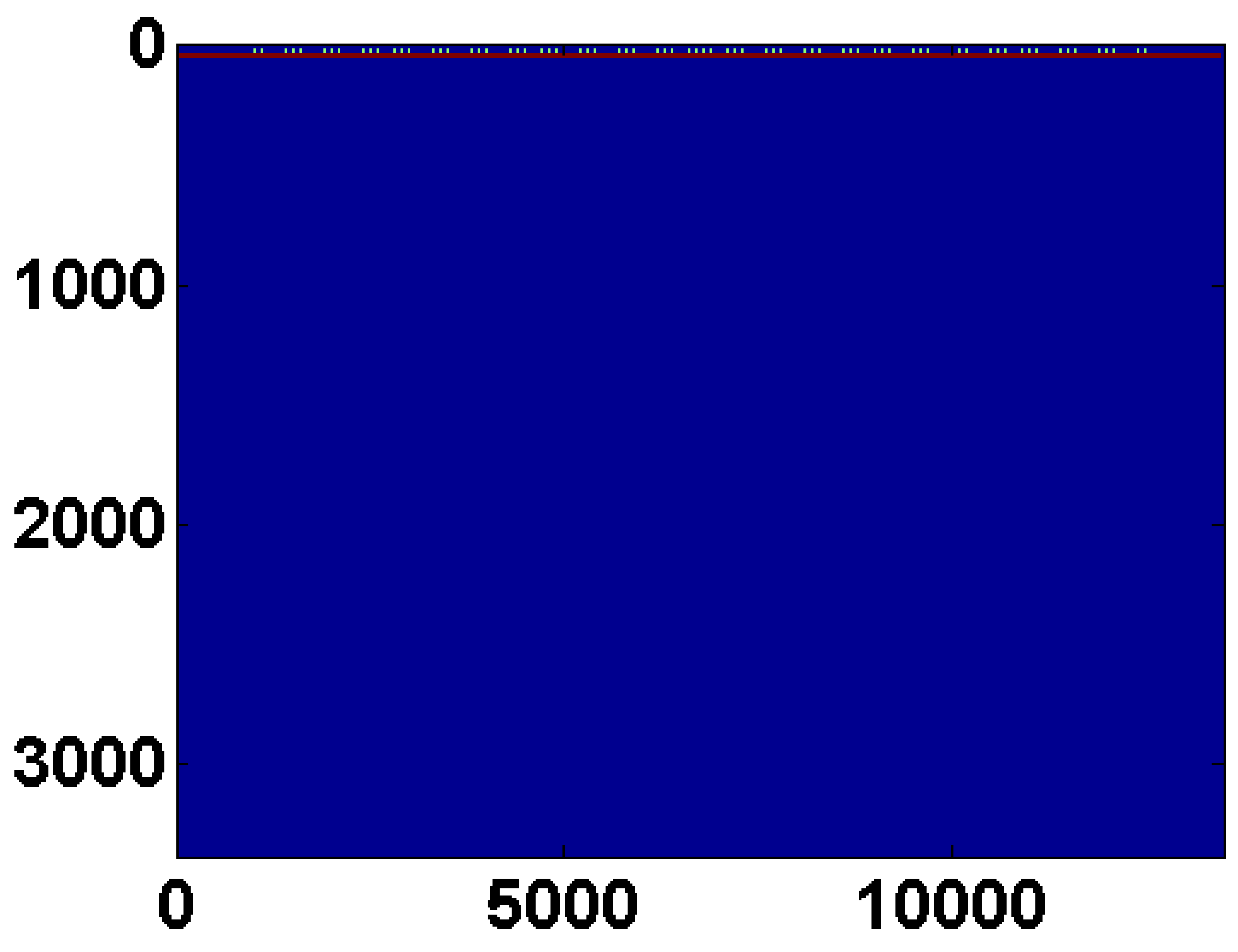
Simultaneous Shot Experiment

SEG/EAGE salt model, sources and receivers near the surface, two simultaneous shots, and a very good initial guess

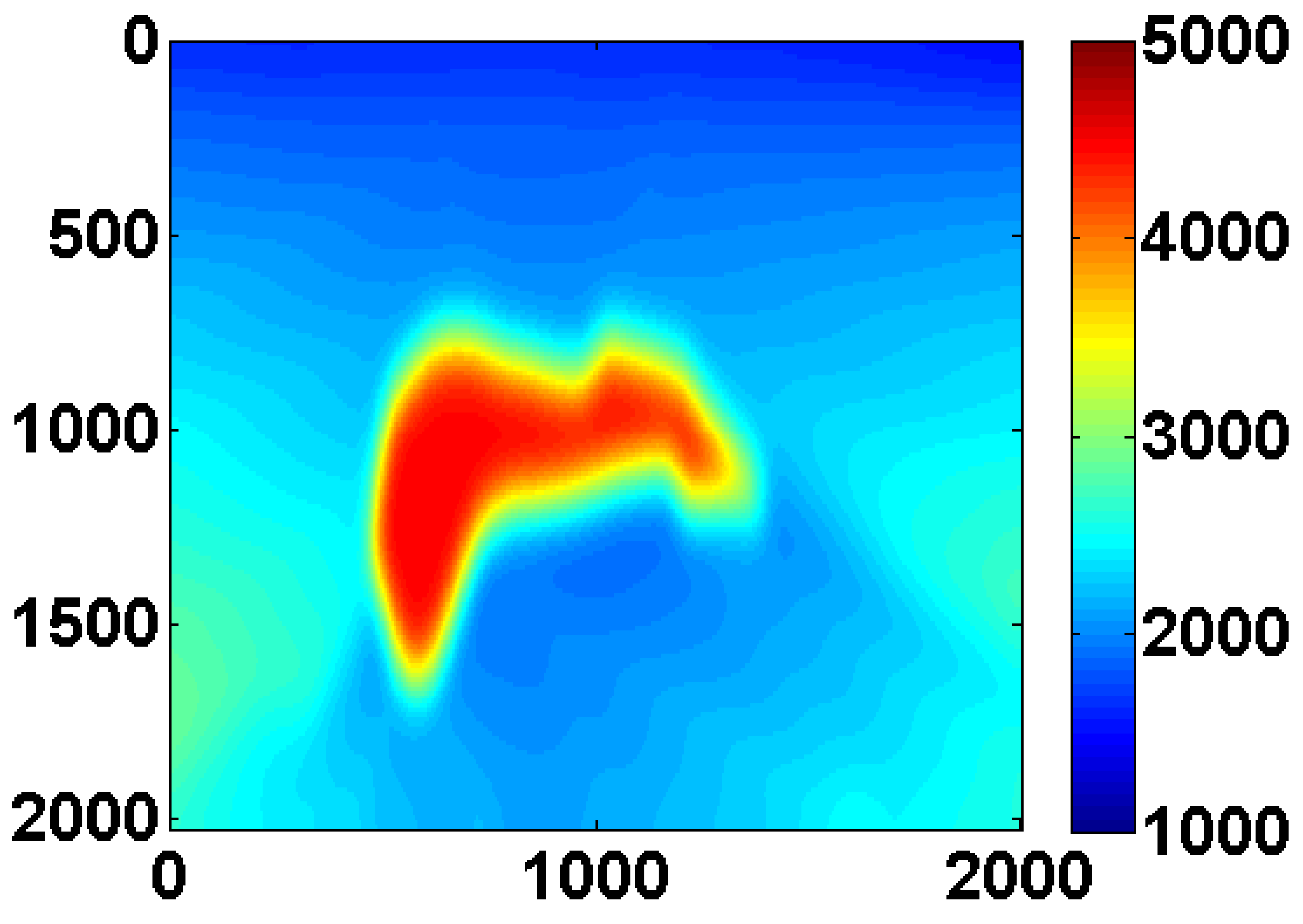
$$\bar{q}_{jv} = \sum_{s=1}^{N_s} w_{js} q_{sv} \quad j = 1, 2 \qquad w_{js} \in \mathcal{N}(0, 1) \qquad \bar{d}_{jv} = P A_v^{-1}(m) \bar{q}_{jv}$$



True velocity



Source and receiver locations

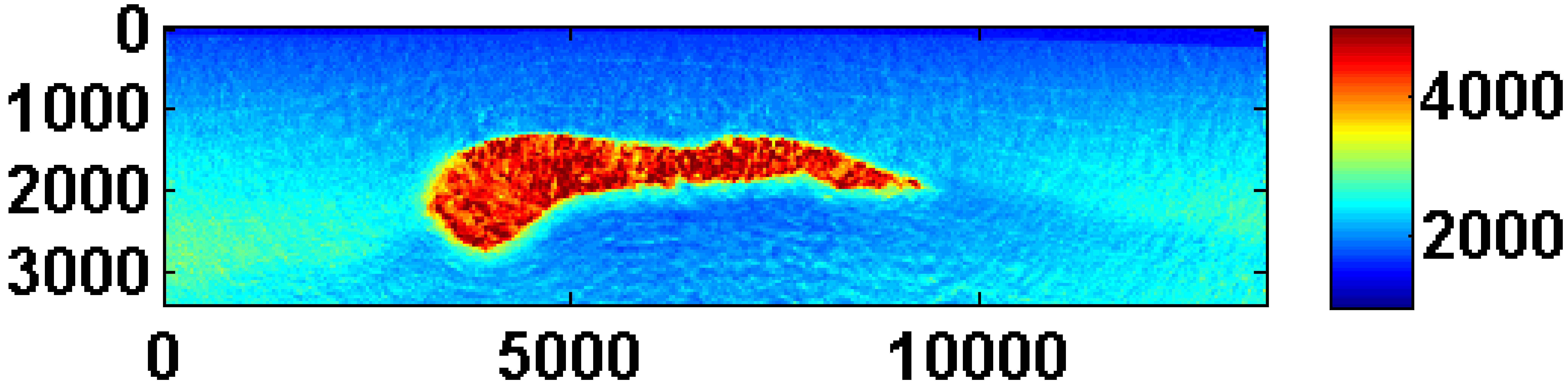


Initial velocity

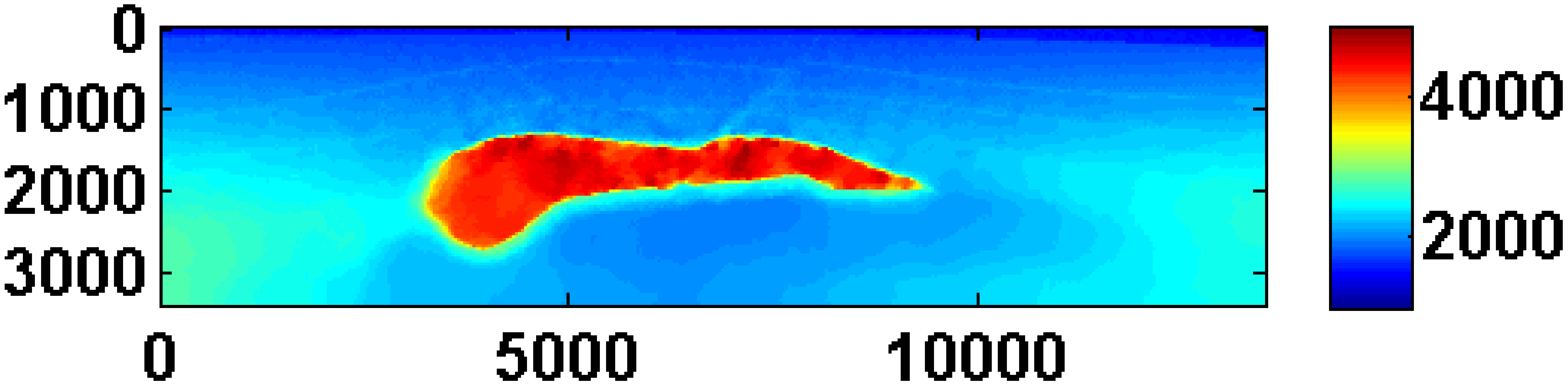
Modeling Details

- model size: 170 by 676
- mesh size: 20m
- number of sources: 116
- number of receivers: 673
- frequency range: 3-33Hz in overlapping batches of 2
- maximum number of outer iterations per frequency batch: 25
- maximum number of inner iterations for convex subproblems: 2000
- known Ricker wavelet sources with 30Hz peak frequency
- two simultaneous shots with Gaussian weights, without redraws
- no added noise

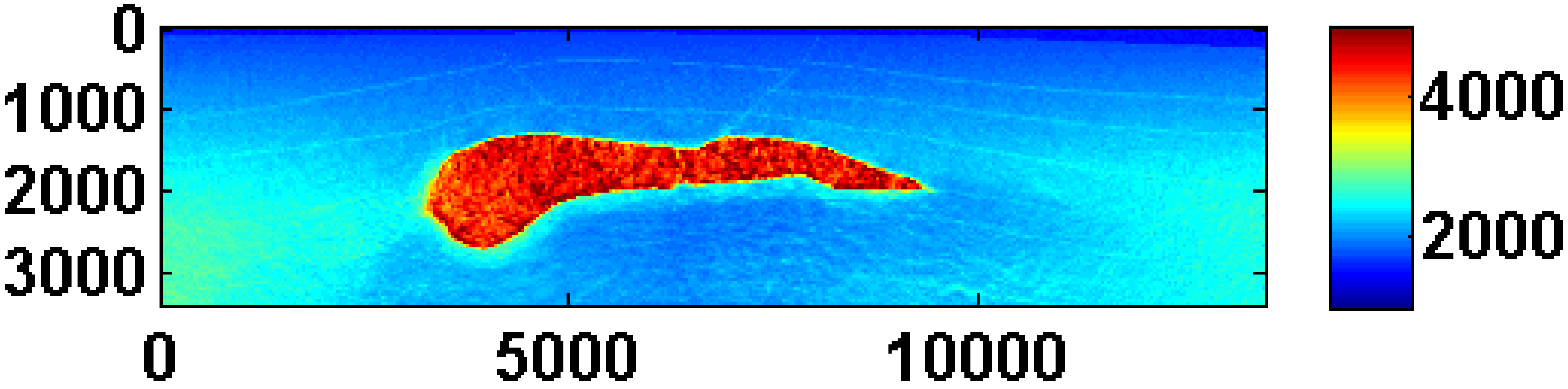
No TV Constraint, Smooth Initial Model



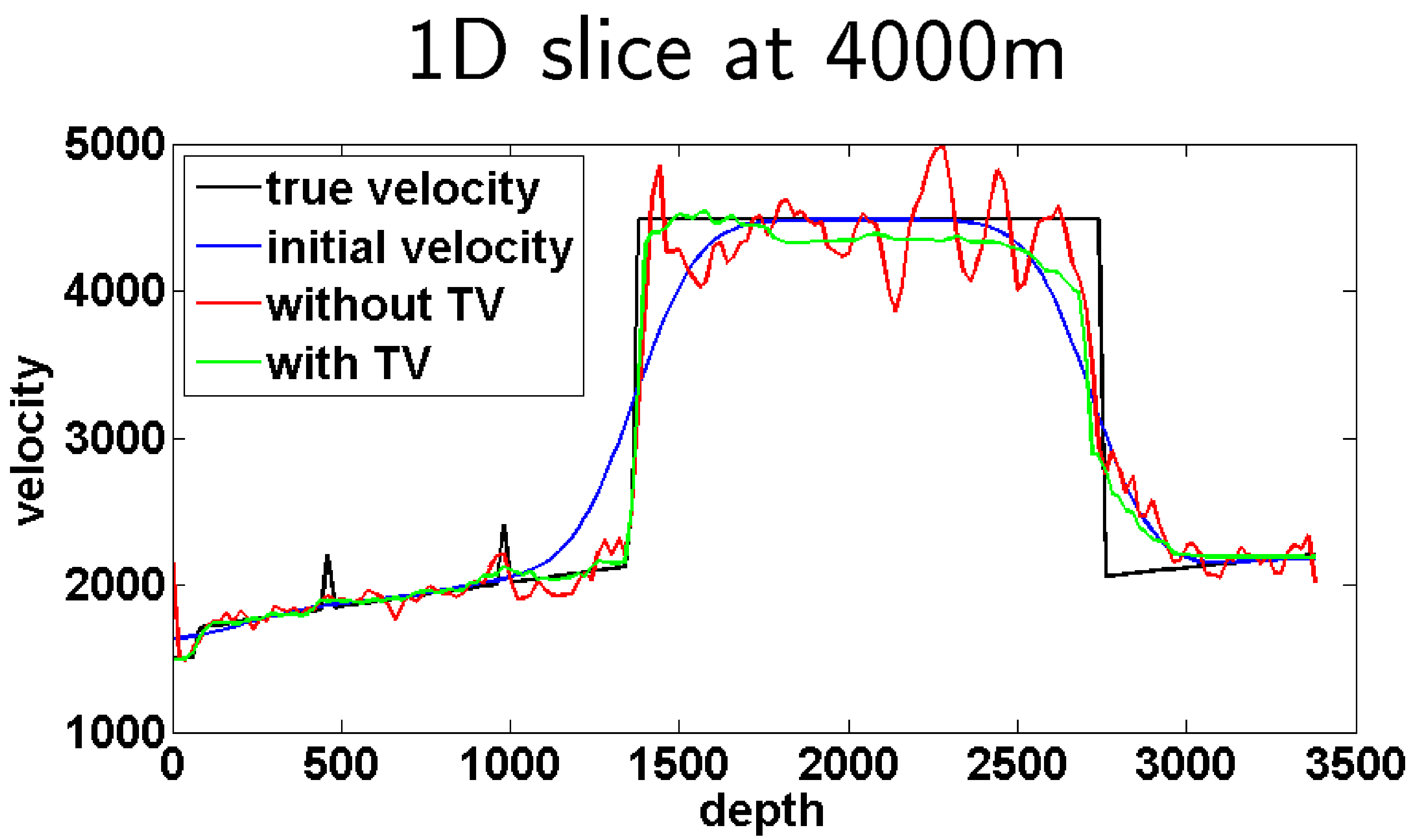
With TV Constraint, Smooth Initial Model

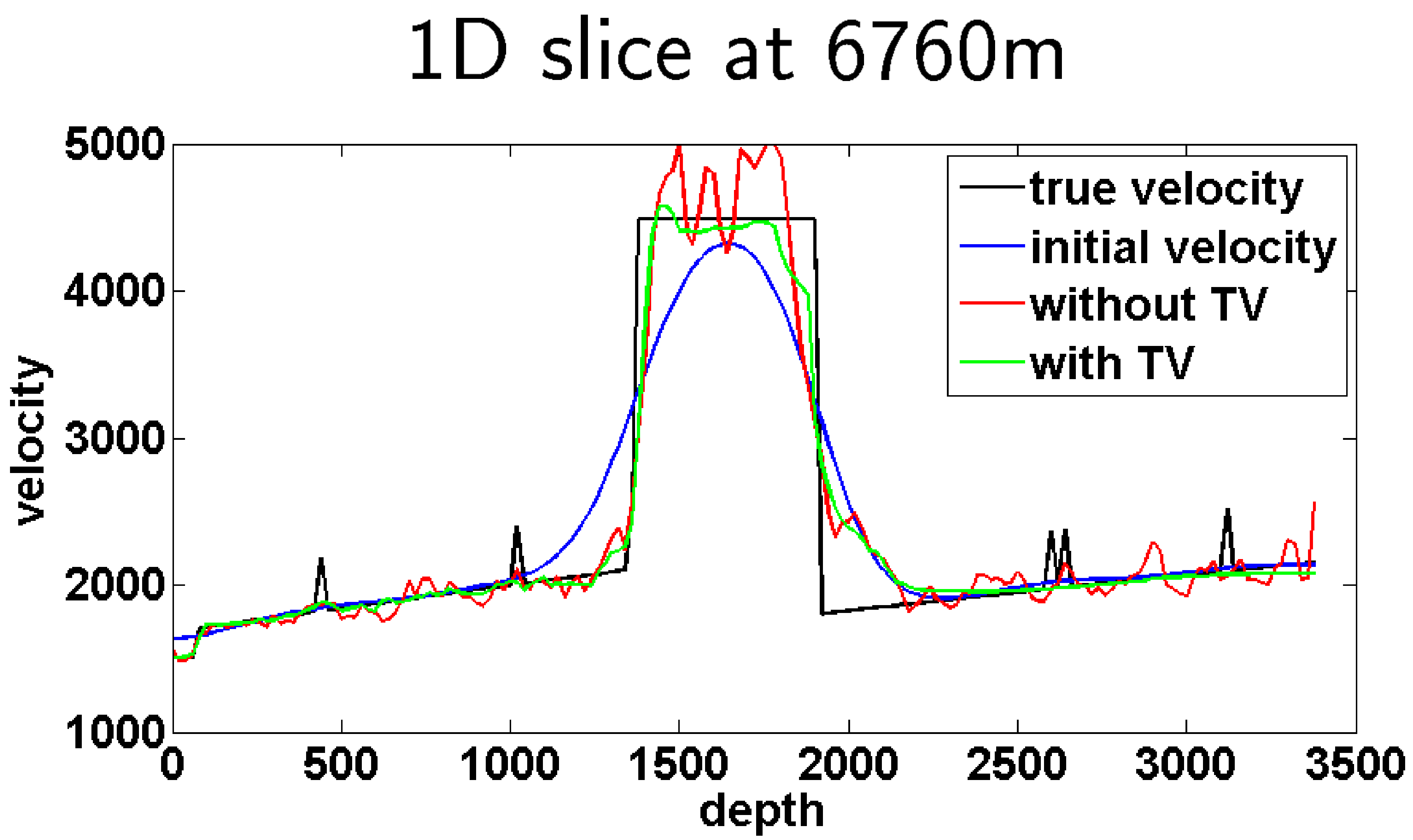


No TV Constraint, Using TV Result as Initial Model

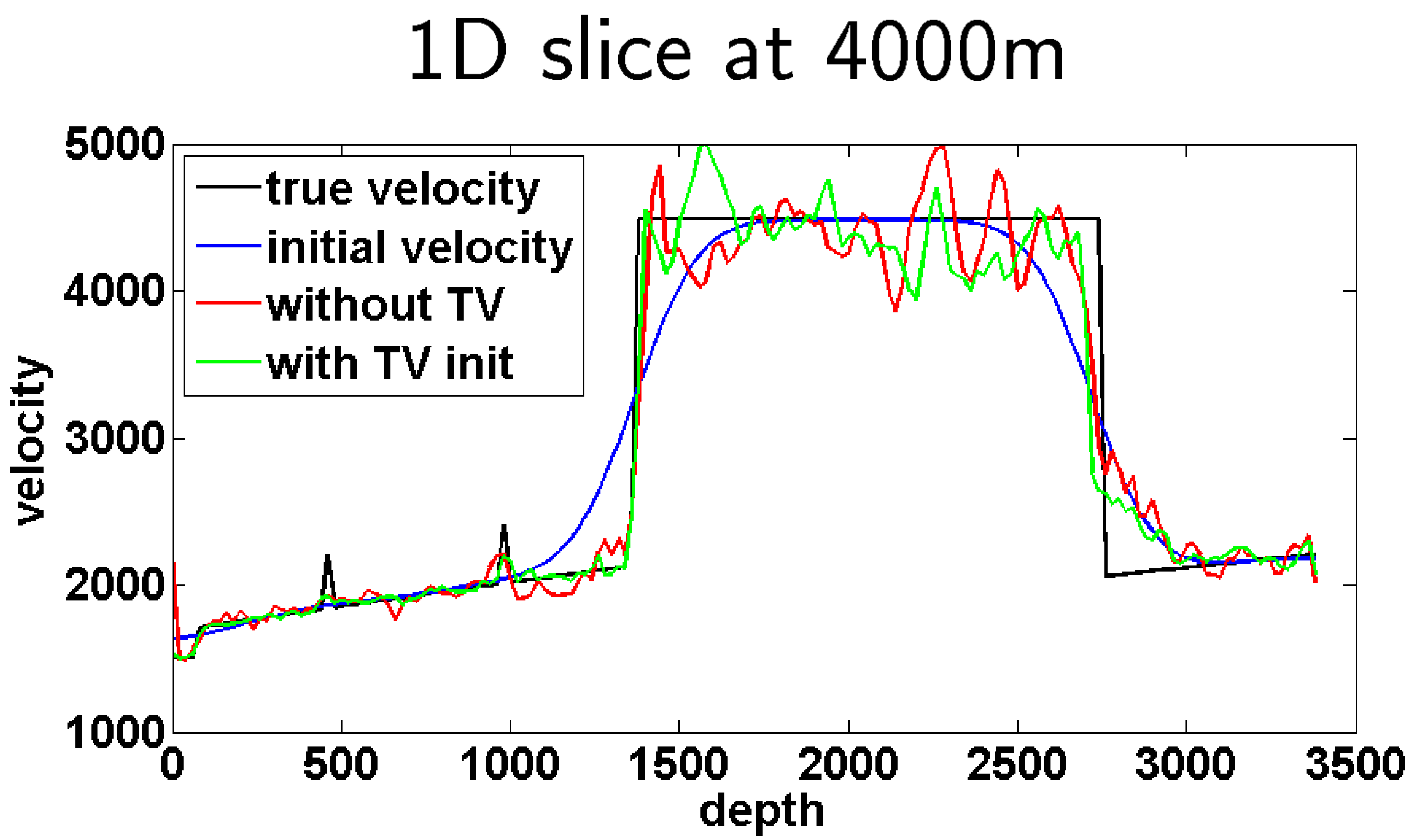


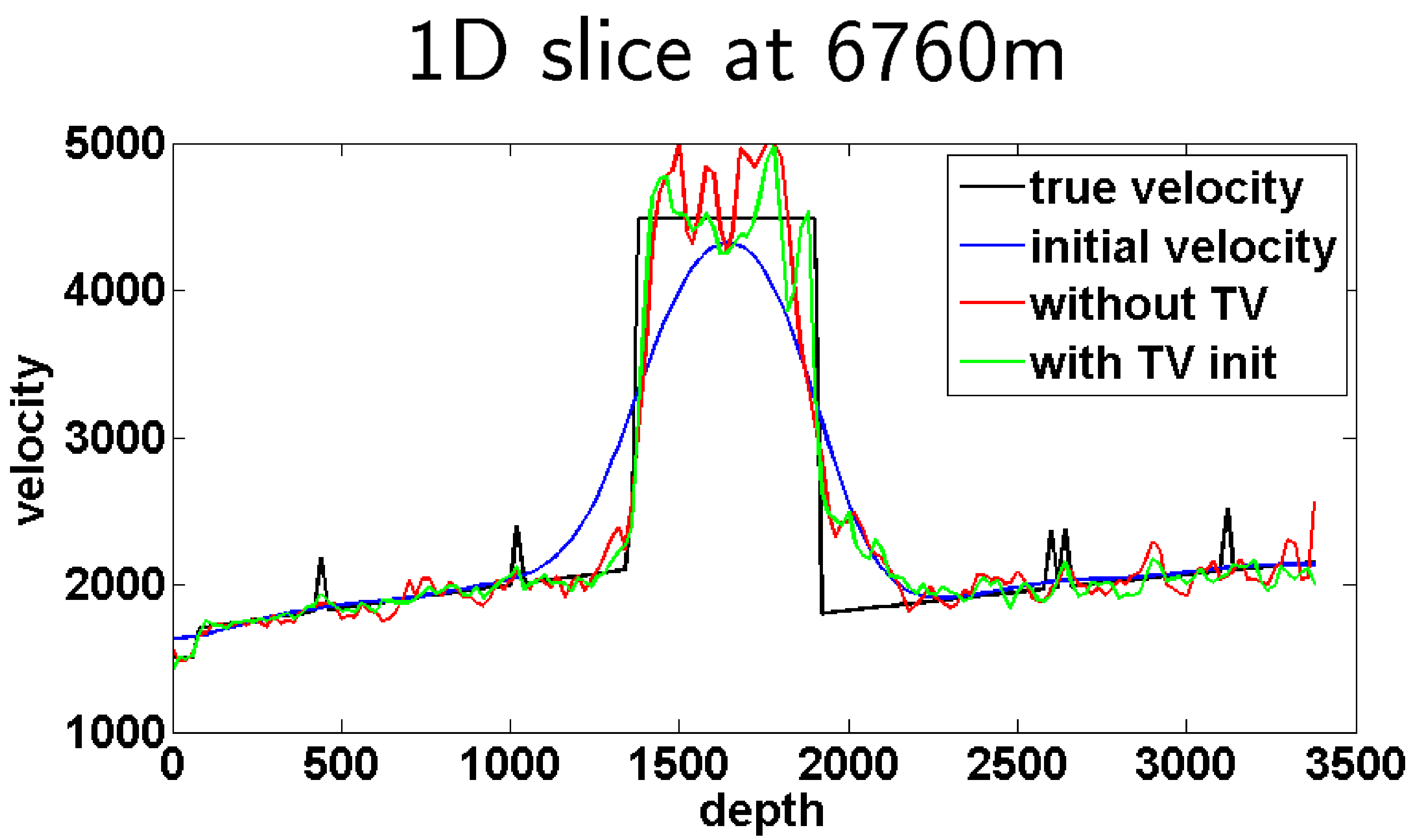
1D Slices from Salt Inversion using TV





1D Slices without TV but using TV Initialization





Conclusions and Future Work

- A scaled gradient projection framework can be used to add convex constraints to the Wavefield Reconstruction Inversion method for acoustic FWI.
- Bound constraints can be added at no additional cost.
- Total variation constraints lead to convex subproblems that are more expensive than simple Gauss Newton updates, but they don't require additional PDE solves.
- We intend to explore other constraints as well as continuation strategies along the lines of using TV constrained WRI as preprocessing to improve the starting model.