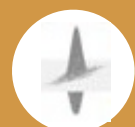


Automatic salt delineation: Wavefield Reconstruction Inversion with Convex Constraints

Ernie Esser, Lluís Guasch*, and Felix J. Herrmann



*Sub Salt Solutions Ltd

SLIM 

University of British Columbia



John “Ernie” Esser (May 19, 1980 – March 8, 2015)

The work presented in this abstract is a tribute to its main author, Ernie Esser, who passed away in tragic circumstances early this year. He was a young and very promising scientist with whom the co-authors had the privilege to collaborate.

Challenge

Salt has

- ▶ sharp edges
- ▶ strong-velocity contrasts w.r.t. sedimentary layers
- ▶ high velocities

Major challenges for (automatic) velocity building...

Strategy

Extend the search space

- ▶ “less” nonlinear
- ▶ ensures data fit & avoids cycle skips

“Squeeze” the extension by

- ▶ enforcing the wave equation to compute model updates
- ▶ imposing *asymmetric* constraints that encode “rudimentary” properties of the geology
- ▶ relaxing the constraints to allow data fits & details to enter the solution

Leverage frequency continuation & warm starts where

- ▶ *sparsity-promoting asymmetric* constraints limit adverse affects of local minima
- ▶ there is hope as long as progress is made towards the solution during each cycle

Outcome: an automatic multi-cycle optimization-driven workflow

Wavefield Reconstruction Inversion – WRI

wave-equation $\times$ wavefield $=$ source

versus

$$\left(\begin{array}{c} \text{wave-equation} \\ \text{-----} \\ \text{sampling operator} \end{array} \right) \times \text{wavefield} = \left(\begin{array}{c} \text{source} \\ \text{-----} \\ \text{data} \end{array} \right)$$

WRI – iterations

WRI method

for each source i

$$\text{solve } \begin{pmatrix} P_i \\ \lambda A_i(\mathbf{m}) \end{pmatrix} \mathbf{u}_{\lambda,i} \approx \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix}$$

$$\mathbf{g} = \mathbf{g} + \lambda^2 \omega^2 \text{diag}(\bar{\mathbf{u}}_{i,\lambda})^* (A(\mathbf{m}) \bar{\mathbf{u}}_{i,\lambda} - \mathbf{q}_i)$$

end

$$\mathbf{m} = \mathbf{m} - \alpha \mathbf{g}$$

correlation proxy
wavefield & PDE
residual

Conventional method

for each source i

$$\text{solve } A(\mathbf{m}) \mathbf{u}_i = \mathbf{q}_i$$

$$\text{solve } A(\mathbf{m})^* \mathbf{v}_i = P_i^* (P_i \mathbf{u}_i - \mathbf{d}_i)$$

$$\mathbf{g} = \mathbf{g} + \omega^2 \text{diag}(\mathbf{u}_i)^* \mathbf{v}_i$$

end

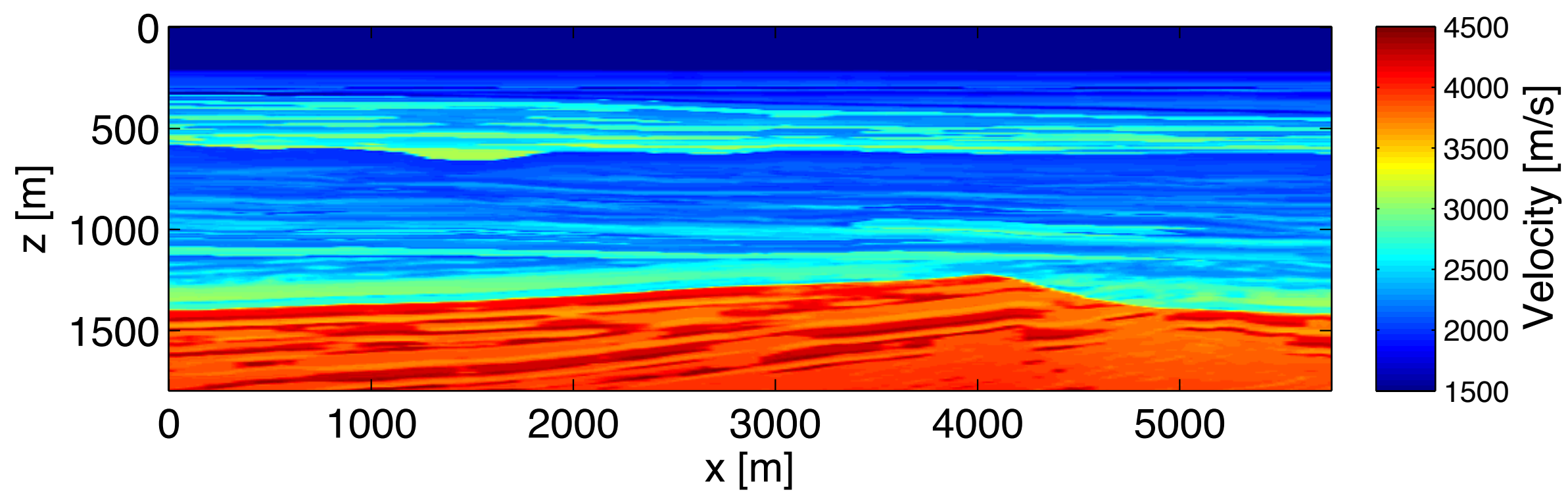
$$\mathbf{m} = \mathbf{m} - \alpha \mathbf{g}$$

correlation
wavefield &
data residual

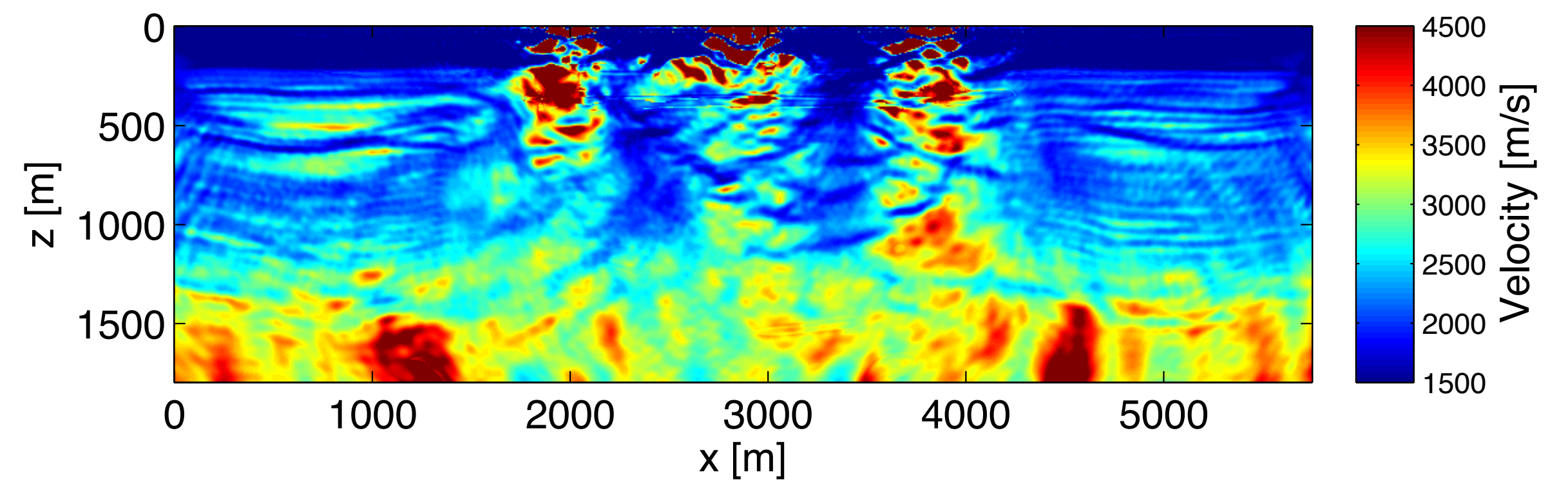
WRI

poor starting model

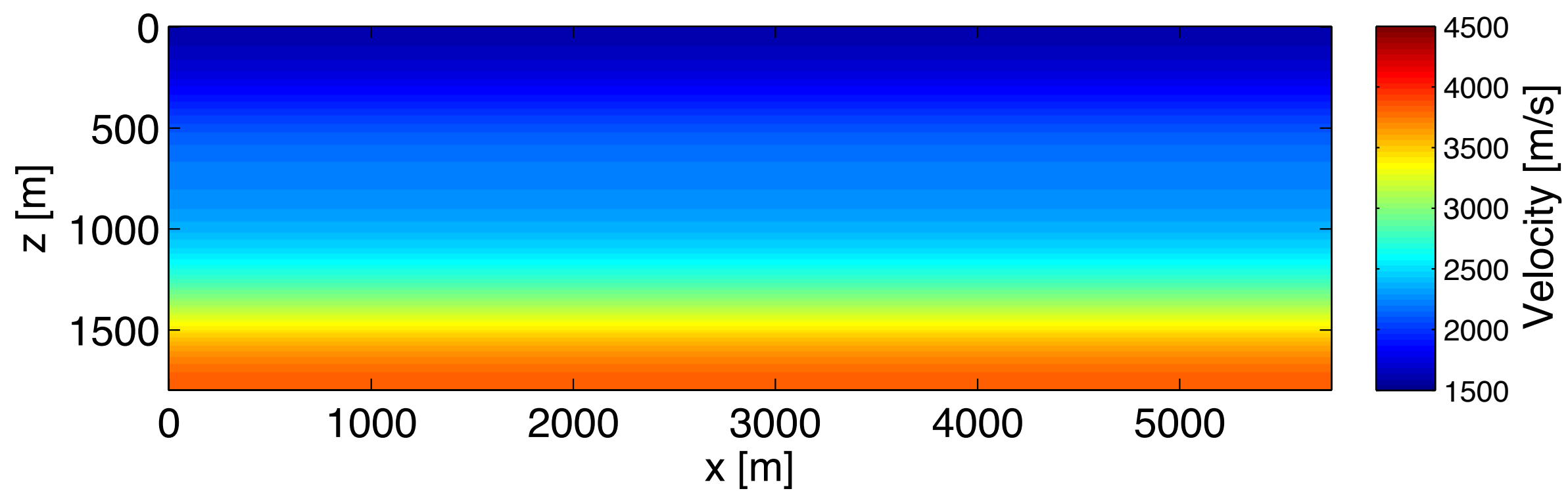
True model



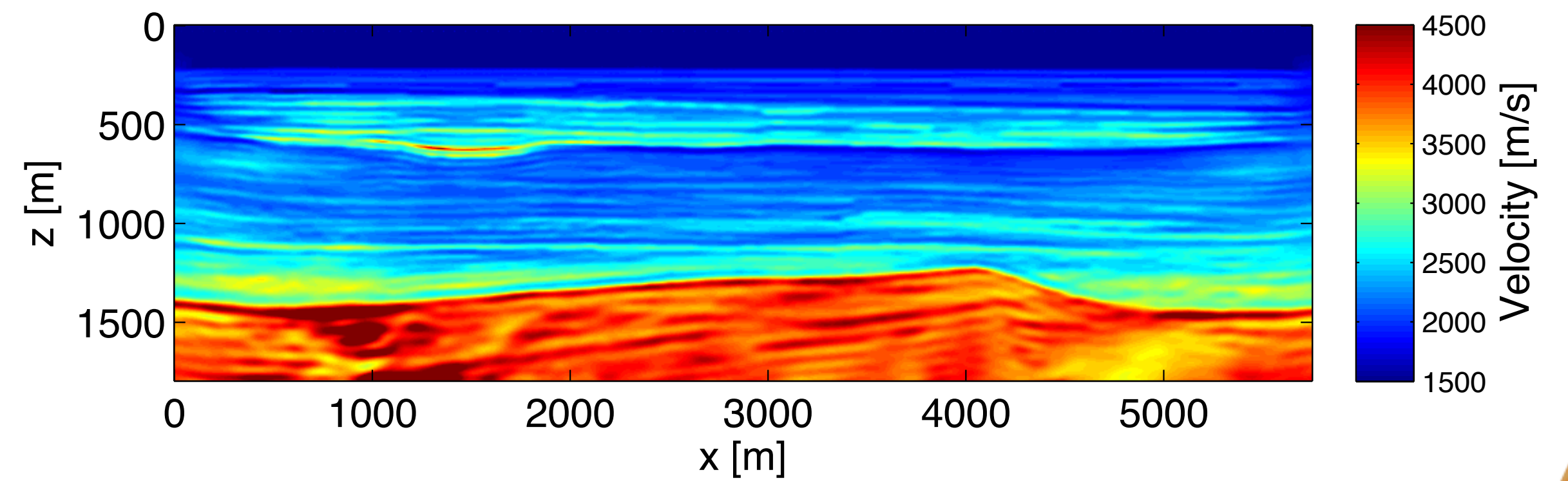
Result FWI



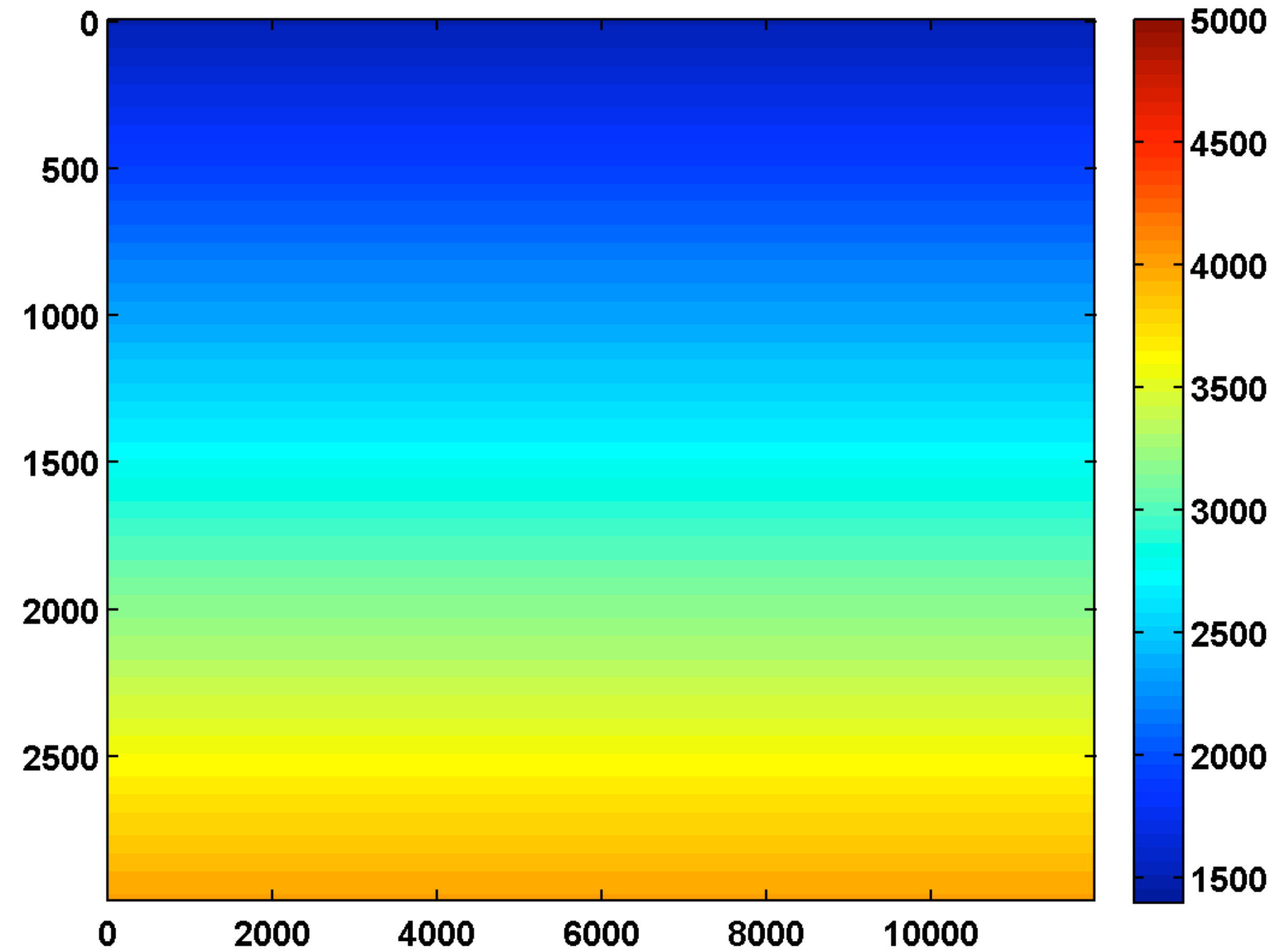
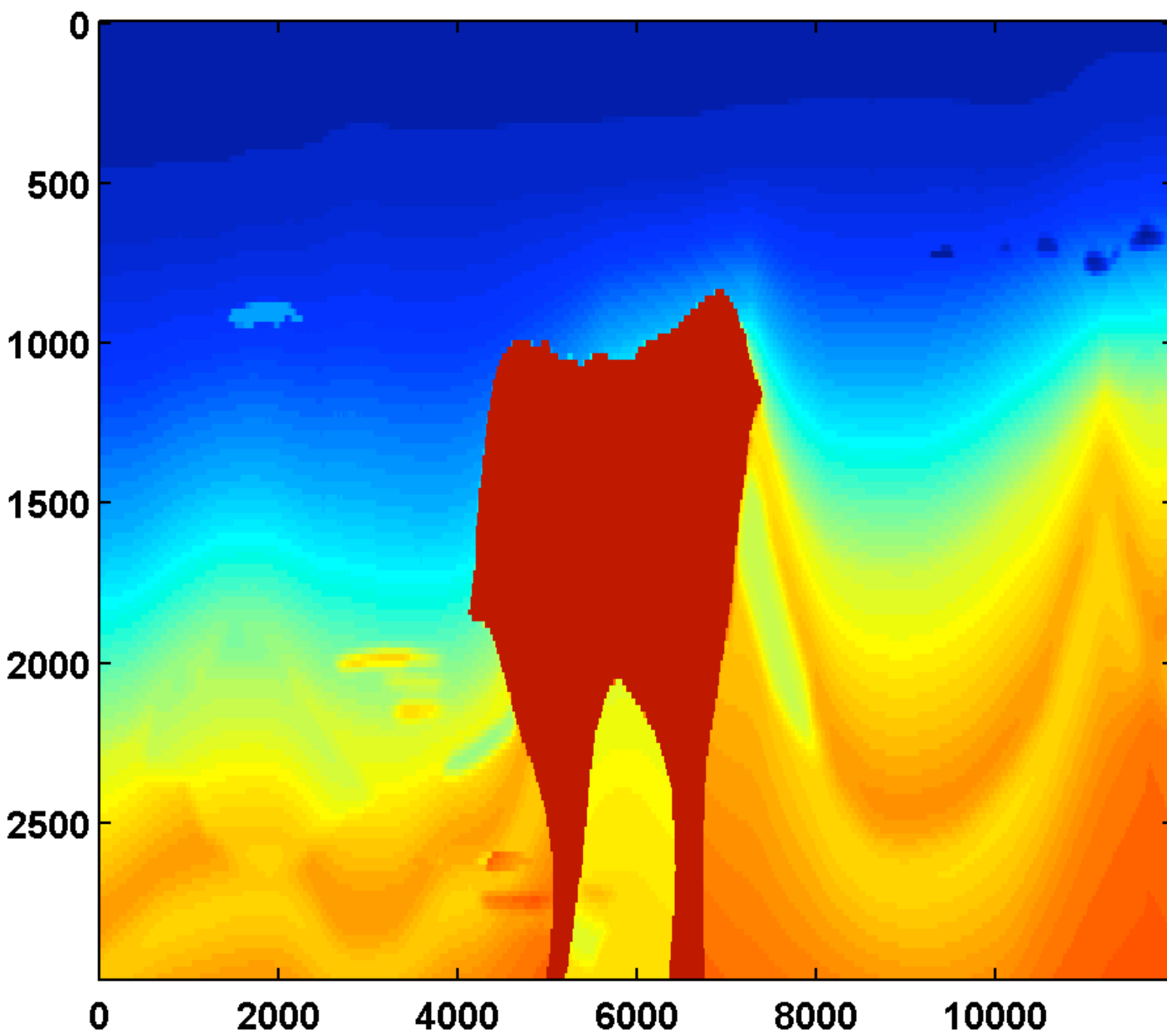
Initial model



Result WRI, $\lambda = 1$

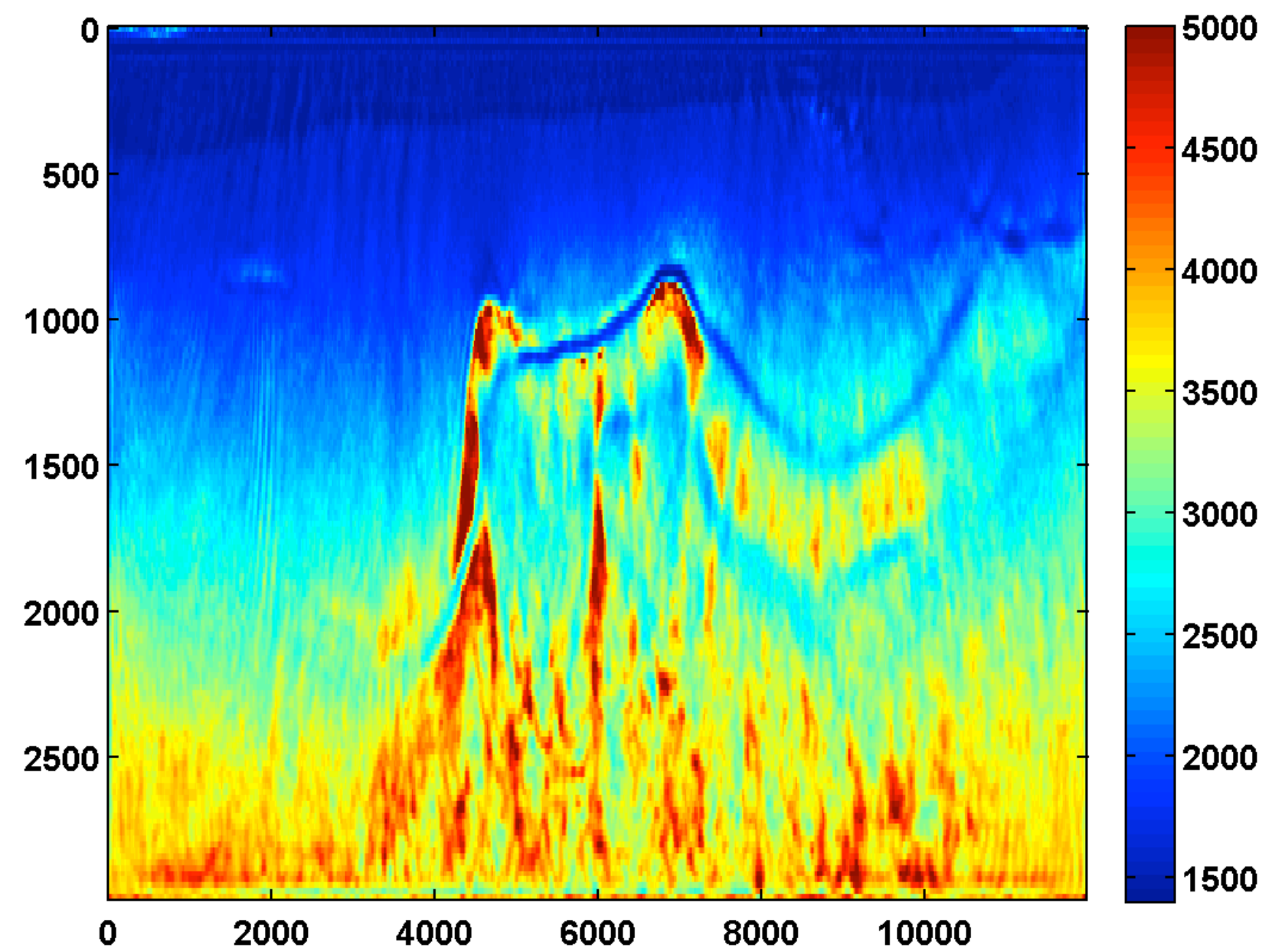


FWI – poor starting model

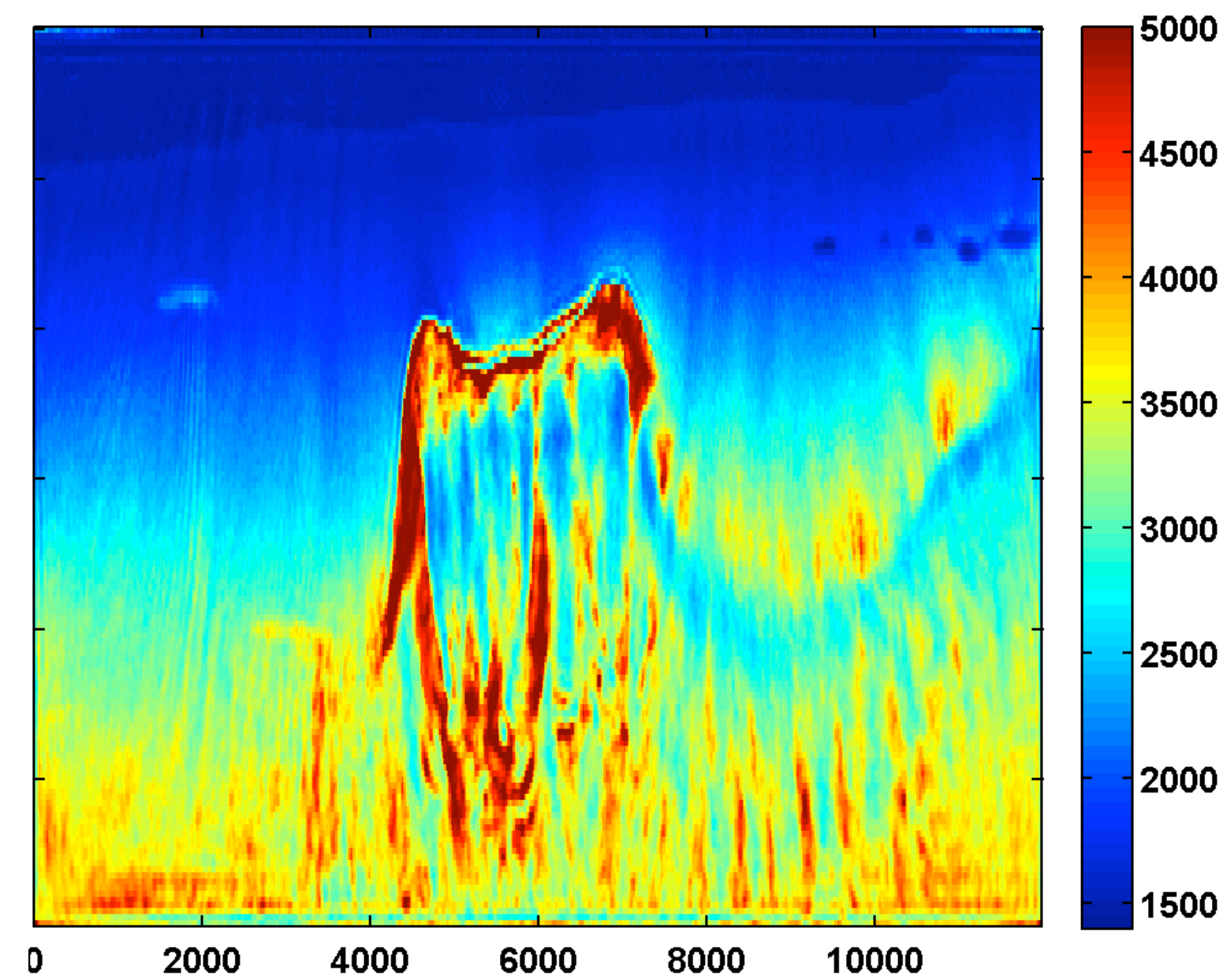


WRI results – w/ source encoding

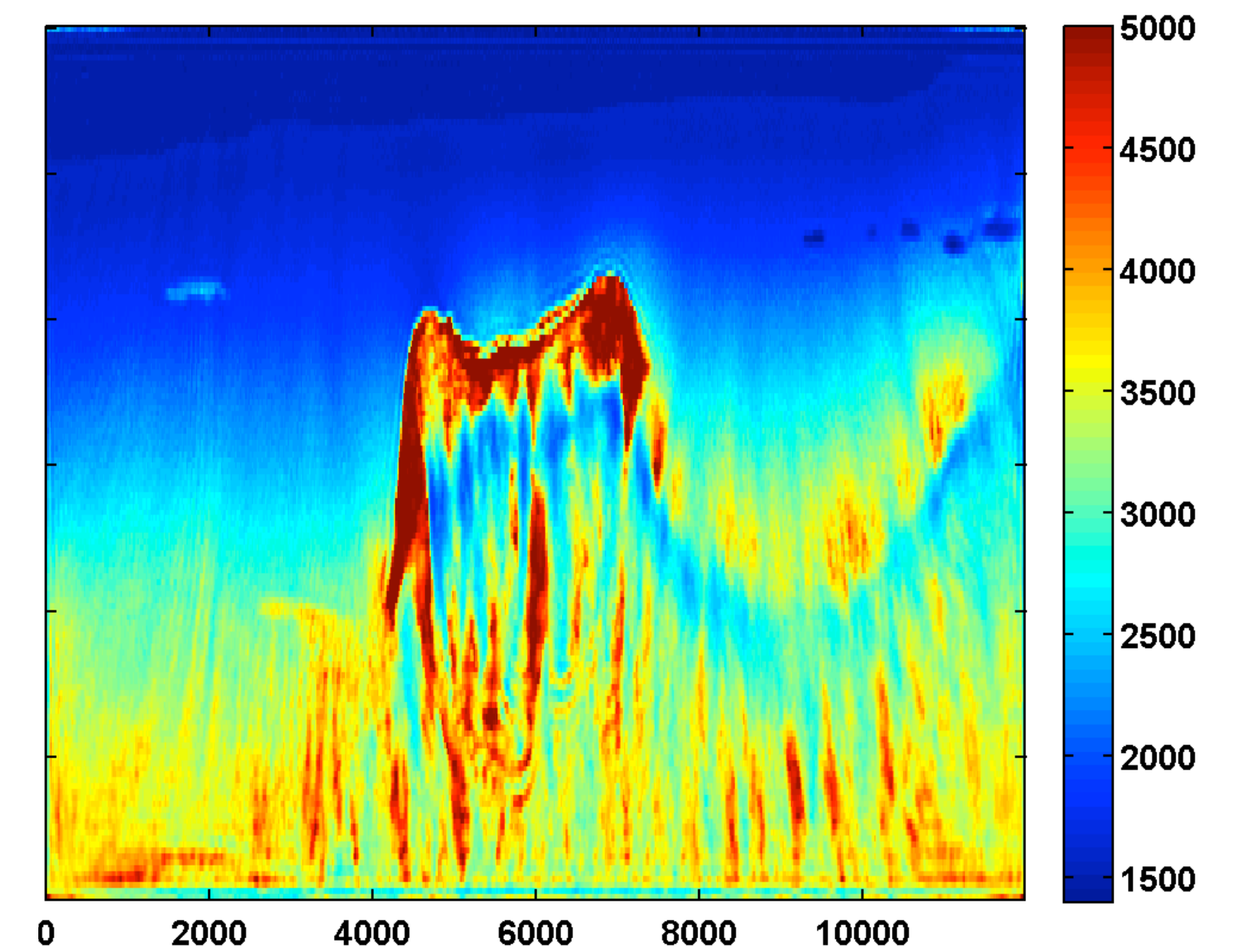
after one cycle through the frequencies



after two cycles through the frequencies



after three cycles through the frequencies



Including constraints

Aim: constrain inverted models to a certain set of models, e.g. via

- ▶ box constraints – ensure point-wise values to be w/i specified interval
- ▶ TV-norm constraints – promote sharp edges & remove unwanted “noise”

Challenge: optimization of large-scale problems

- ▶ intersections of convex sets
- ▶ computational costs

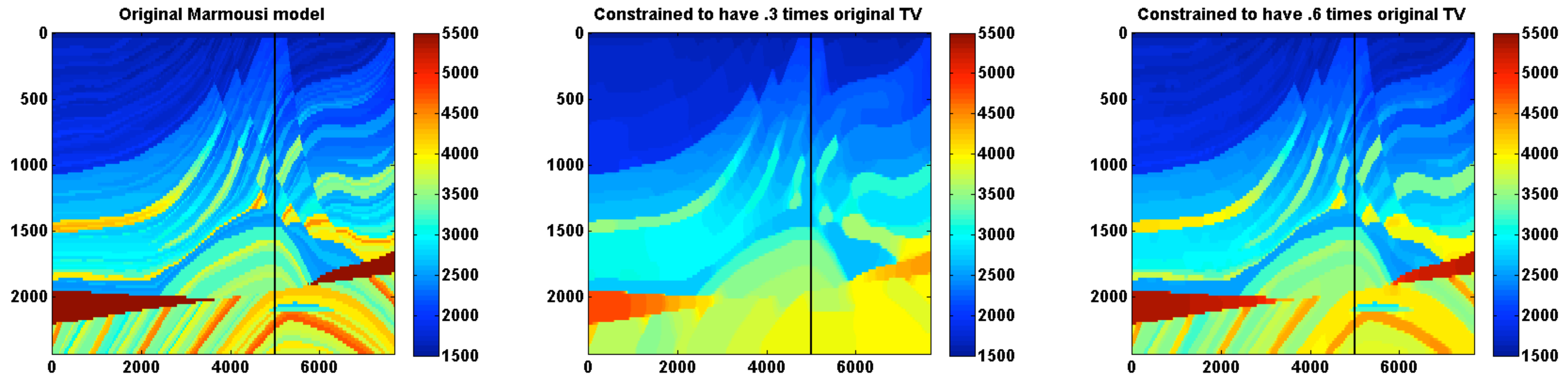
Solution: leverage developments in optimization w/ “simple” constraints

- ▶ saddle-point formulation – primal-dual hybrid gradients
- ▶ leverage WRI’s cheap GN Hessians in inner-/outer-loop structure

Edge-preserving Total-Variation projections

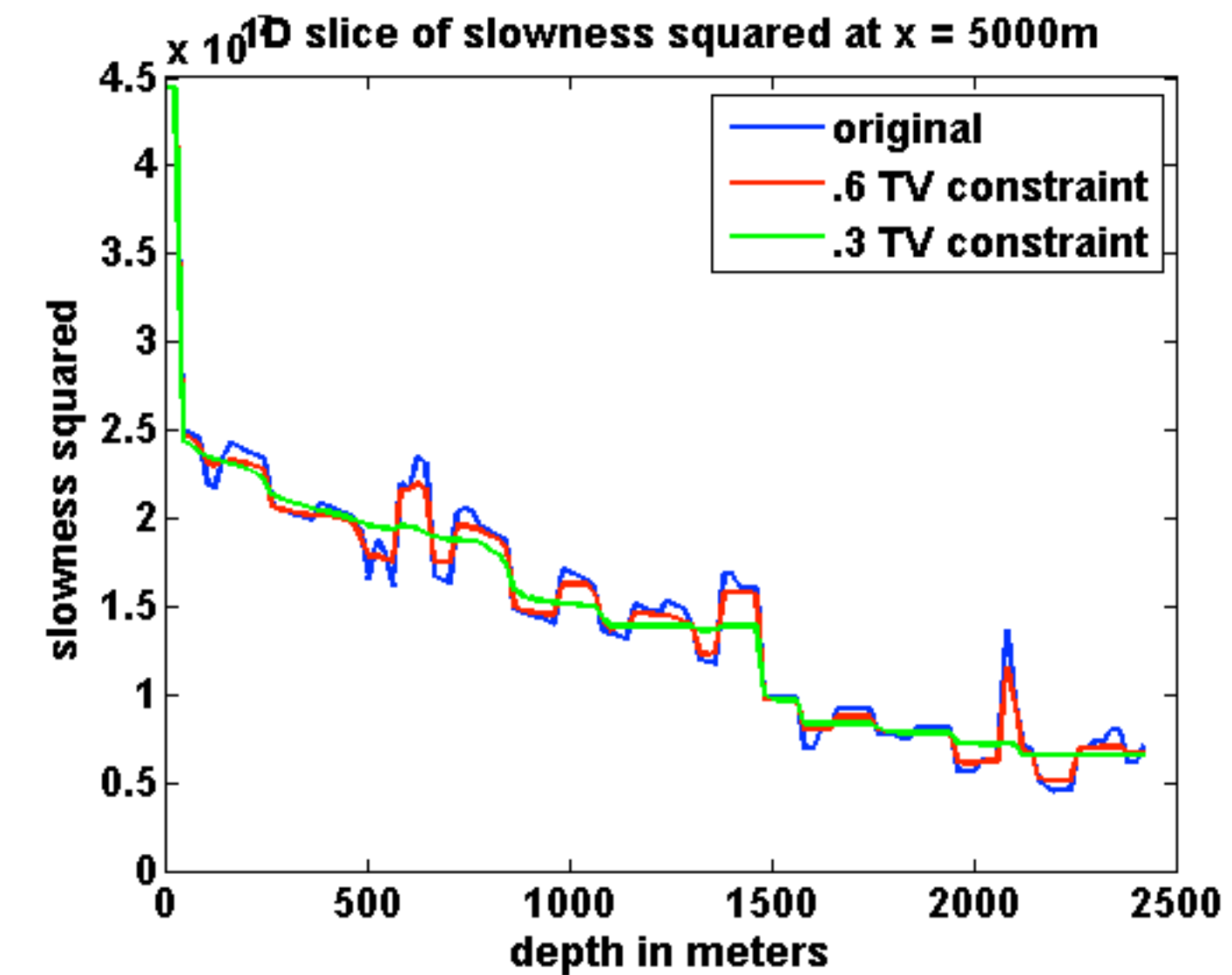
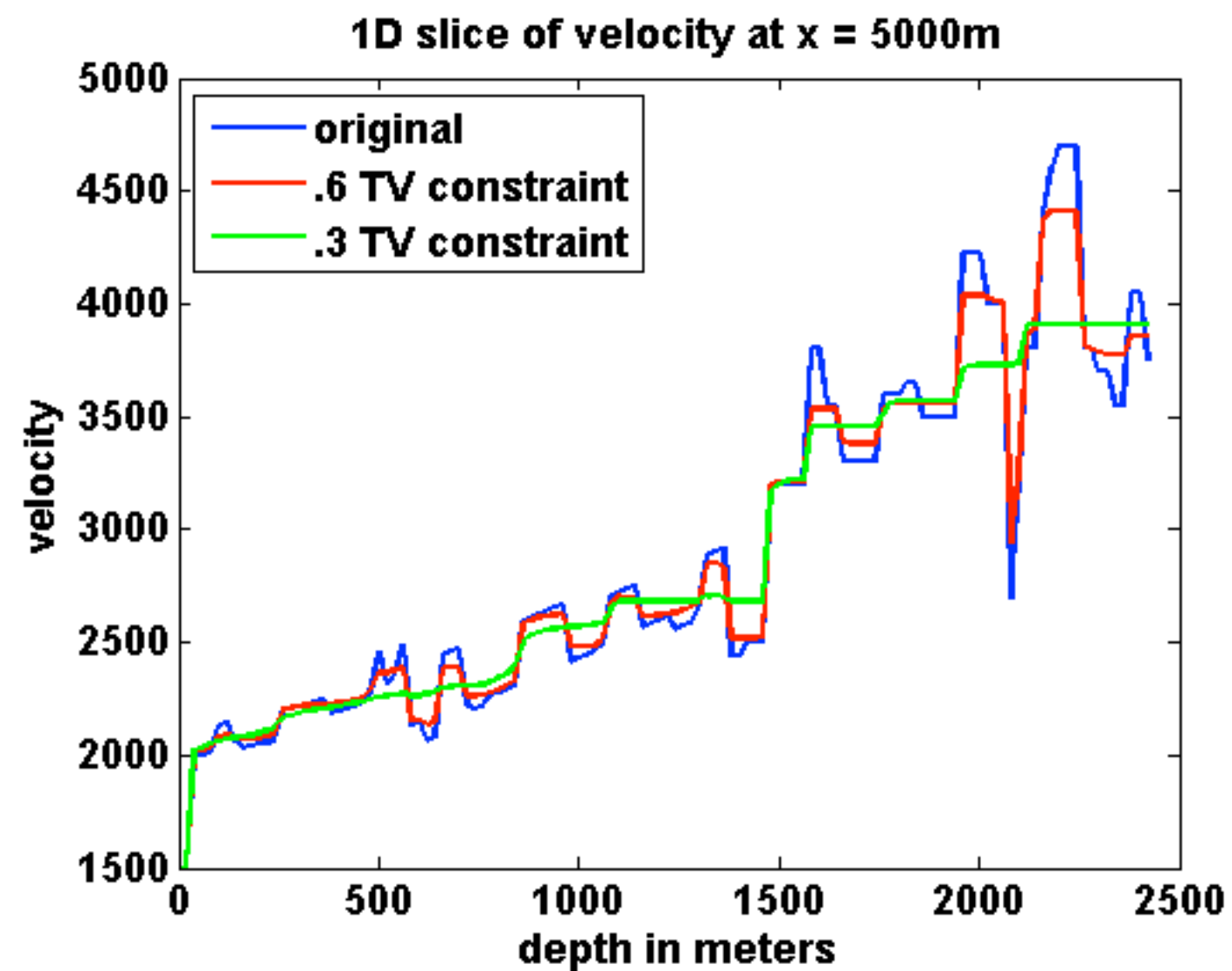
$v_{\min} = 1500$, $v_{\max} = 5500$, and $\tau = \{0.3\tau_0, 0.6\tau_0\}$

$$\Pi_C(\mathbf{m}_0) = \arg \min_{\mathbf{m}} \frac{1}{2} \|\mathbf{m} - \mathbf{m}_0\|^2 \quad \text{subject to} \quad \mathbf{m}_i \in [B_i^l, B_i^u] \quad \text{and} \quad \|\mathbf{m}\|_{TV} \leq \tau$$



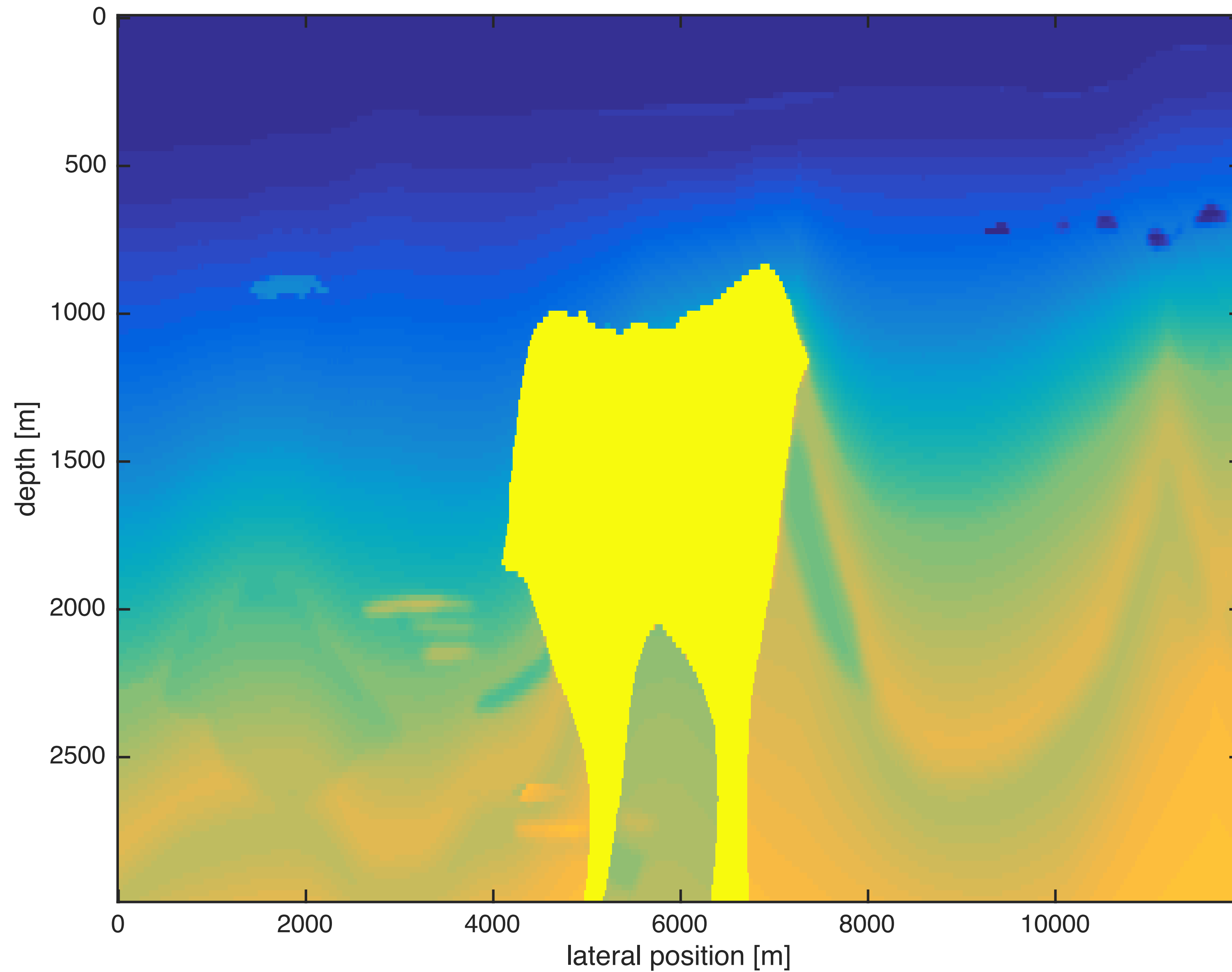
Edge-preserving Total-Variation projections

$v_{\min} = 1500$, $v_{\max} = 5500$, and $\tau = \{0.3\tau_0, 0.6\tau_0\}$

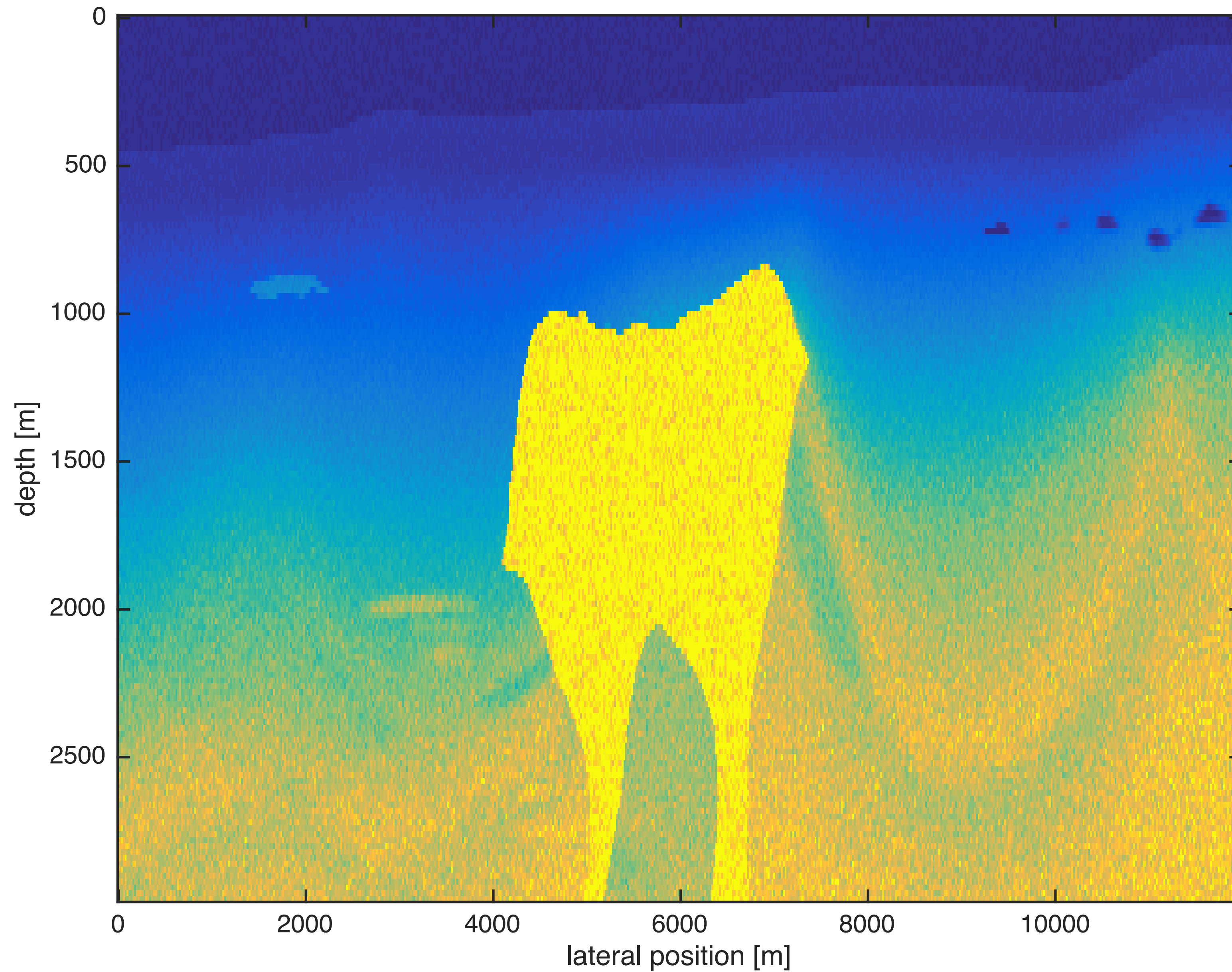


- ▶ reduces complexity of the model
- ▶ preserves blockiness of reflectors
- ▶ relaxing the constraint allows for more layers

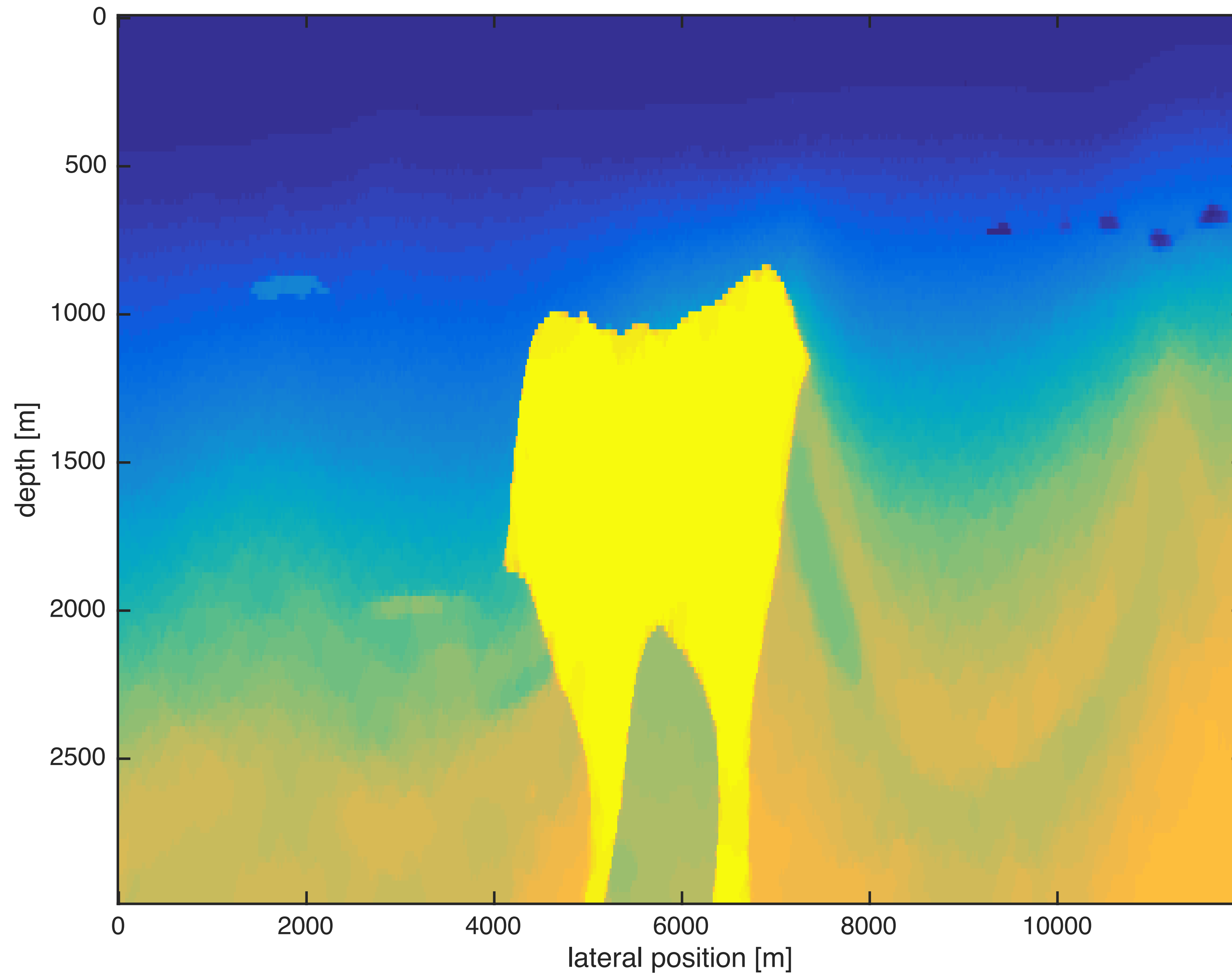
Subset BP 2004 Salt model – original



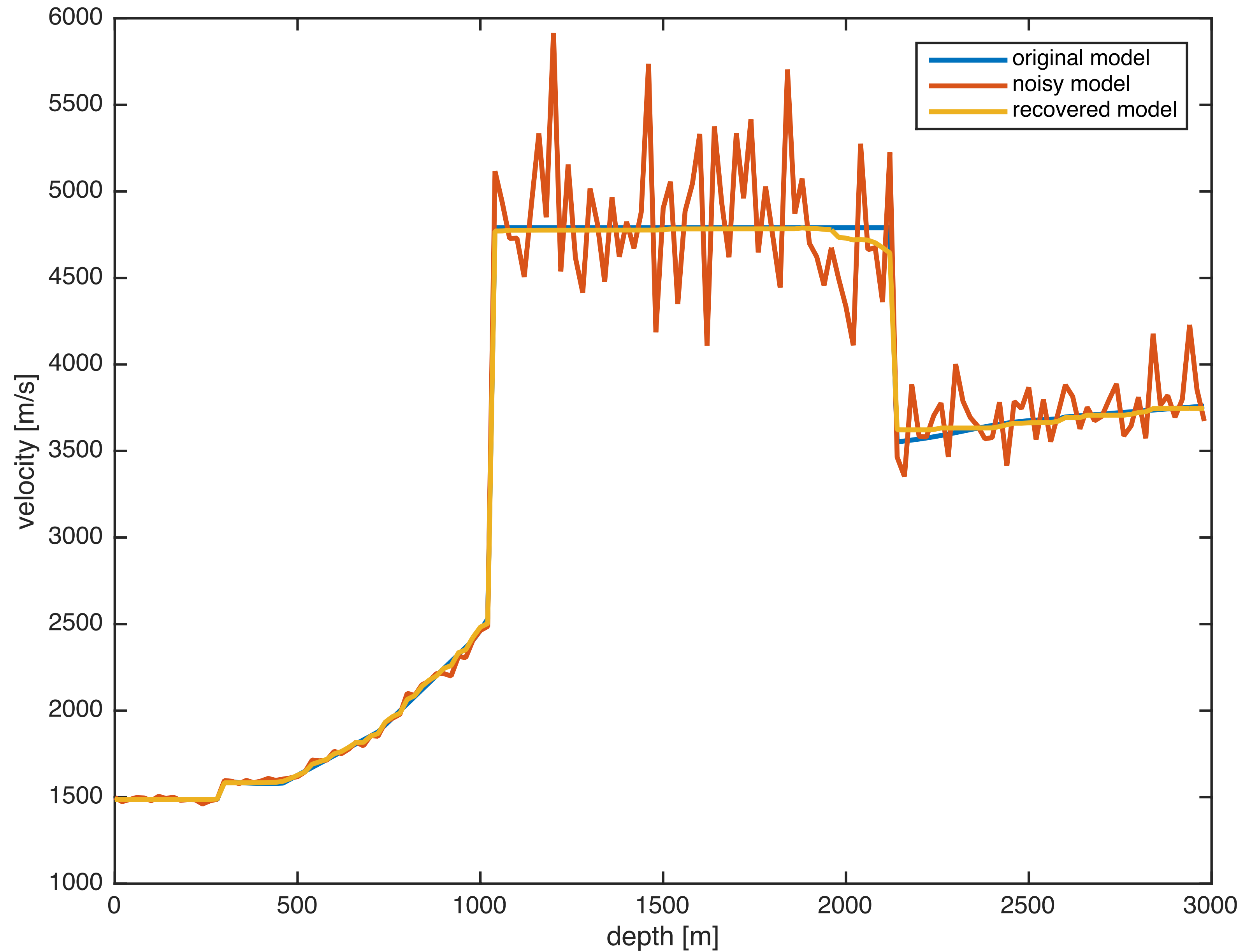
Subset BP 2004 Salt model – noisy



Subset BP 2004 Salt model – TV-constraint imposed



Subset BP 2004 Salt model – TV-constraint imposed



WRI – diagonal Gauss-Newton Hessian

WRI method

for each source i

$$\text{solve } \begin{pmatrix} P_i \\ \lambda A_i(\mathbf{m}) \end{pmatrix} \mathbf{u}_{\lambda,i} \approx \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix}$$

$$\mathbf{g} = \mathbf{g} + \lambda^2 \omega^2 \text{diag}(\bar{\mathbf{u}}_{i,\lambda})^* (A(\mathbf{m}) \bar{\mathbf{u}}_{i,\lambda} - \mathbf{q}_i)$$

$$H_{GN} = H_{GN} + \lambda^2 \omega^4 \text{diag}(\bar{\mathbf{u}}_{i,\lambda})^* \text{diag}(\bar{\mathbf{u}}_{i,\lambda})$$

end

$$\mathbf{m} = \mathbf{m} - \alpha H_{GN}^{-1} \mathbf{g}$$

replace by inner
loop that imposes
convex constraints

diagonal Hessian
=
pseudo Hessian

Conventional method

for each source i

$$\text{solve } A(\mathbf{m}) \mathbf{u}_i = \mathbf{q}_i$$

$$\text{solve } A(\mathbf{m})^* \mathbf{v}_i = P_i^* (P_i \mathbf{u}_i - \mathbf{d}_i)$$

$$\mathbf{g} = \mathbf{g} + \omega^2 \text{diag}(\mathbf{u}_i)^* \mathbf{v}_i$$

end

$$\mathbf{m} = \mathbf{m} - \alpha \mathbf{g}$$

dense Hessian
&
too expensive

WRI – scaled gradient descent

Minimize the reduced objective for given $\bar{\mathbf{u}}_{sv}$'s

$$\Phi(\mathbf{m}) = \sum_{sv} \frac{1}{2} \|P\bar{\mathbf{u}}_{sv}(\mathbf{m}) - \mathbf{d}_{sv}\|^2 + \frac{\lambda^2}{2} \|A_v(\mathbf{m})\bar{\mathbf{u}}_{sv}(\mathbf{m}) - \mathbf{q}_{sv}\|^2$$

via scaled gradient descents at iteration n [Bertsekas '99]

$$\Delta \mathbf{m} = \arg \min_{\Delta \mathbf{m} \in \mathbb{R}^N} \Delta \mathbf{m}^T \mathbf{g}^n + \frac{1}{2} \Delta \mathbf{m}^T \overset{\text{cheap}}{H_{GN}^n} \Delta \mathbf{m} + c_n \Delta \mathbf{m}^T \Delta \mathbf{m}$$

expensive but fixed
cheap
damped

$$\mathbf{m}^{n+1} = \mathbf{m}^n + \Delta \mathbf{m} \text{ with } c_n \geq 0$$

Project onto convex sets

Solve

$$\underset{\mathbf{m}}{\text{minimize}} \Phi(\mathbf{m}) \quad \text{subject to} \quad \mathbf{m}^{n+1} \in C_{\text{box}} \cap C_{\text{TV}}$$

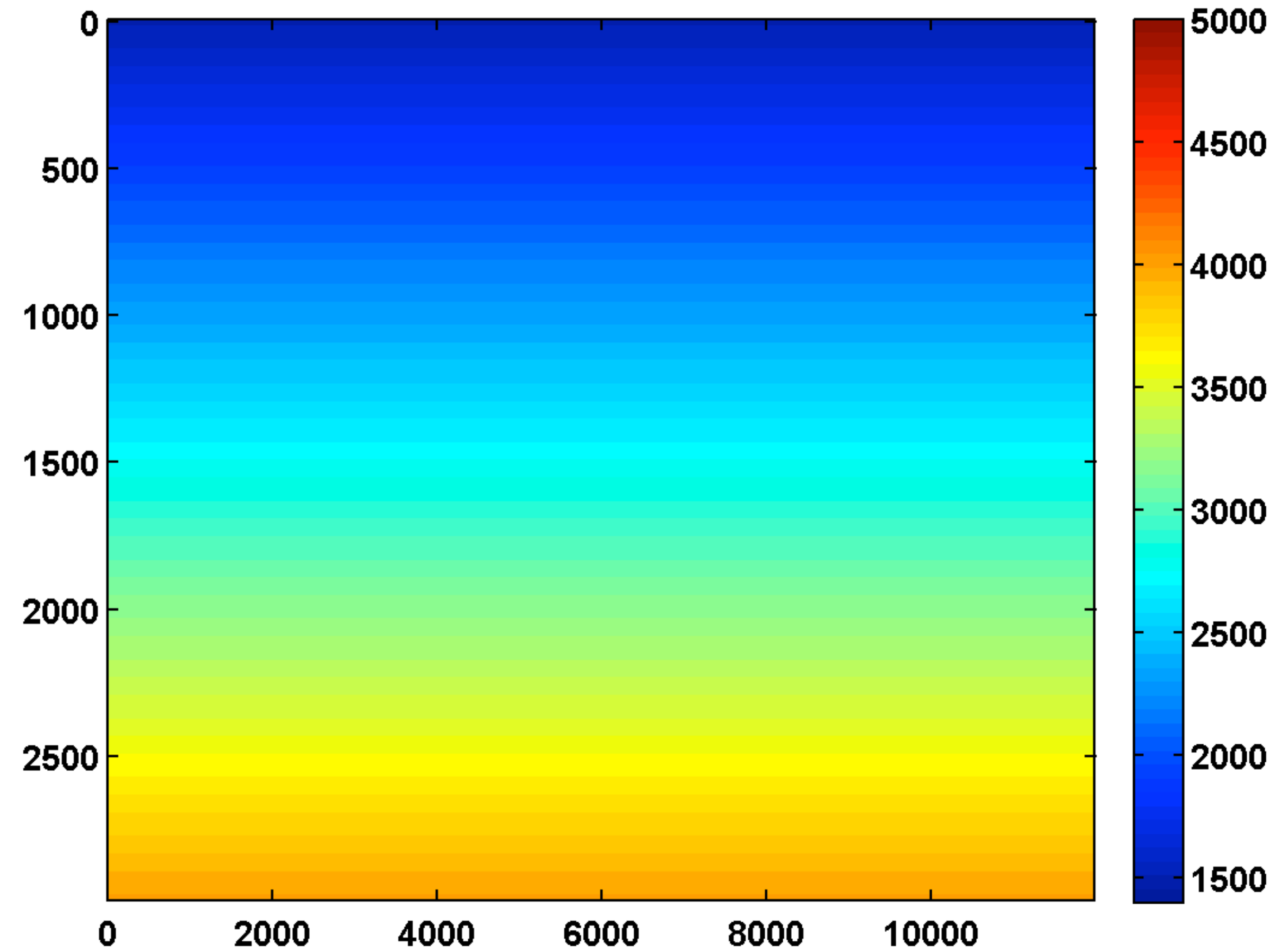
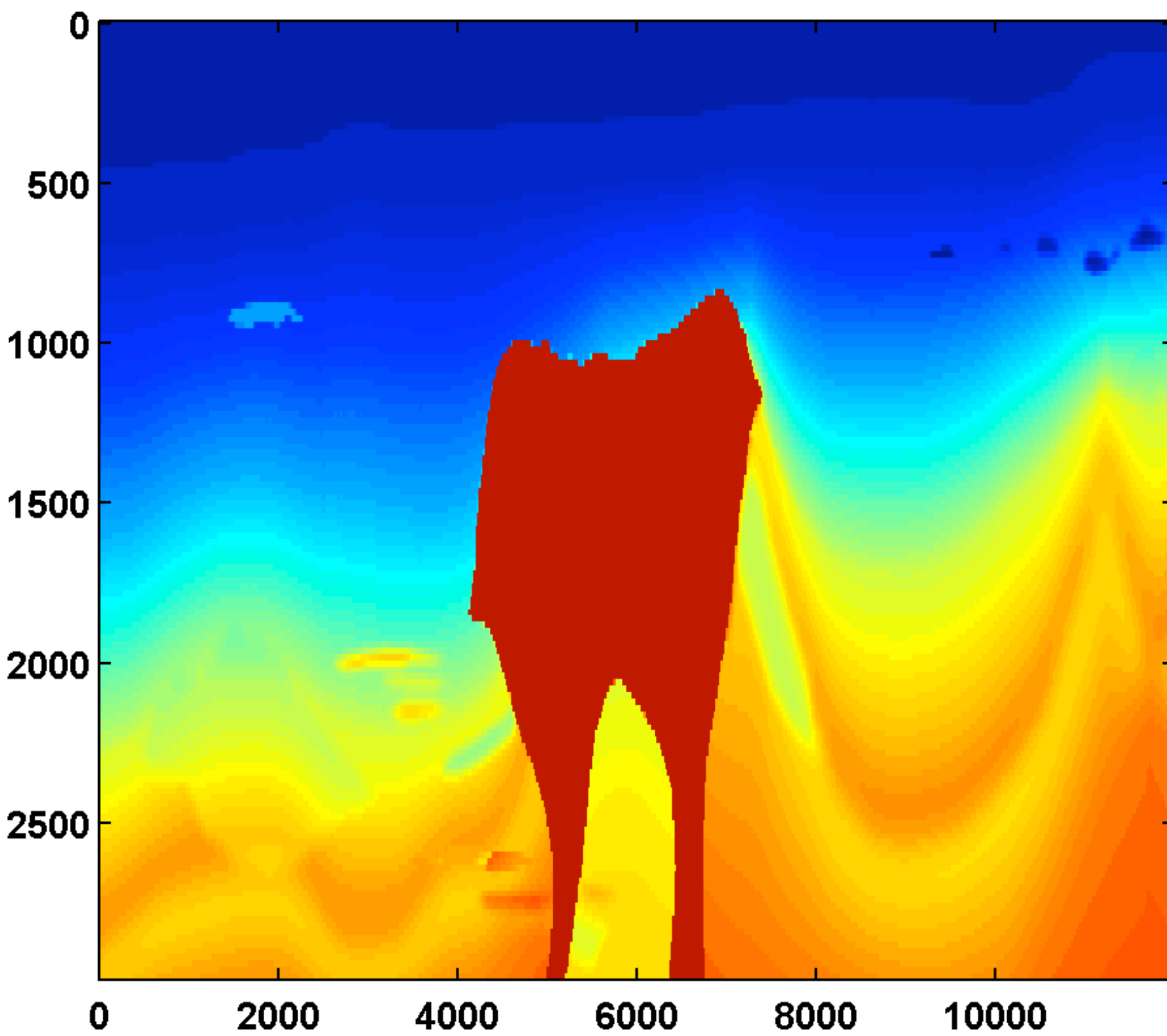
by iterating

$$\Delta \mathbf{m} = \arg \min_{\Delta \mathbf{m}} \Delta \mathbf{m}^T \mathbf{g}^n + \frac{1}{2} \Delta \mathbf{m}^T \mathbf{H}_{GN}^n \Delta \mathbf{m} + c_n \Delta \mathbf{m}^T \Delta \mathbf{m}$$

$$\text{subject to} \quad \mathbf{m}_i^n + \Delta \mathbf{m}_i \in [B_i^l, B_i^u] \quad \text{and} \quad \|\mathbf{m}^n + \Delta \mathbf{m}\|_{\text{TV}} \leq \tau$$

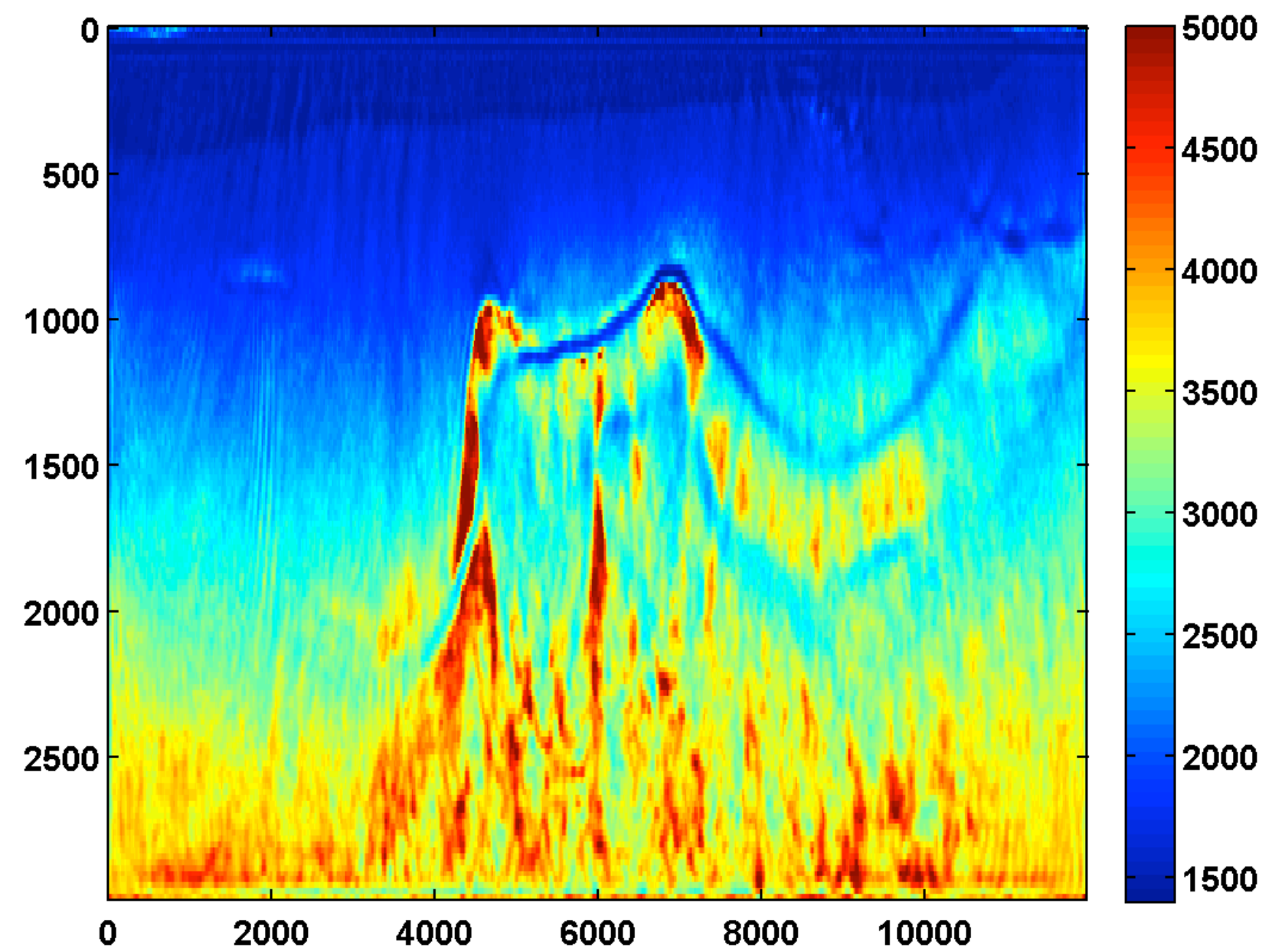
$$\mathbf{m}^{n+1} = \mathbf{m}^n + \Delta \mathbf{m}$$

Salt – poor starting model

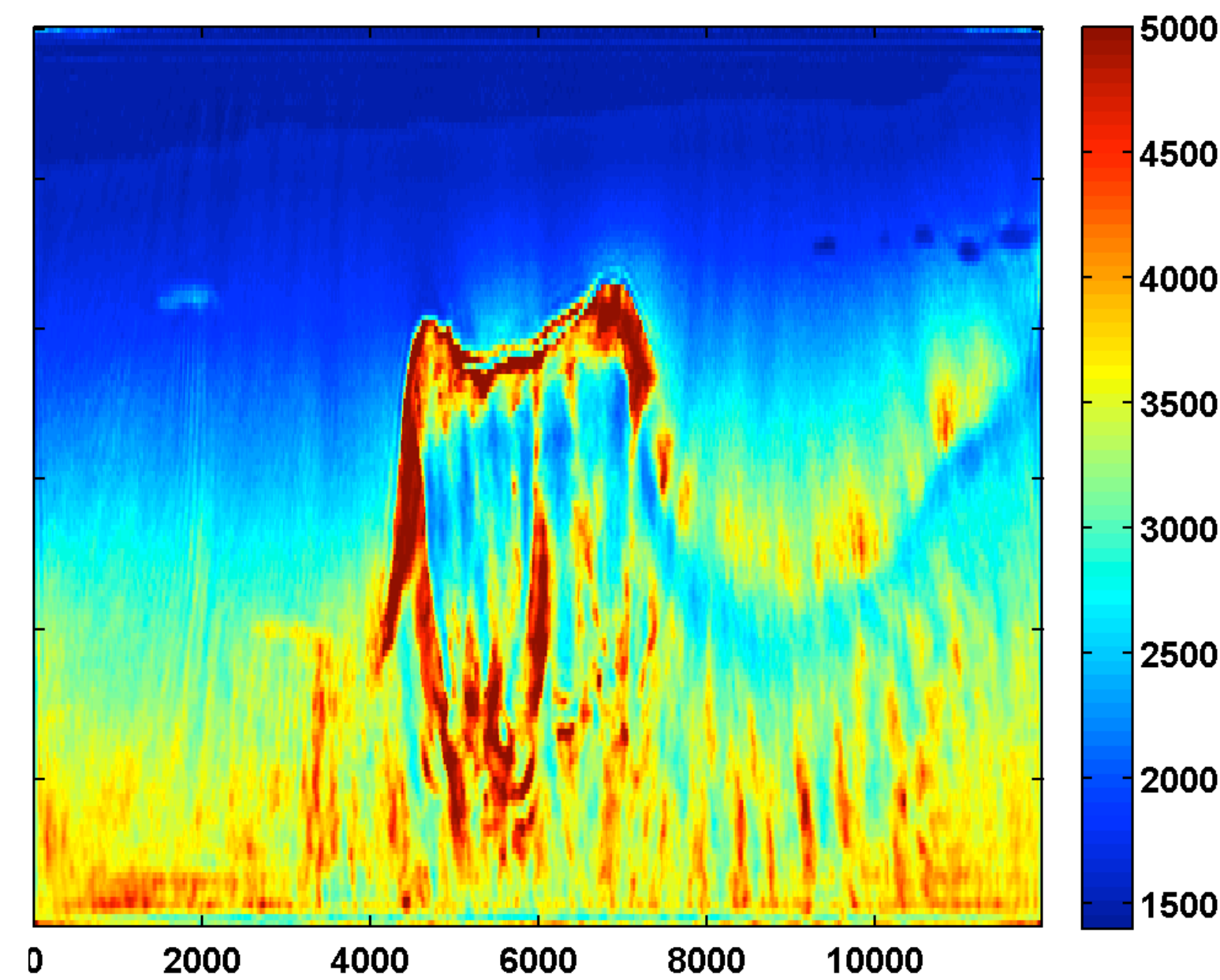


Results w/o TV

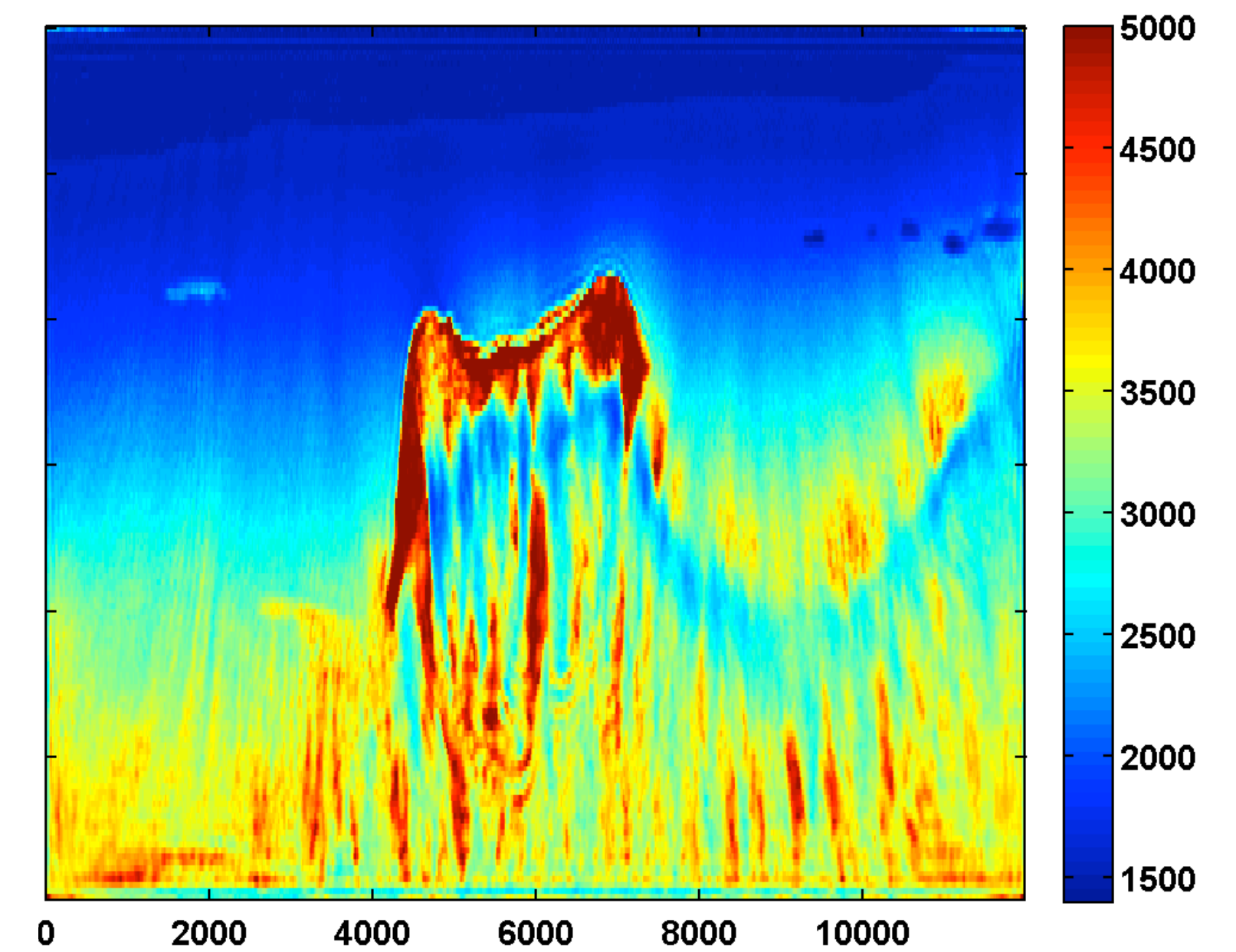
after one cycle through the frequencies



after two cycles through the frequencies

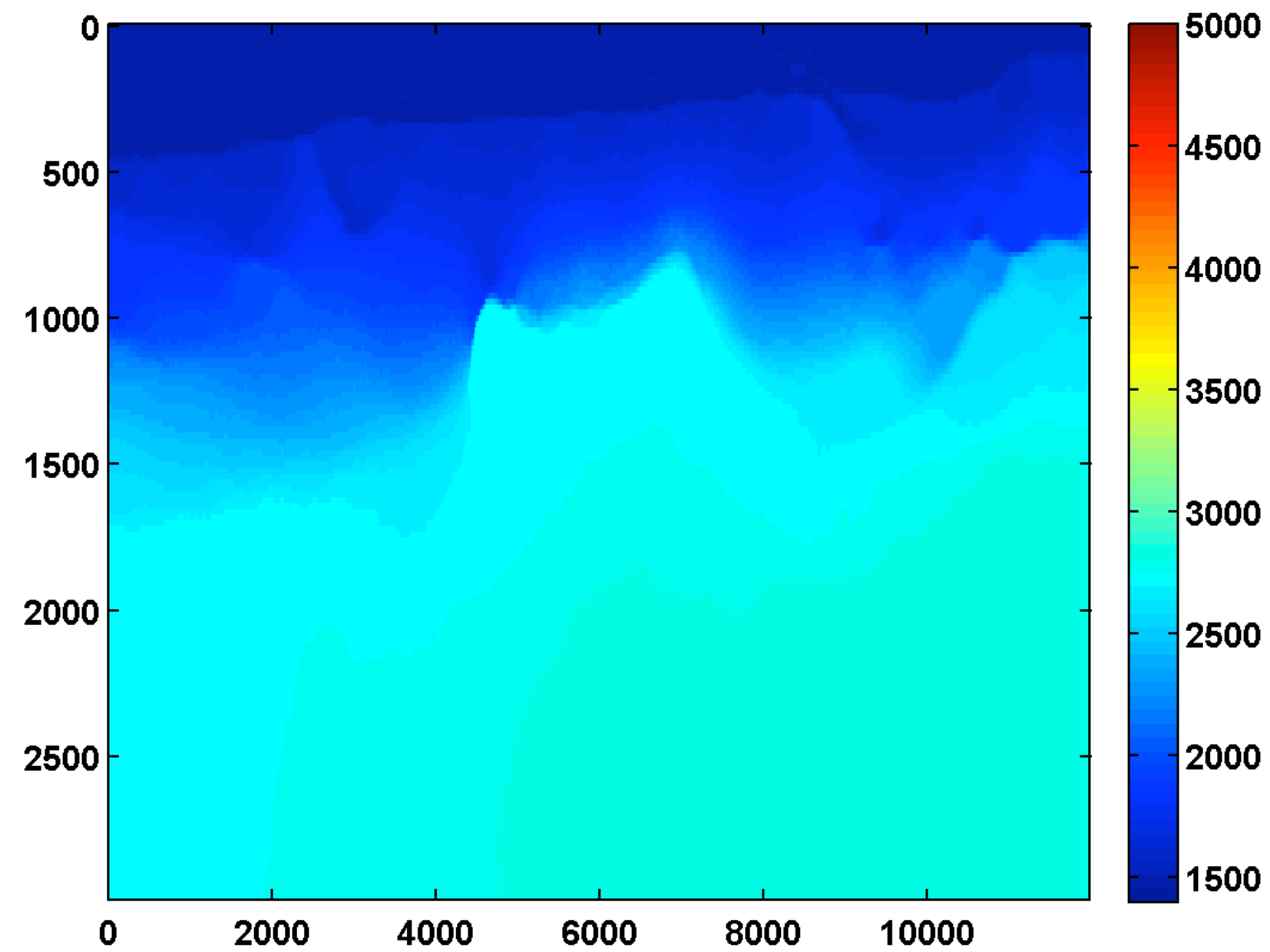


after three cycles through the frequencies

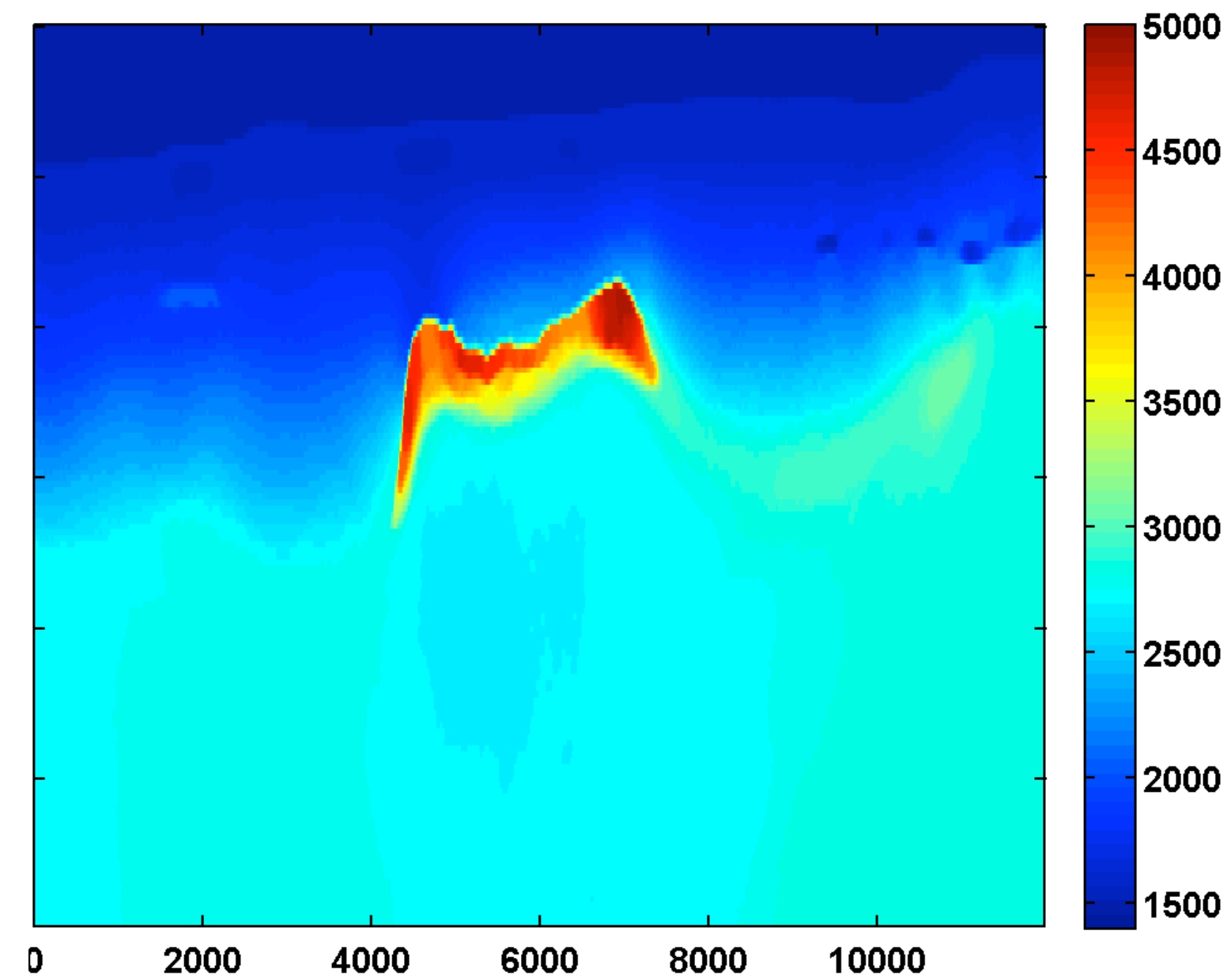


Results w/ TV

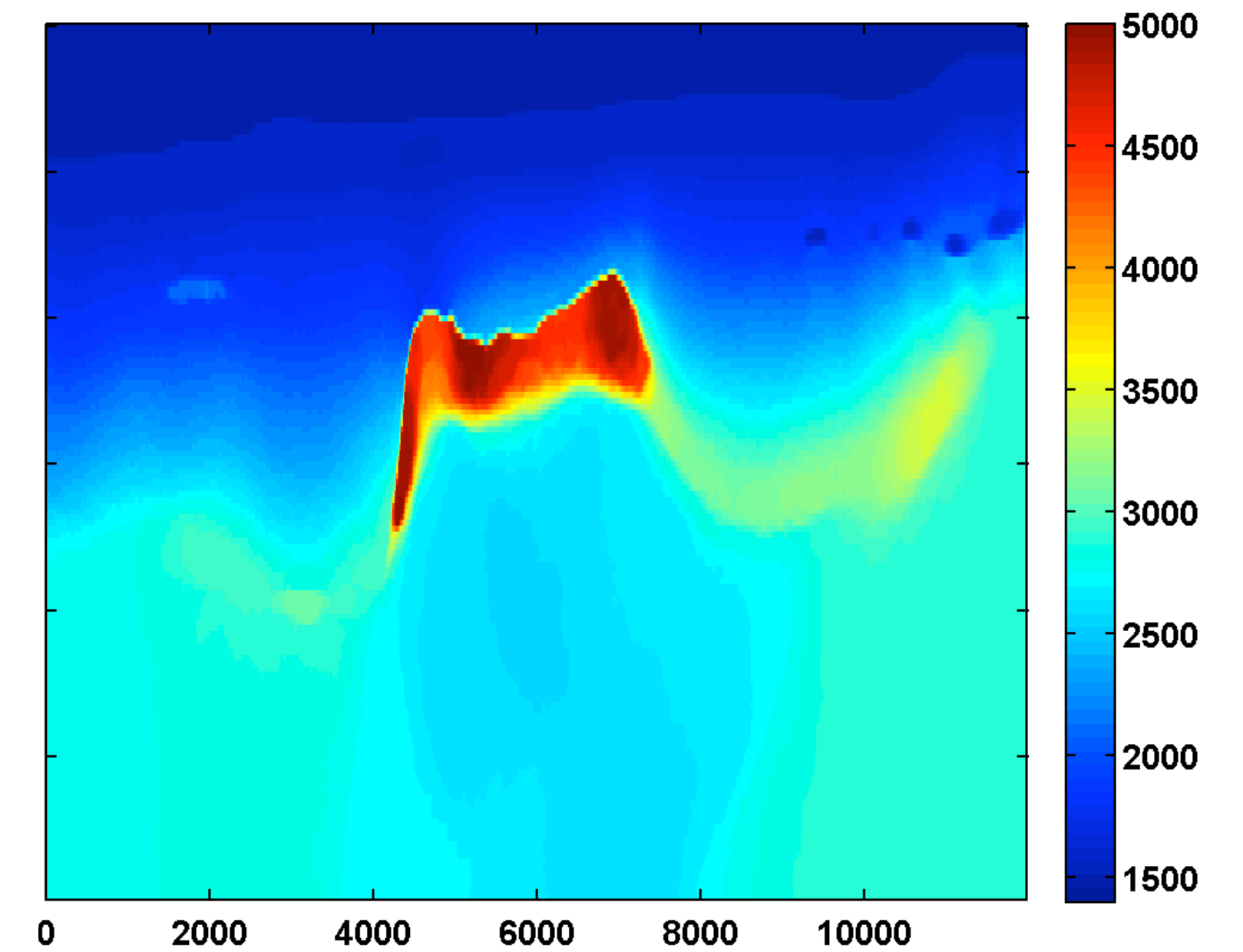
after one cycle through the frequencies



after two cycles through the frequencies



after three cycles through the frequencies



Hinge loss

one-sided TV constraint

Mitigate erroneous velocity model updates by

- ▶ using the fact that vertical slowness profiles tend to decrease w/ depth
- ▶ making it less probable that velocities step down along the vertical

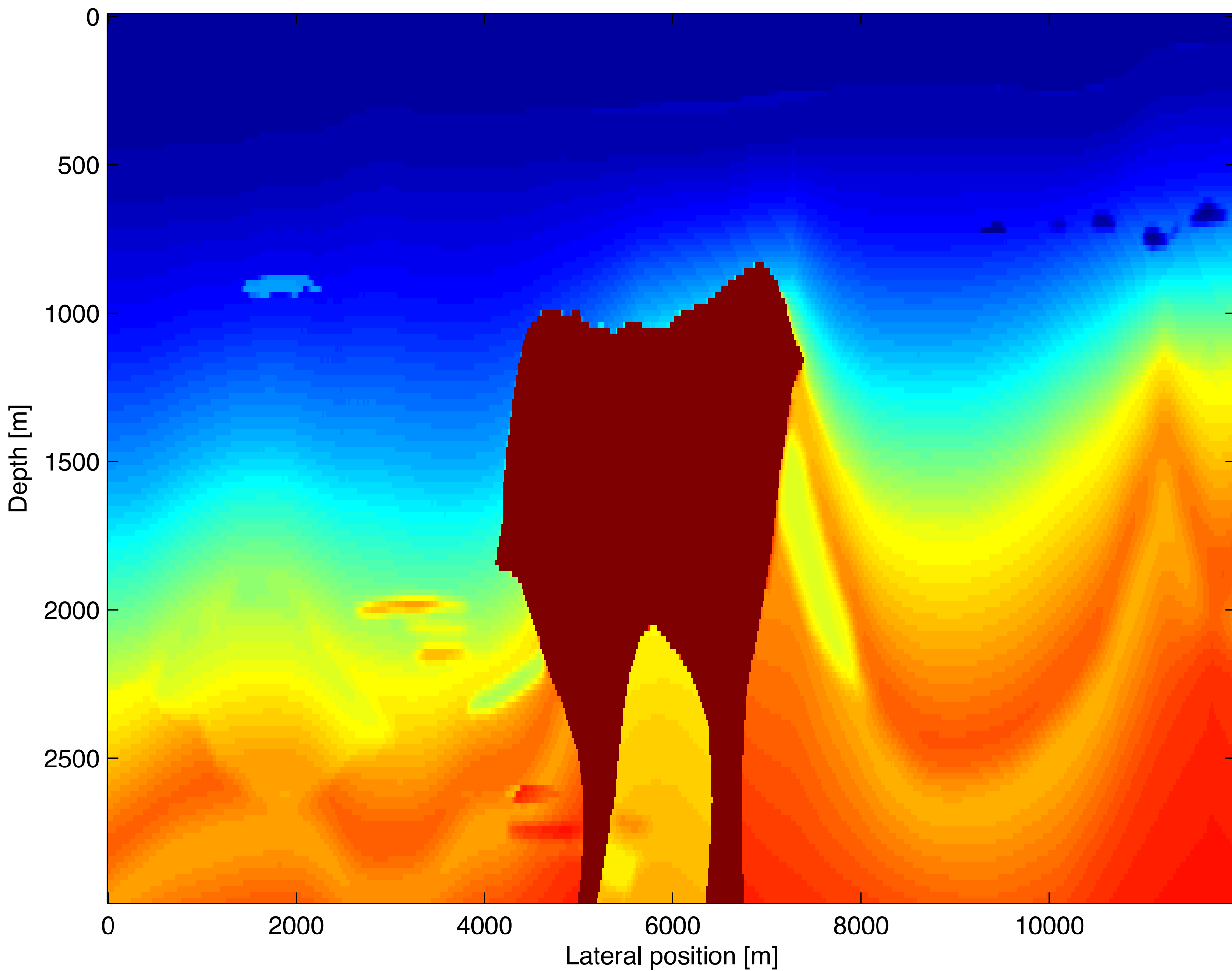
Mathematically expressed as the one-norm of a hinge-loss function

$$\| \max(0, D_z \mathbf{m}) \|_1 \leq \xi$$

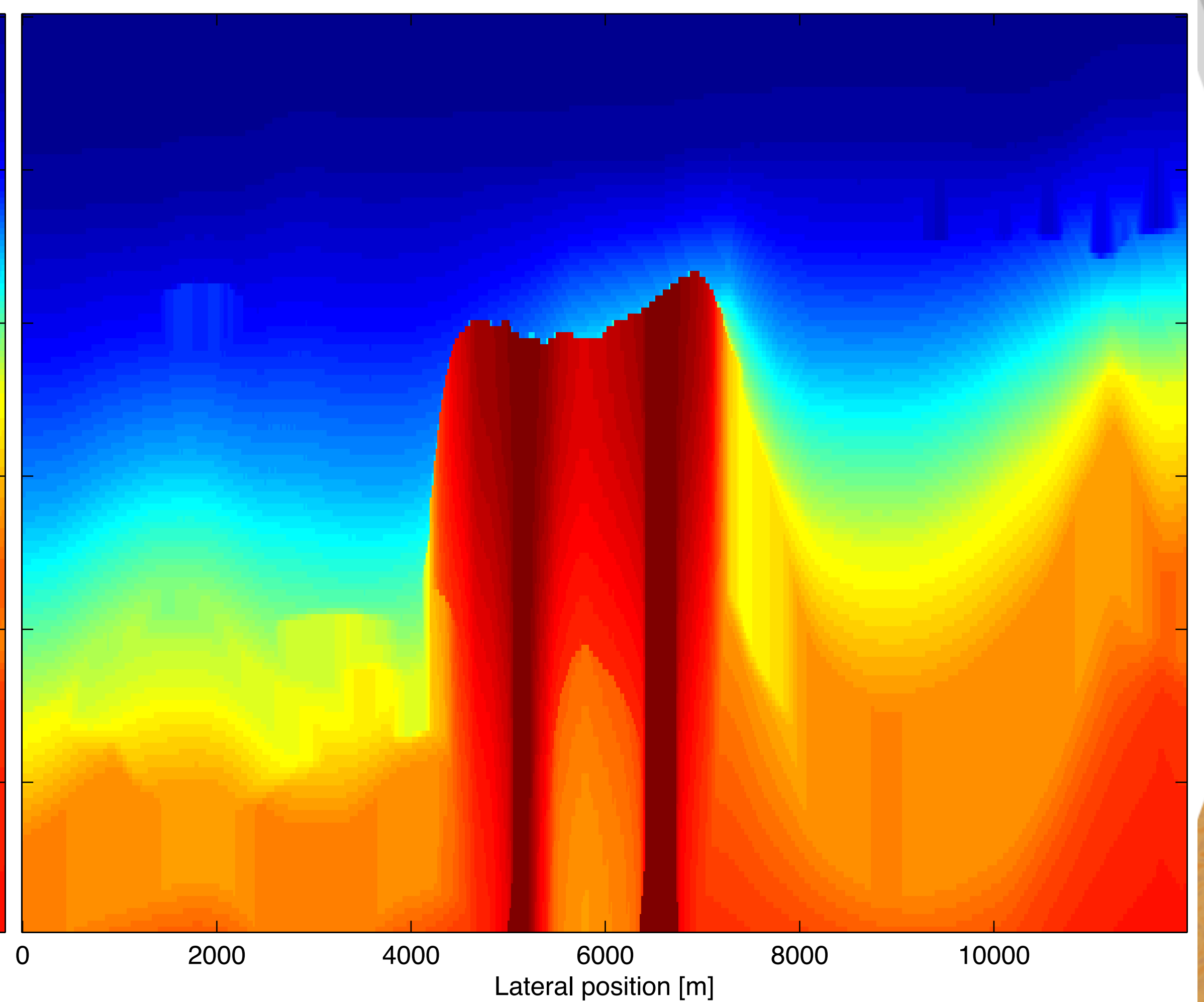
- ▶ for ξ small slowness is unlikely to step up
- ▶ extended to a weighted directional gradient
- ▶ combined w/ omni-directional TV and bound constraints

Hinge loss

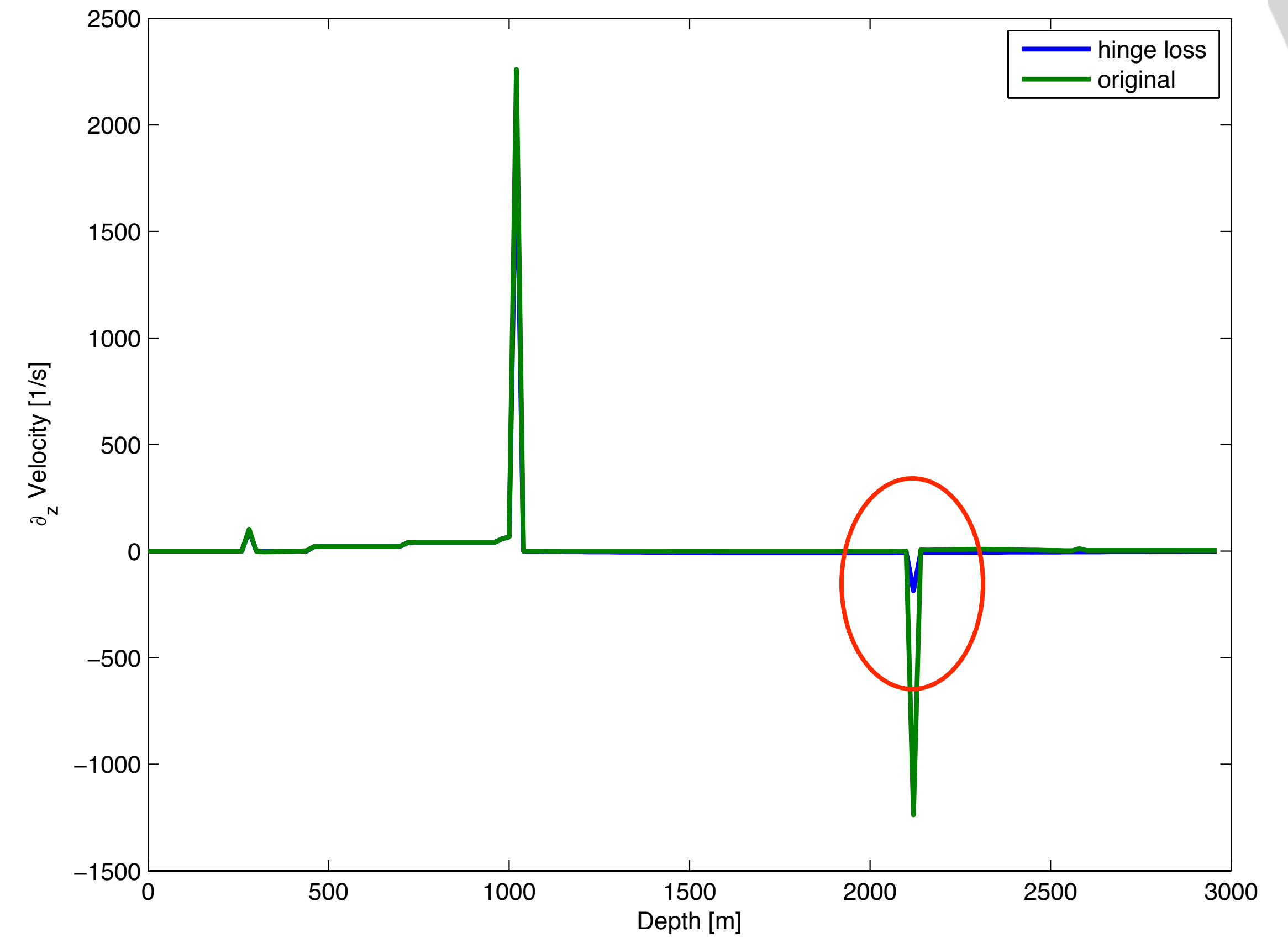
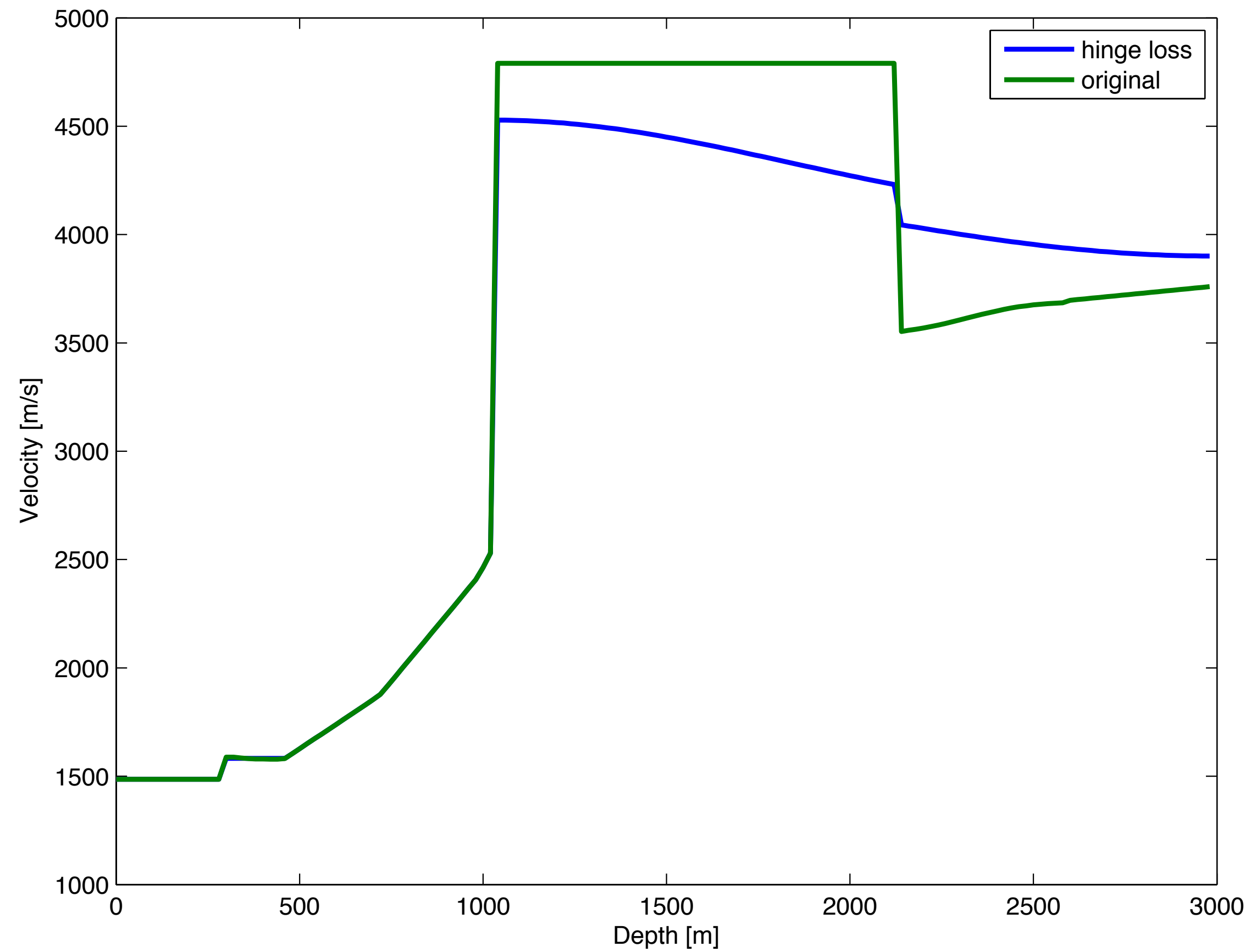
original



small hinge



Hinge loss – vertical profile



Scaled-gradient projections

w/ convex total-variation, box & hinge-loss constraints

Solve for given $\bar{\mathbf{u}}_\lambda$

$$\min_{\mathbf{m}} \phi(\mathbf{m}) \quad \text{subject to} \quad \begin{cases} m_i \in [B_1, B_2] \\ \|\mathbf{m}\|_{TV} \leq \tau \\ \|\mathbf{m}\|_{\text{Hinge}} \leq \xi \end{cases}$$

with

$$\|\mathbf{m}\|_{TV} = \sum_{ij} \frac{1}{h} \left\| \begin{bmatrix} m_{i,j+1} - m_{i,j} \\ m_{i+1,j} - m_{i,j} \end{bmatrix} \right\|$$

and

$$\|\mathbf{m}\|_{\text{Hinge}} = \|\max(0, D_z \mathbf{m})\|_1$$

Inner iterations

Solve

$$\underset{\mathbf{m}}{\text{minimize}} \Phi(\mathbf{m}) \quad \text{subject to} \quad \mathbf{m}^{n+1} \in C_{\text{box}} \cap C_{\text{TV}} \cap C_{\text{Hinge}}$$

by iterating

$$\mathbf{p}_1^{k+1} = \mathbf{p}_1^k + \delta D(\mathbf{m}^n + \Delta \mathbf{m}^k) - \Pi_{\|\cdot\|_{1,2} \leq \tau \delta}(\mathbf{p}_1^k + \delta D(\mathbf{m}^n + \Delta \mathbf{m}^k))$$

$$\mathbf{p}_2^{k+1} = \mathbf{p}_2^k + \delta D_z(\mathbf{m}^n + \Delta \mathbf{m}^k) - \Pi_{\|\max(0, \cdot)\|_1 \leq \xi \delta}(\mathbf{p}_2^k + \delta D_z(\mathbf{m}^n + \Delta \mathbf{m}^k))$$

$$\Delta \mathbf{m}_i^{k+1} = \max((B_i^l - \mathbf{m}_i^n), B_i)$$

where

$$B_i = \min \left((B_i^u - \mathbf{m}_i^n), \left[\left(\mathbf{H}_{GN}^n + \left(c_n + \frac{1}{\alpha} \right) \mathbf{I} \right)^{-1} \left(-\mathbf{g}^n + \frac{\Delta \mathbf{m}^k}{\alpha} - D^T(2\mathbf{p}_1^{k+1} - \mathbf{p}_1^k) - D_z^T(2\mathbf{p}_2^{k+1} - \mathbf{p}_2^k) \right) \right]_i \right)$$

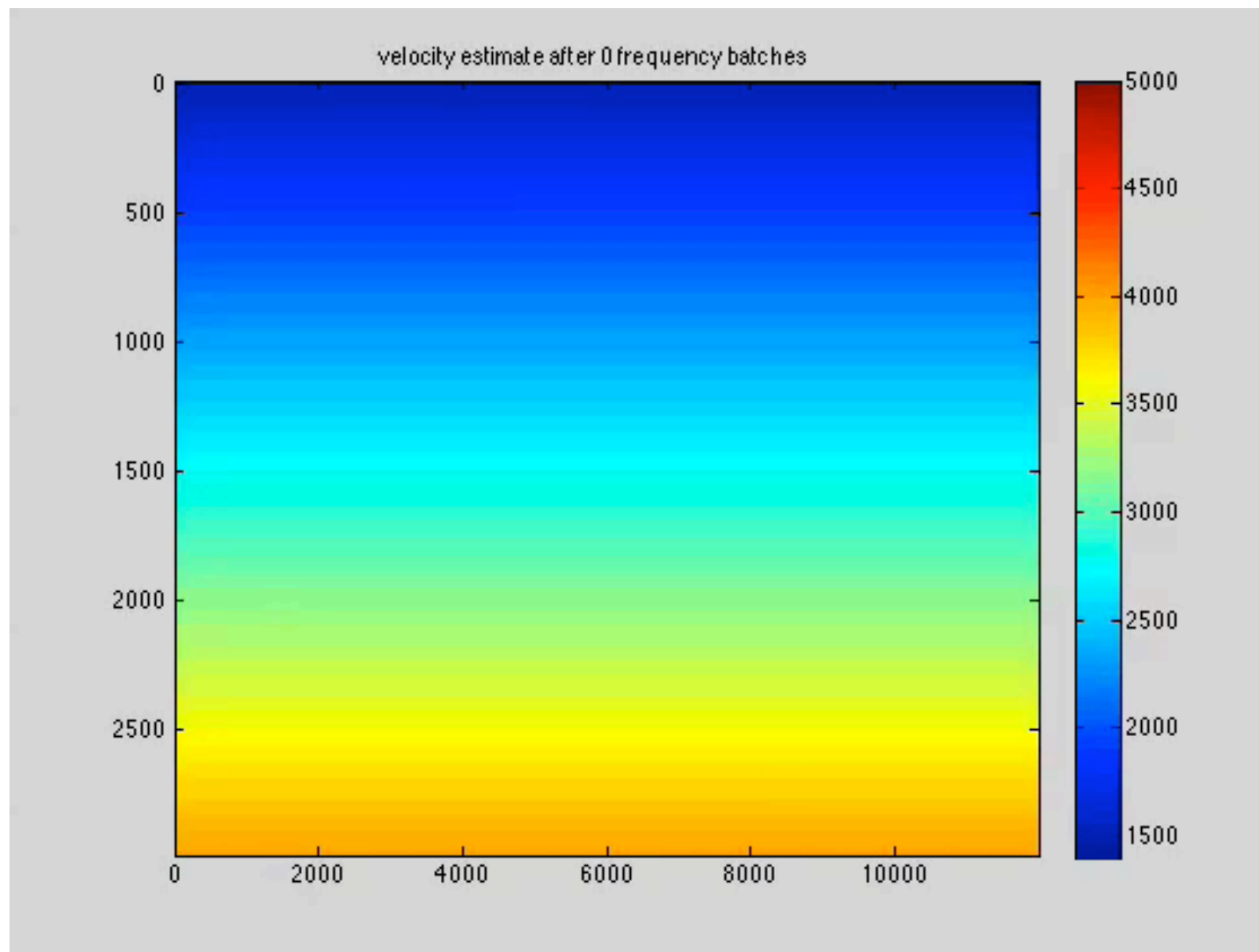
BP model – inversion parameters

- ▶ number of sources: 126
- ▶ number of receivers: 299
- ▶ frequency continuation over 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 15Hz peak frequency
- ▶ two simultaneous shots with Gaussian weights w/ redraws
- ▶ noise-free & inversion crime

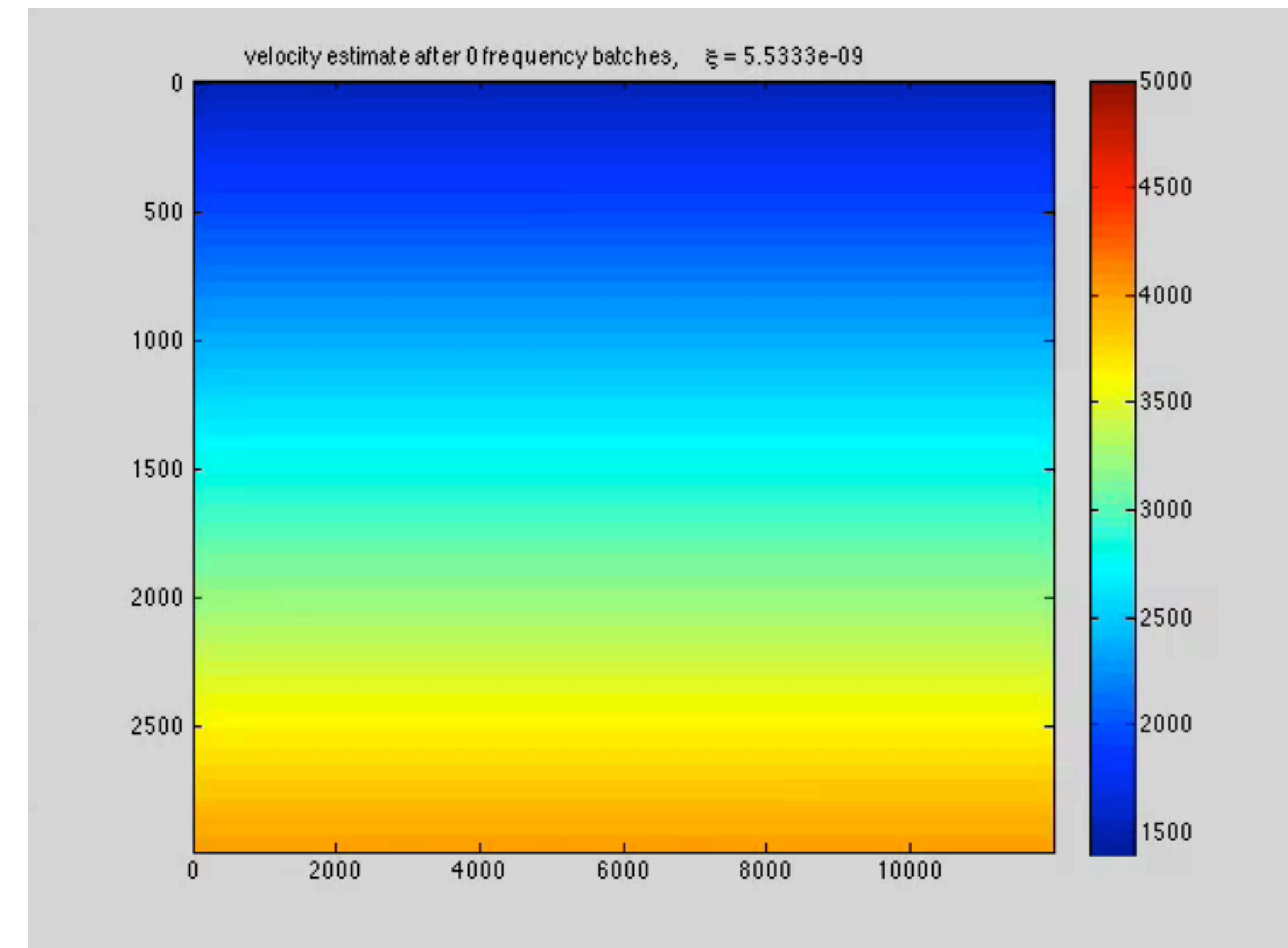
WRI

w/ or w/o TV-norm & hinge-loss projections & poor starting model

w/o TV & Hinge



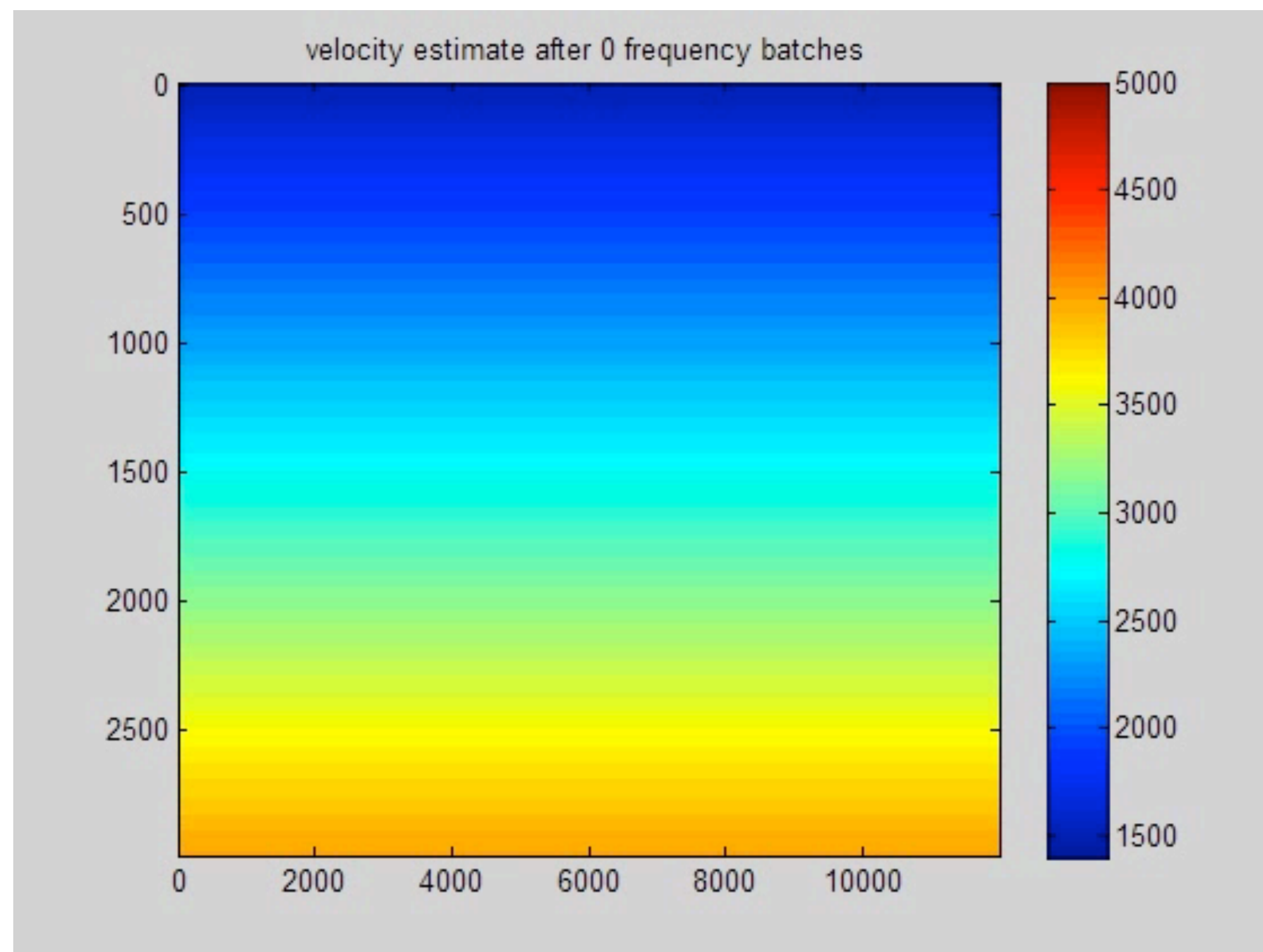
w/ TV & Hinge



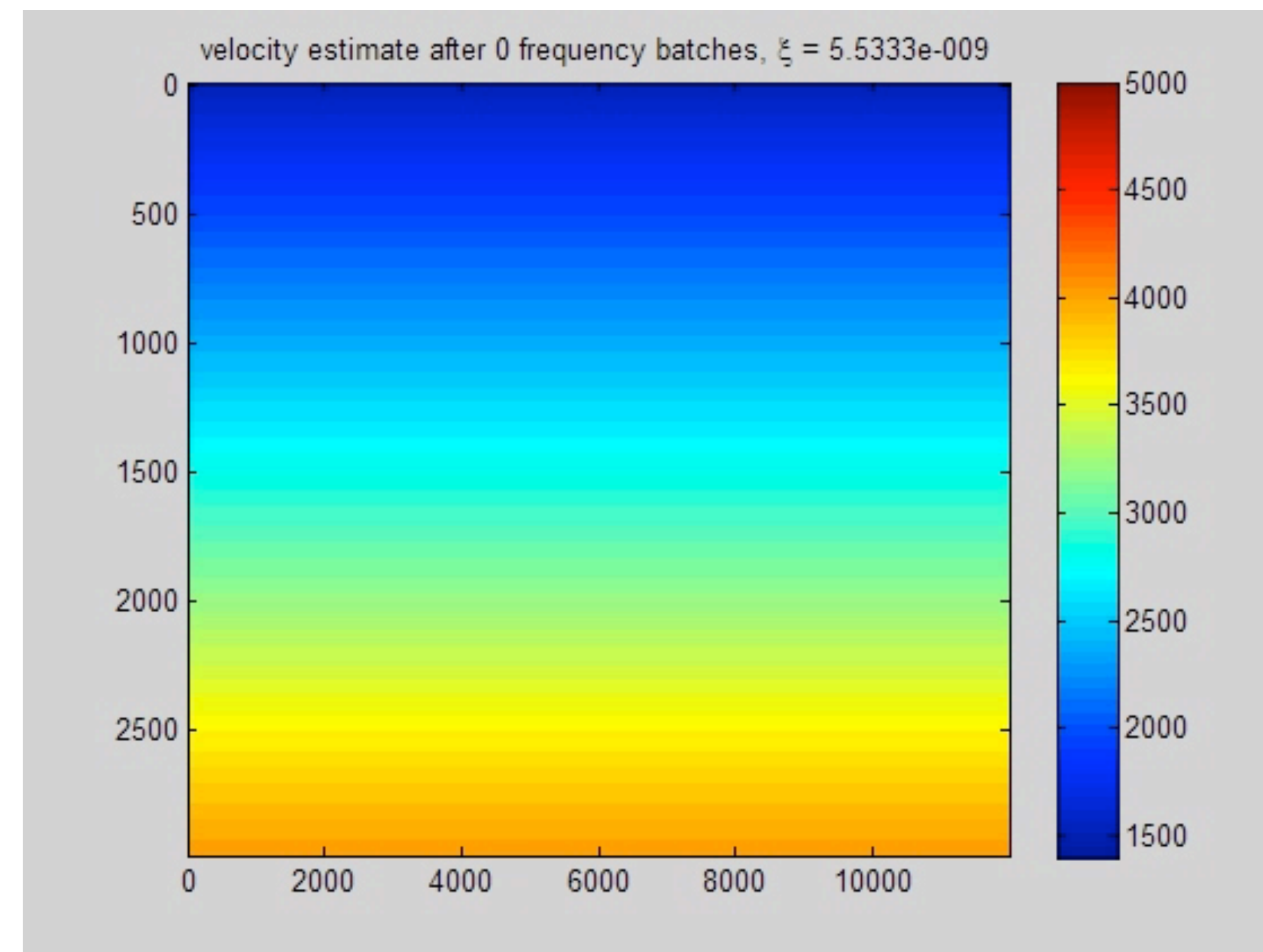
Adjoint-state FWI

w/ or w/o TV-norm & hinge-loss projections & poor starting model

w/o TV & Hinge

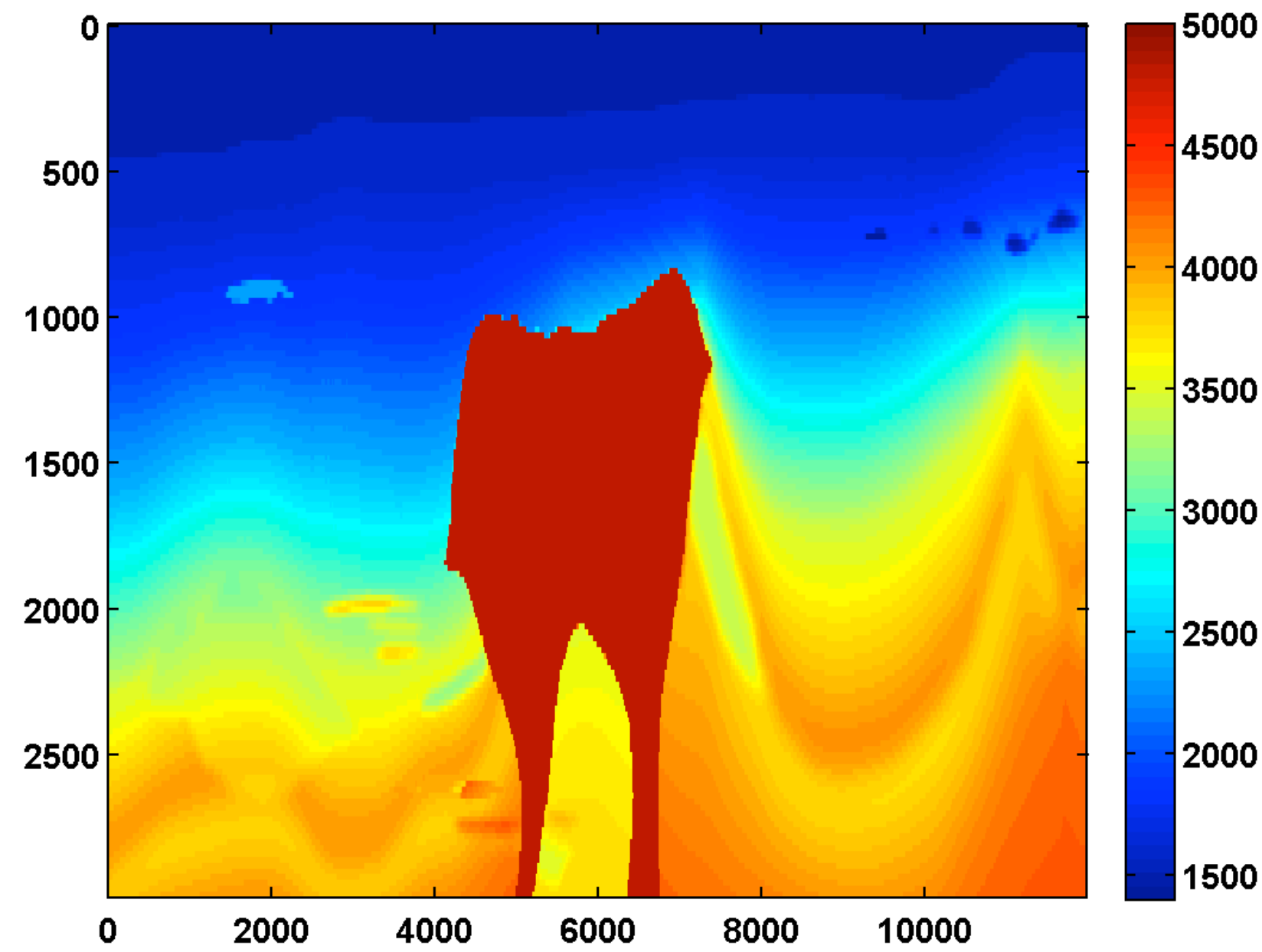


w/ TV & Hinge

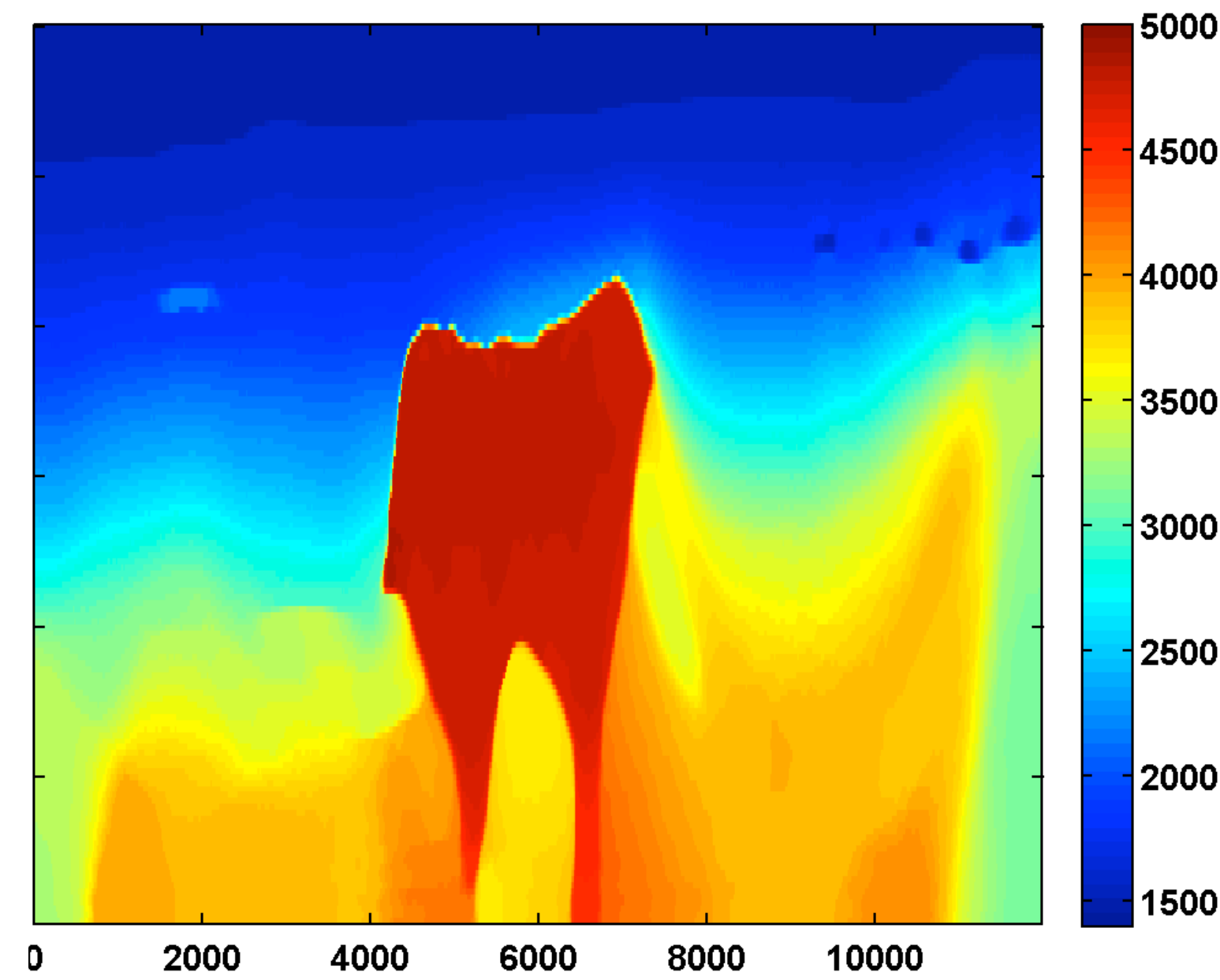


WRI vs adjoint-state

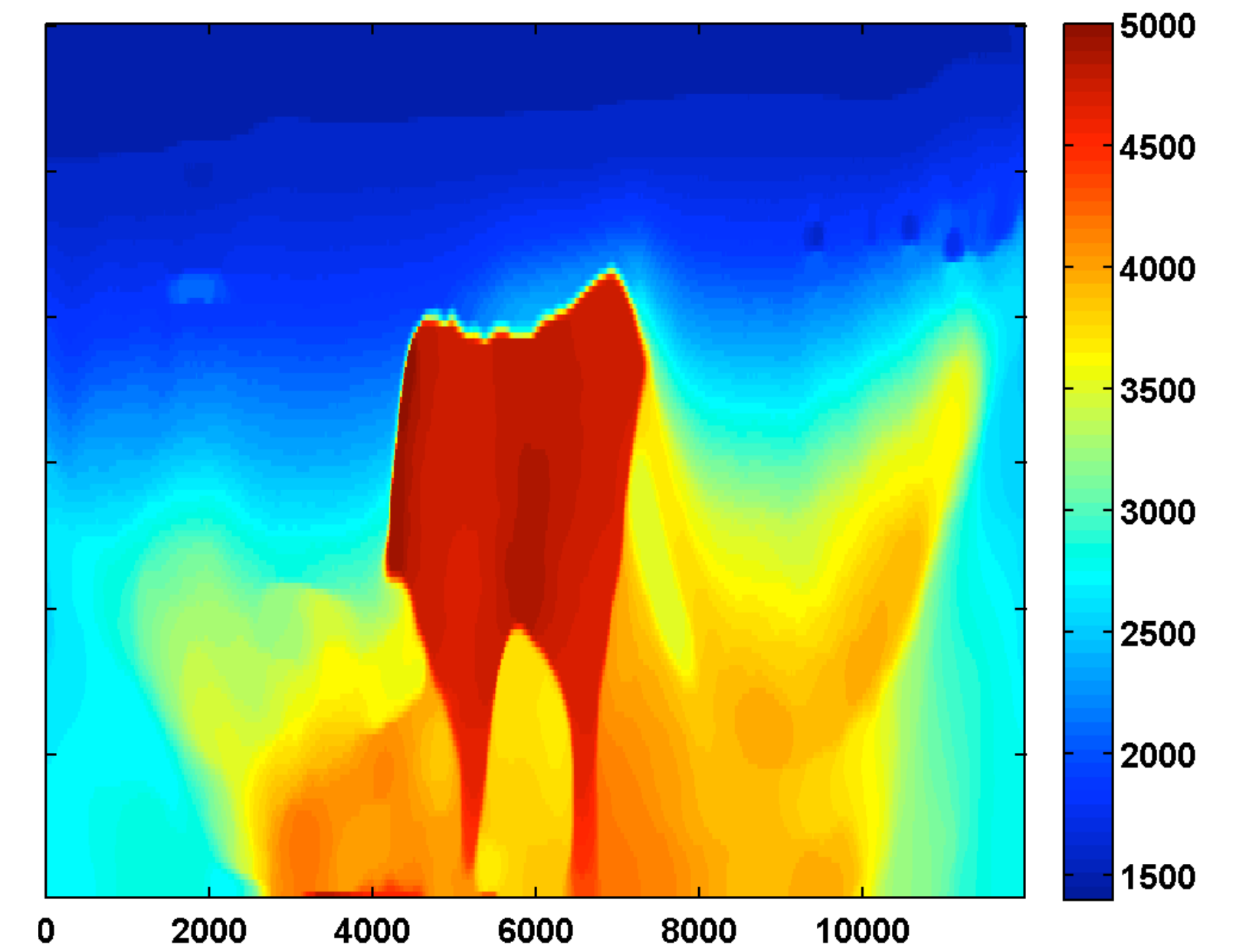
initial model



WRI



adjoint-state



Why this works?

Combination of

- ▶ multiscale frequency cycles
- ▶ relaxation of *asymmetric* constraints

works when

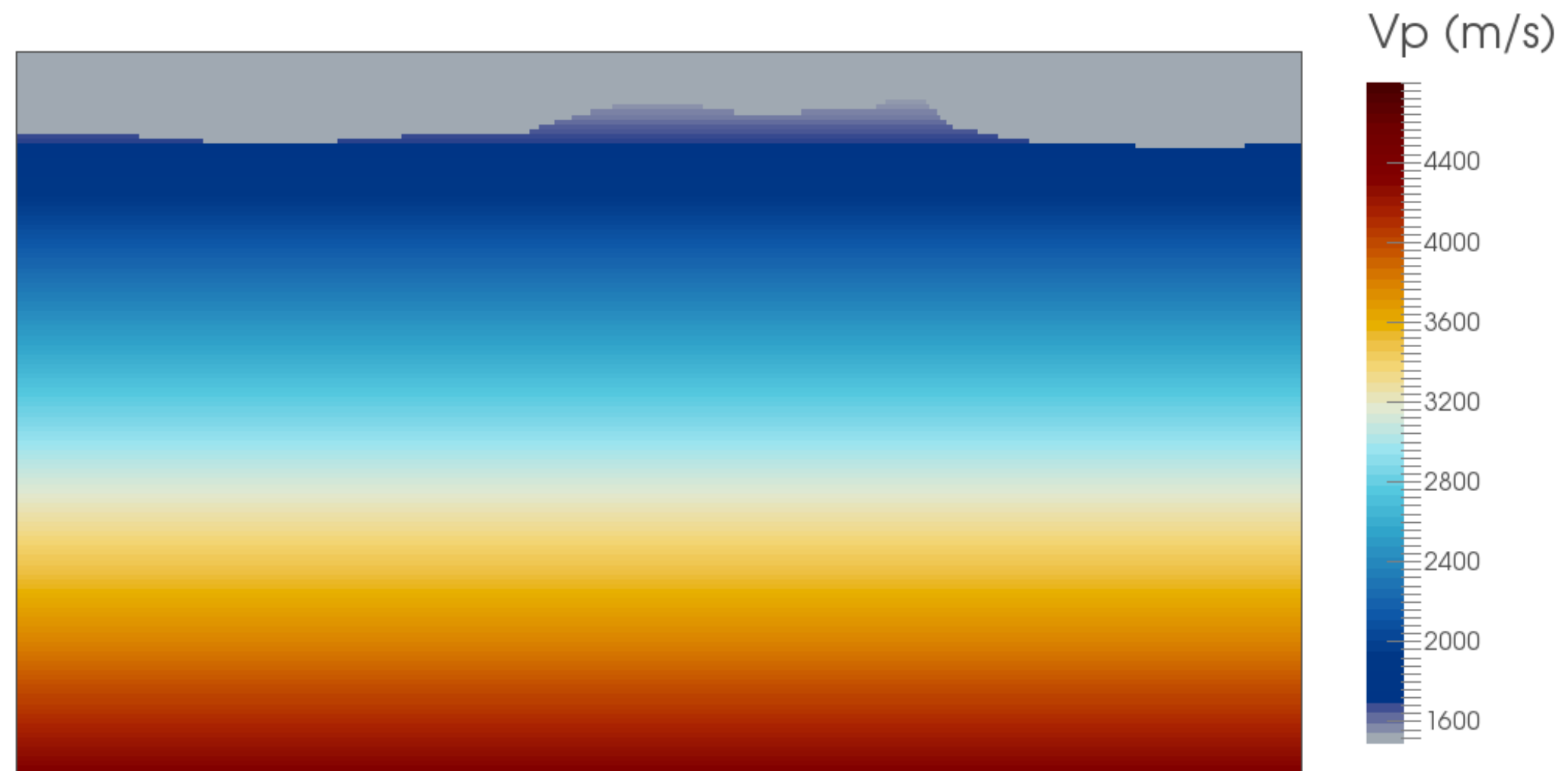
- ▶ sharp reflectors are introduced that are of the correct sign
- ▶ progress is made during previous cycles
- ▶ adverse affects of local minima are controlled by constraints
- ▶ “fine-scales” contribute to “coarse-scales” of the next cycle

“Deep” SEAM model – inversion parameters

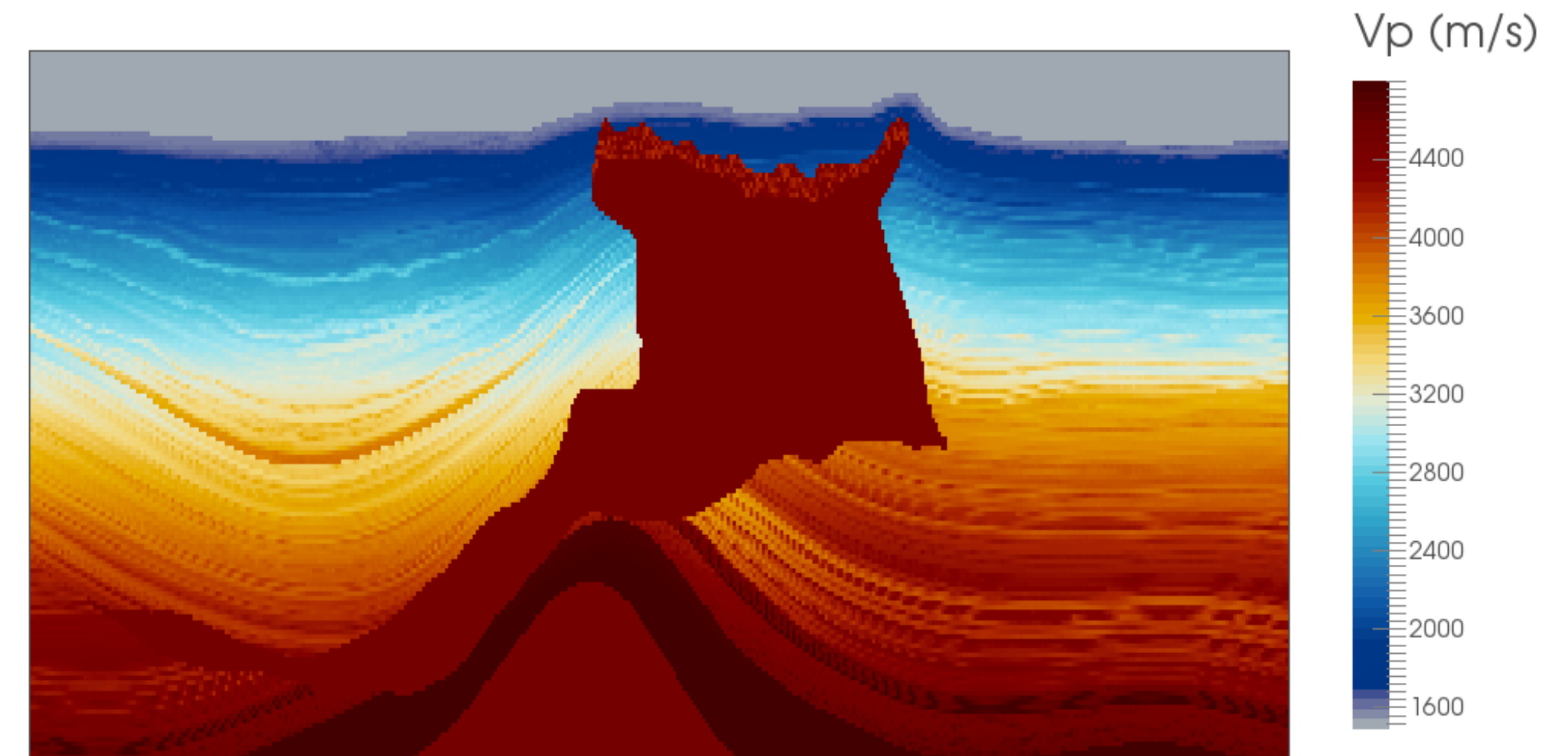
- ▶ model size 401x151
- ▶ grid size: 20 m
- ▶ frequencies: 3 – 12 Hz
- ▶ source: 10 Hz Ricker wavelet
- ▶ 98 sources
- ▶ 196 receivers
- ▶ two simultaneous shots with Gaussian weights w/ redraws
- ▶ noise-free & inversion crime

“Deep” SEAM model

starting model

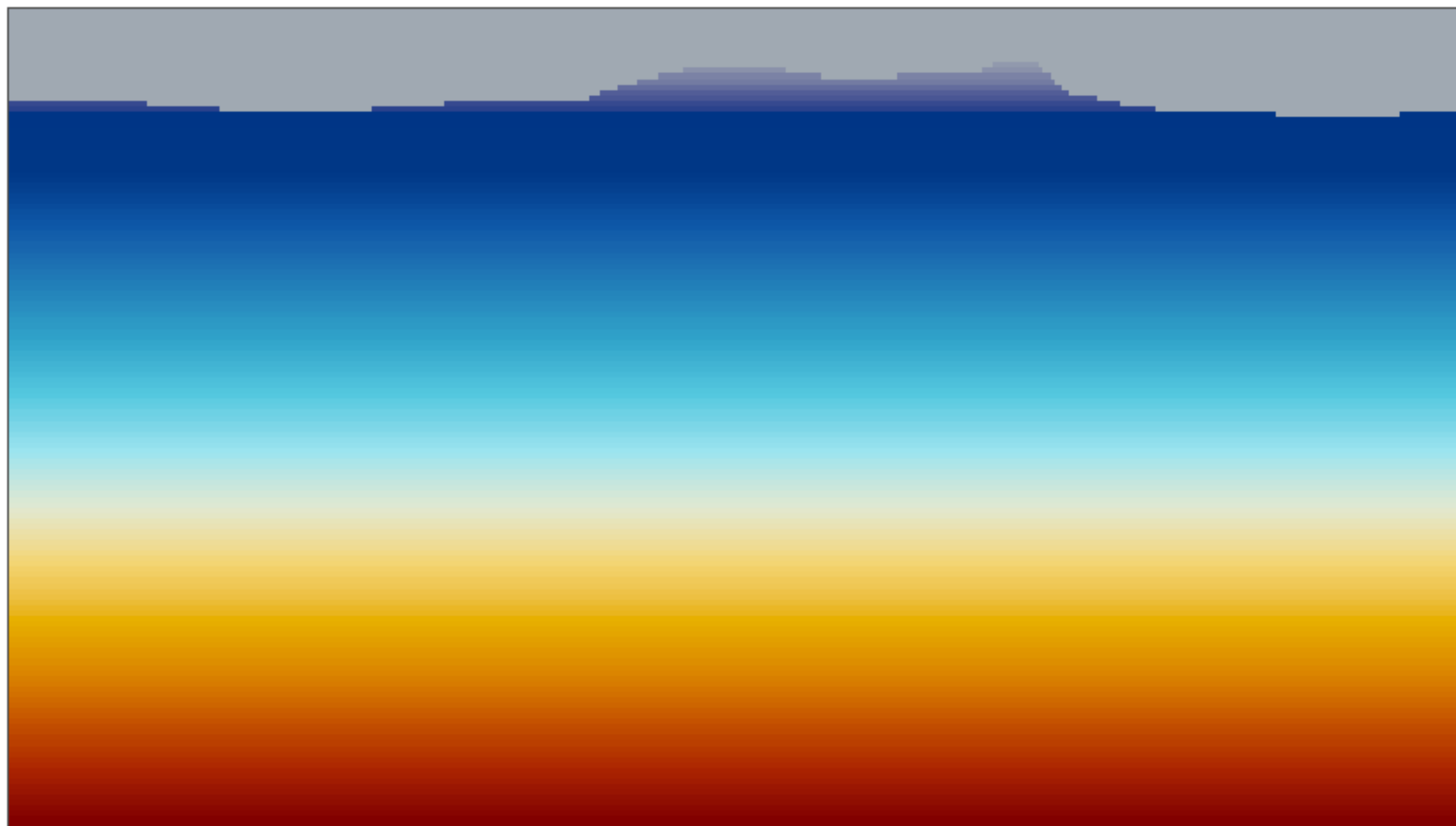


true model

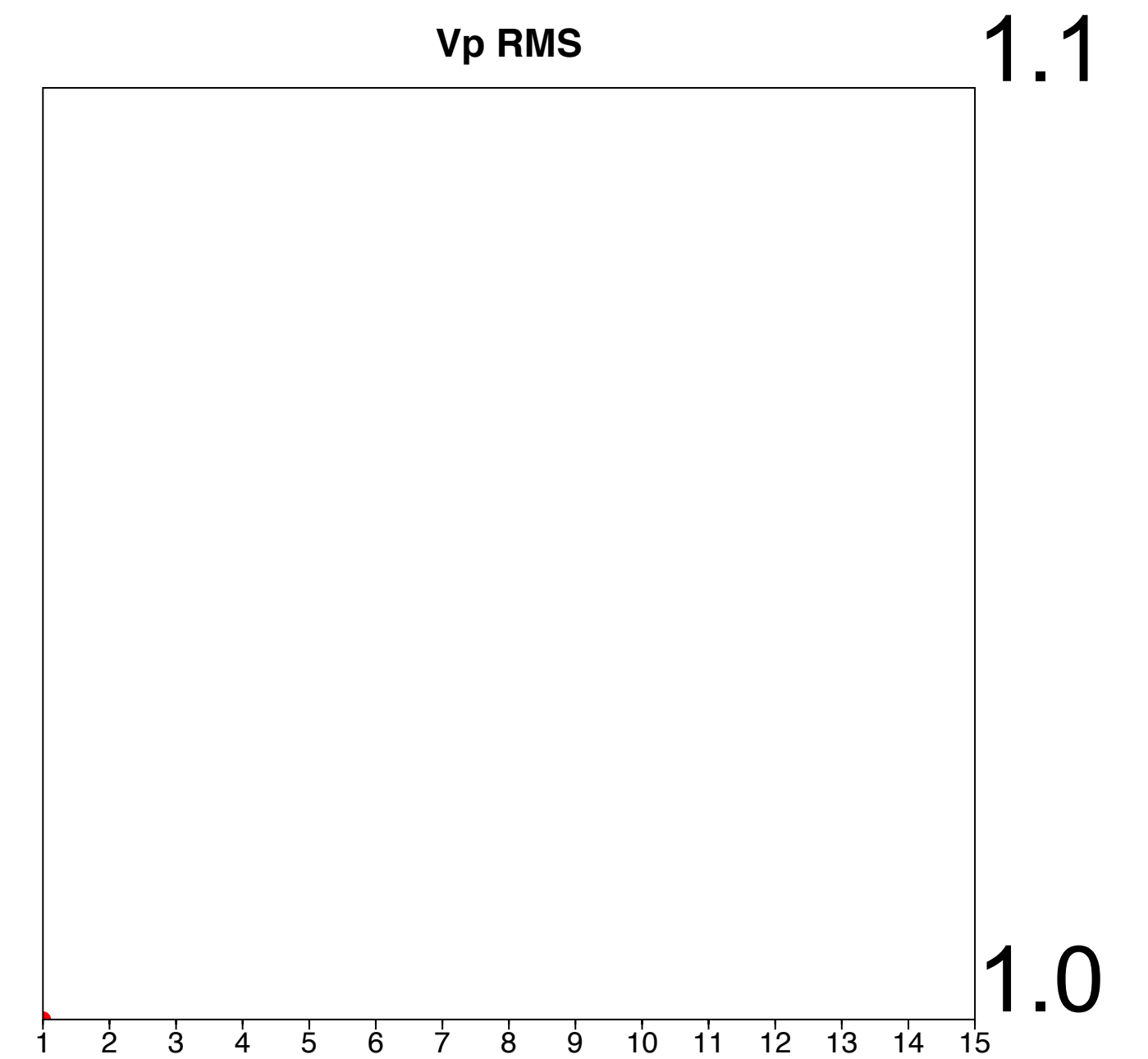
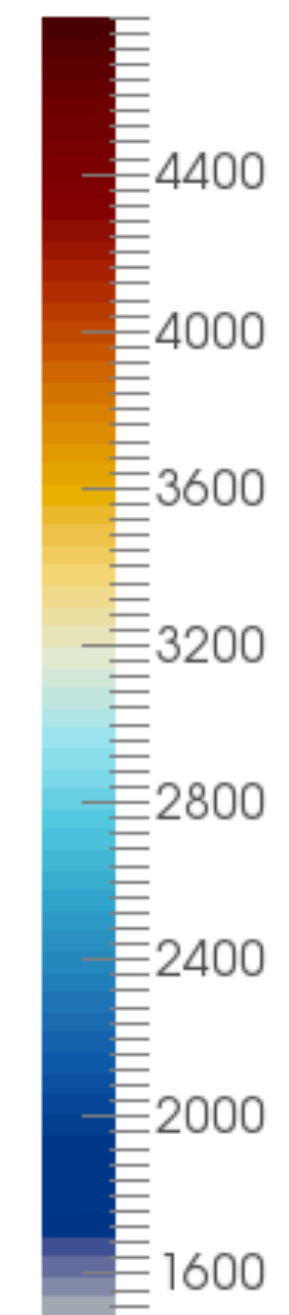


Models from SEG Advanced Modeling Corporation (SEAM)

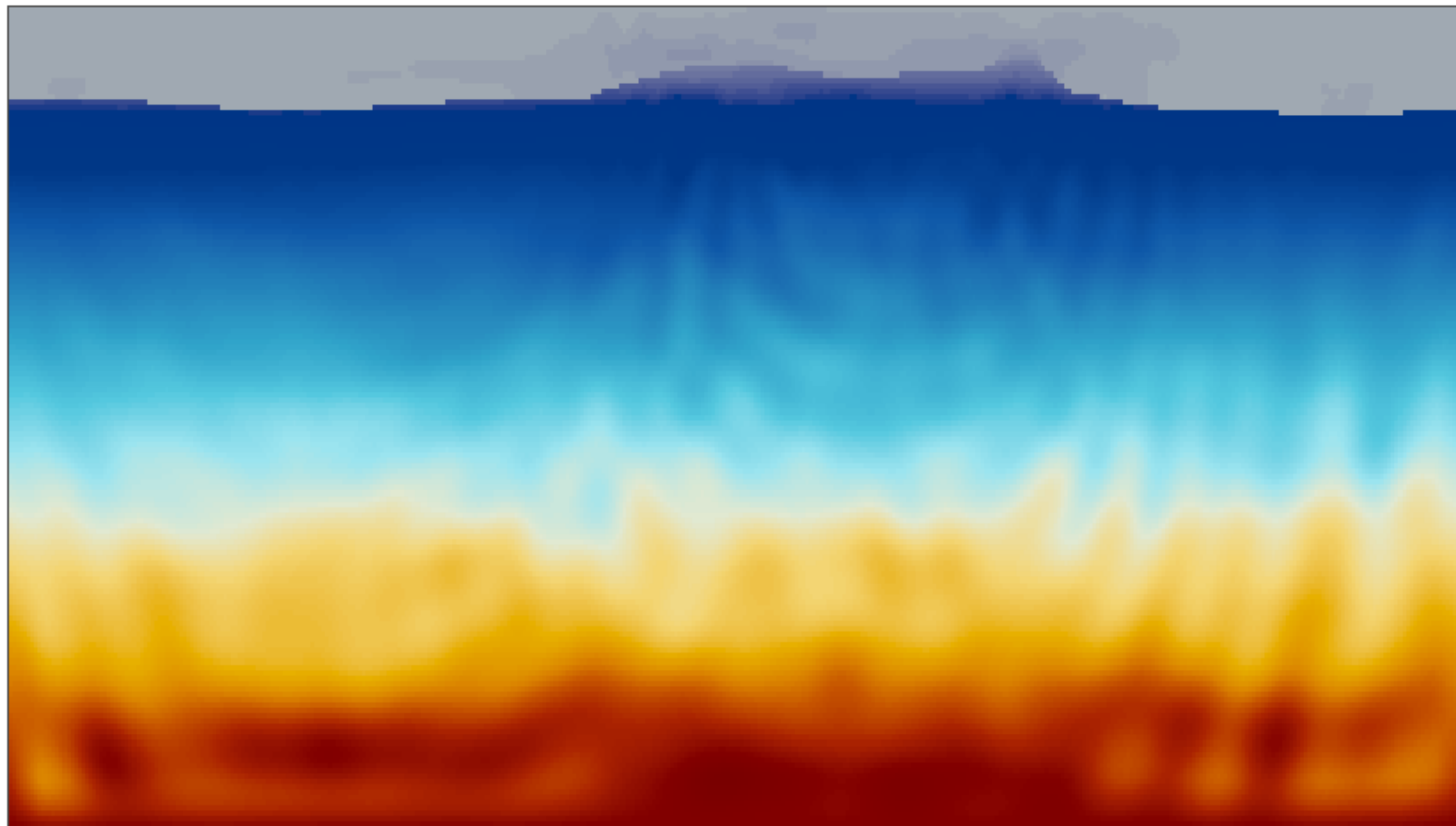
Adjoint-state – w/o constraints



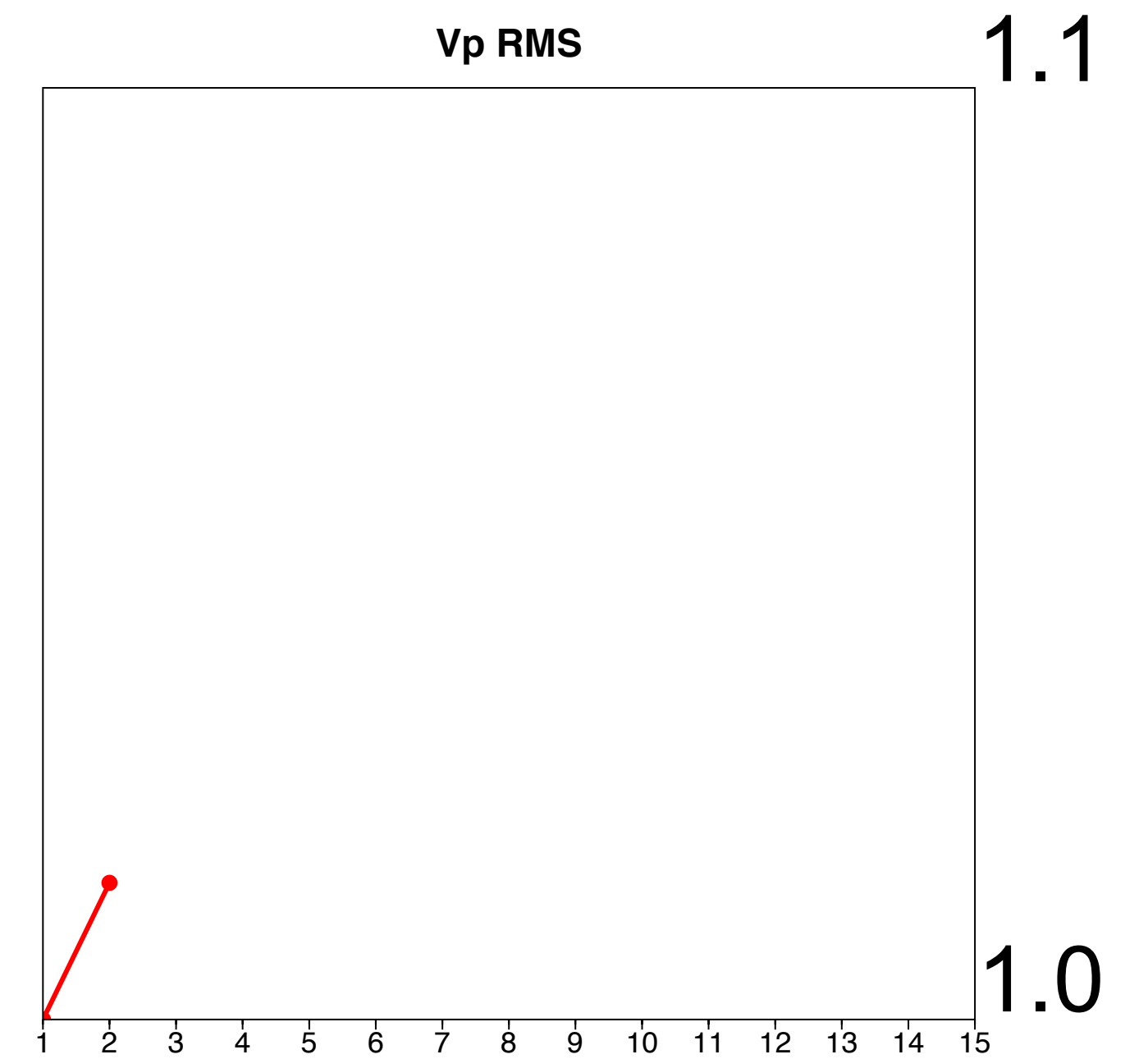
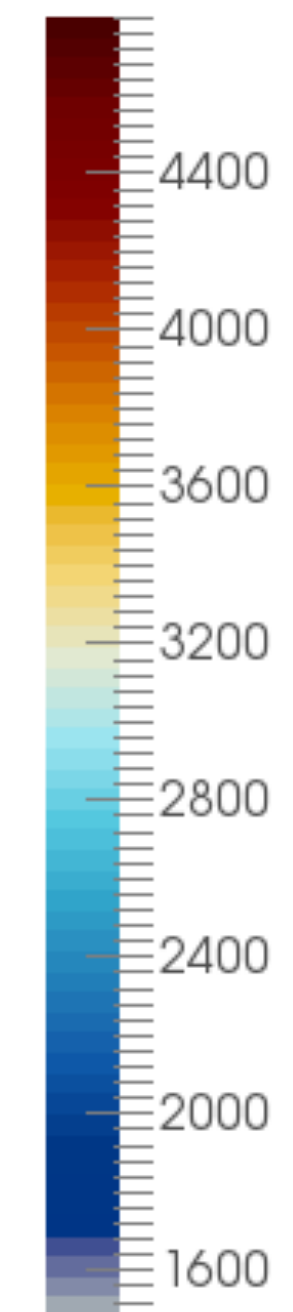
Vp (m/s)



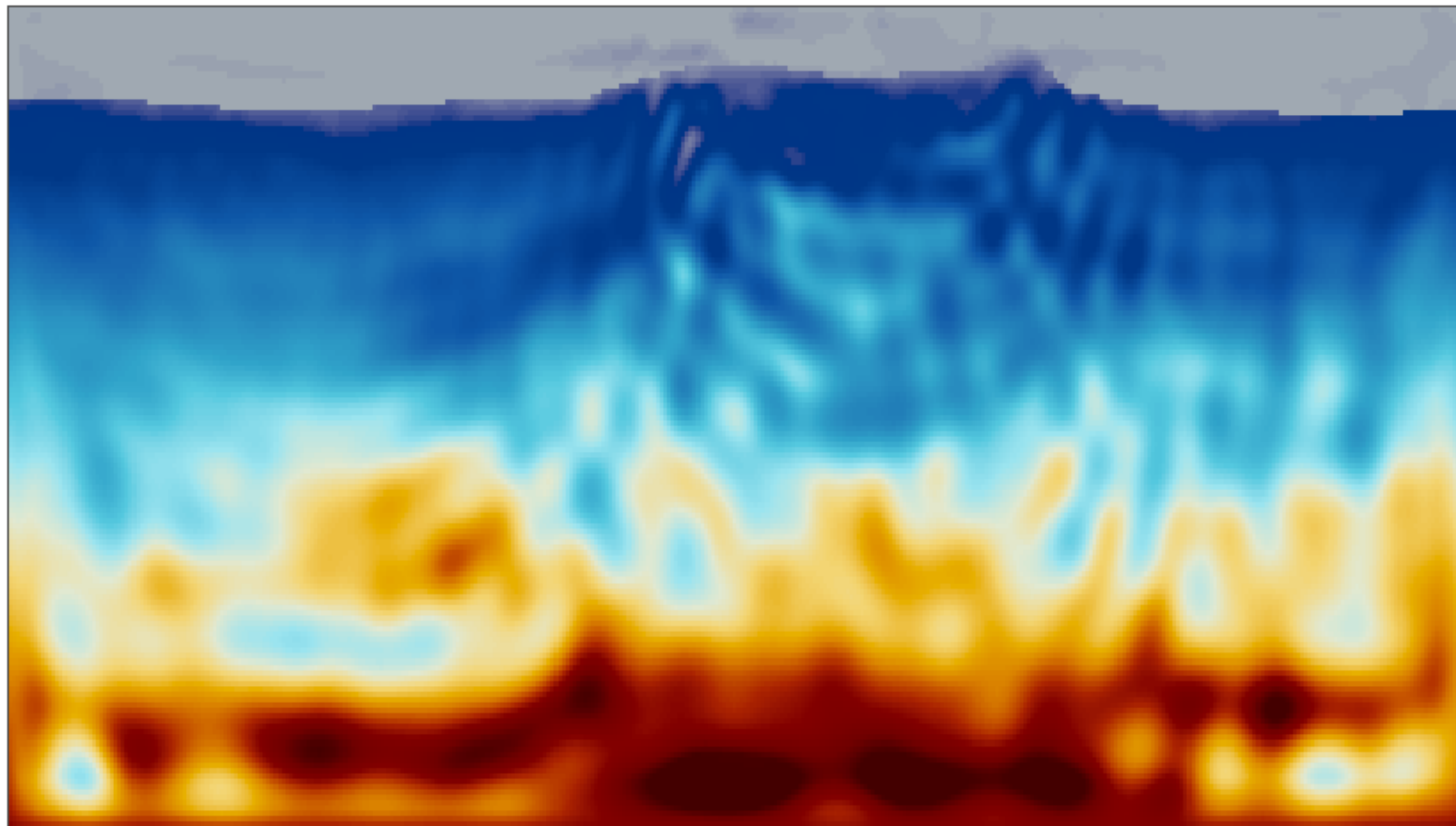
Adjoint-state – w/o constraints



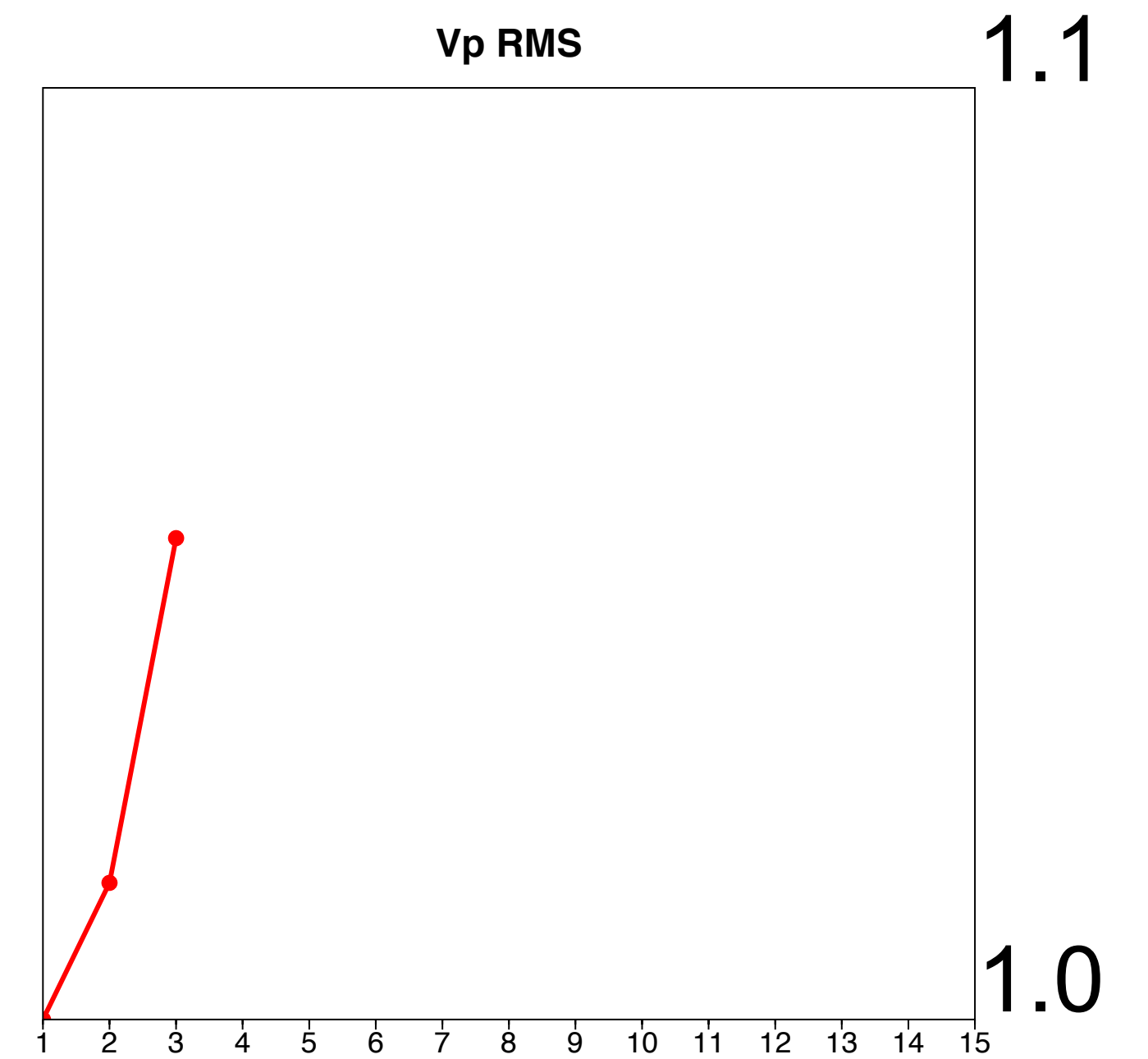
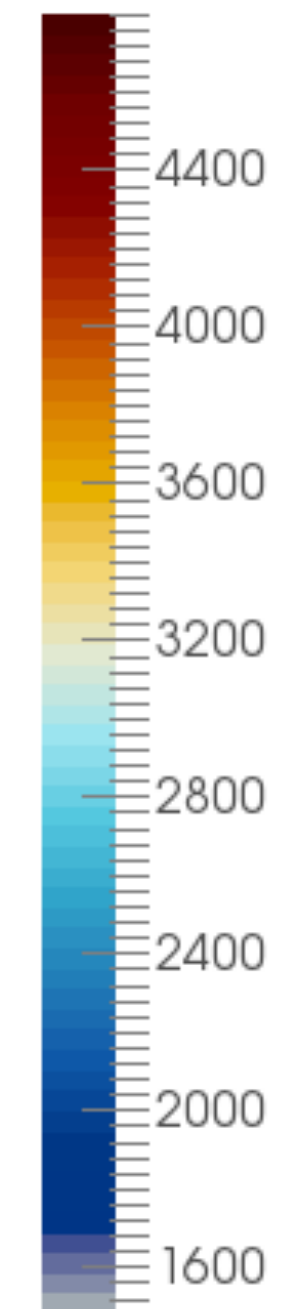
V_p (m/s)



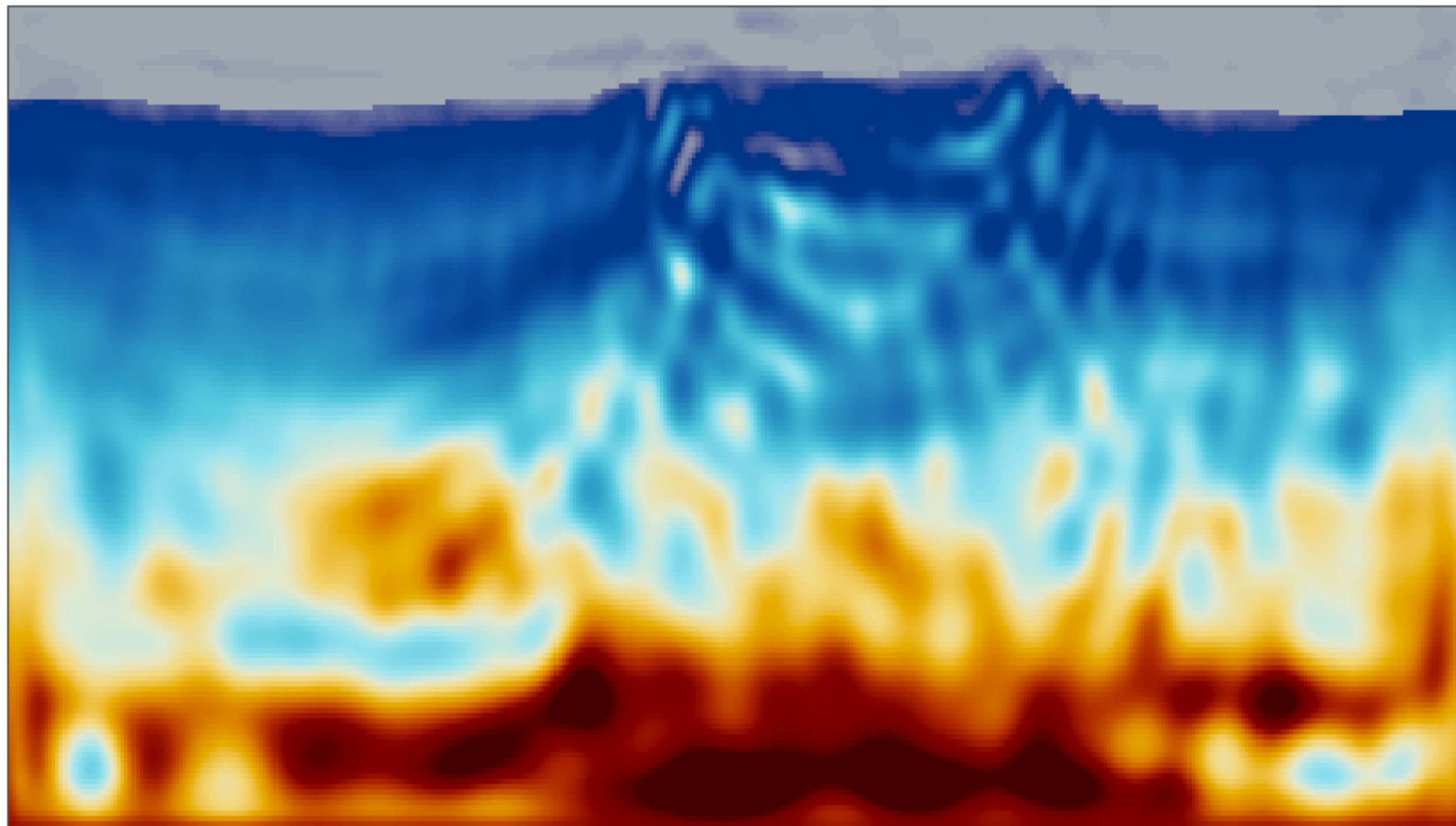
Adjoint-state – w/o constraints



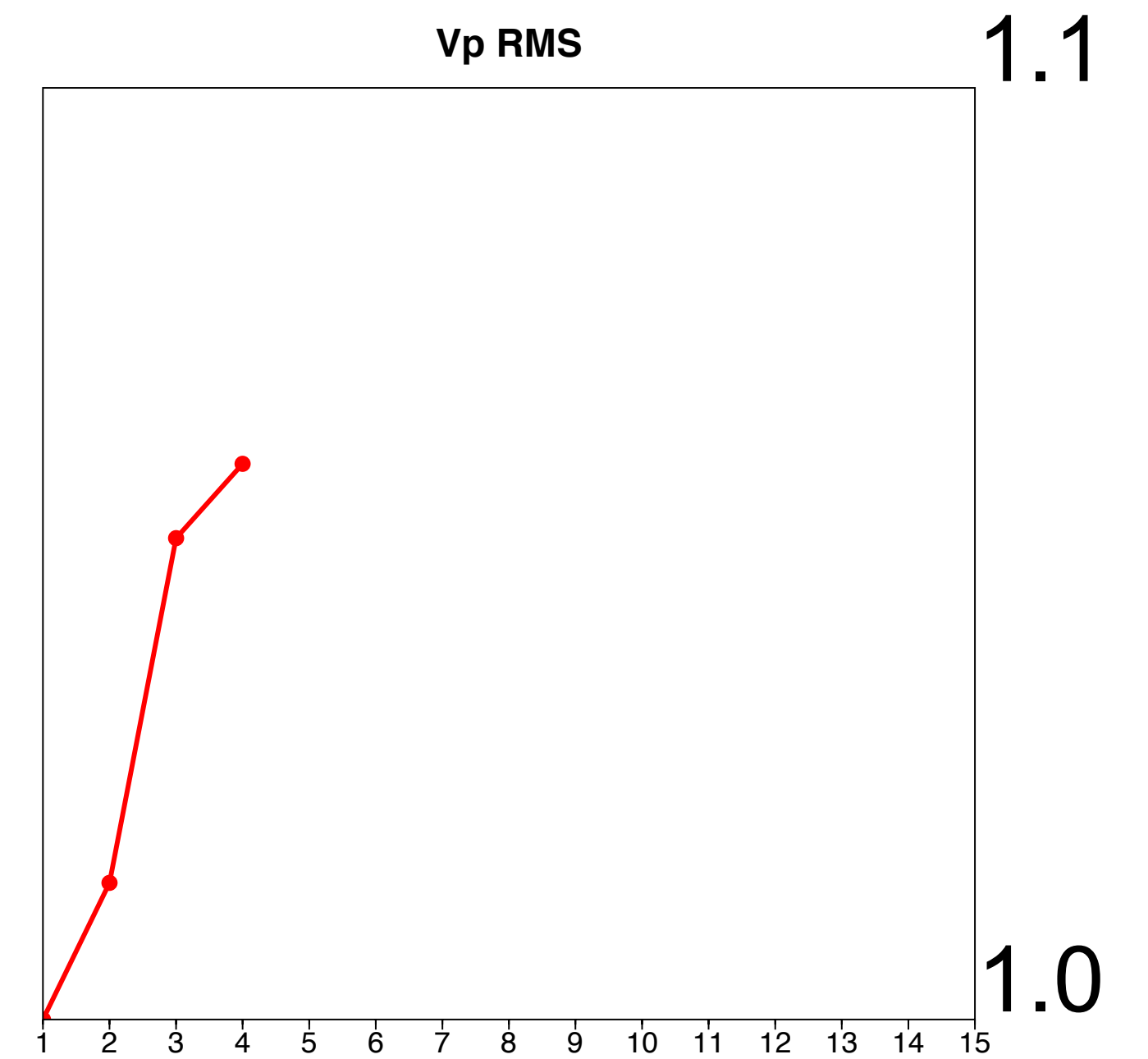
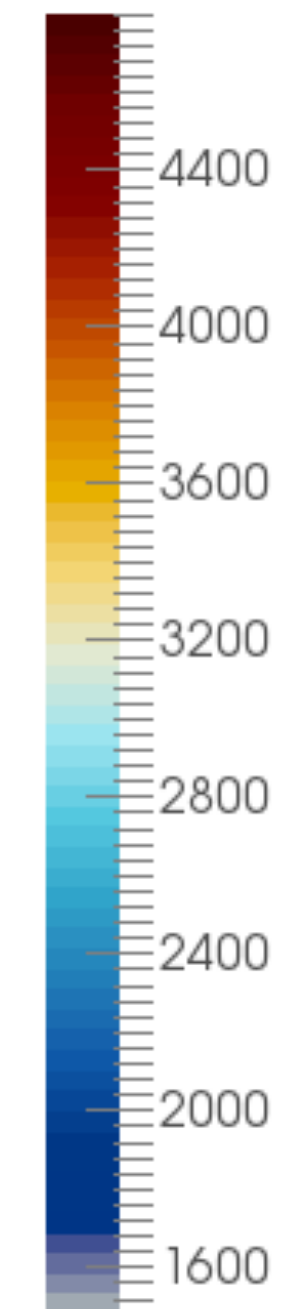
Vp (m/s)



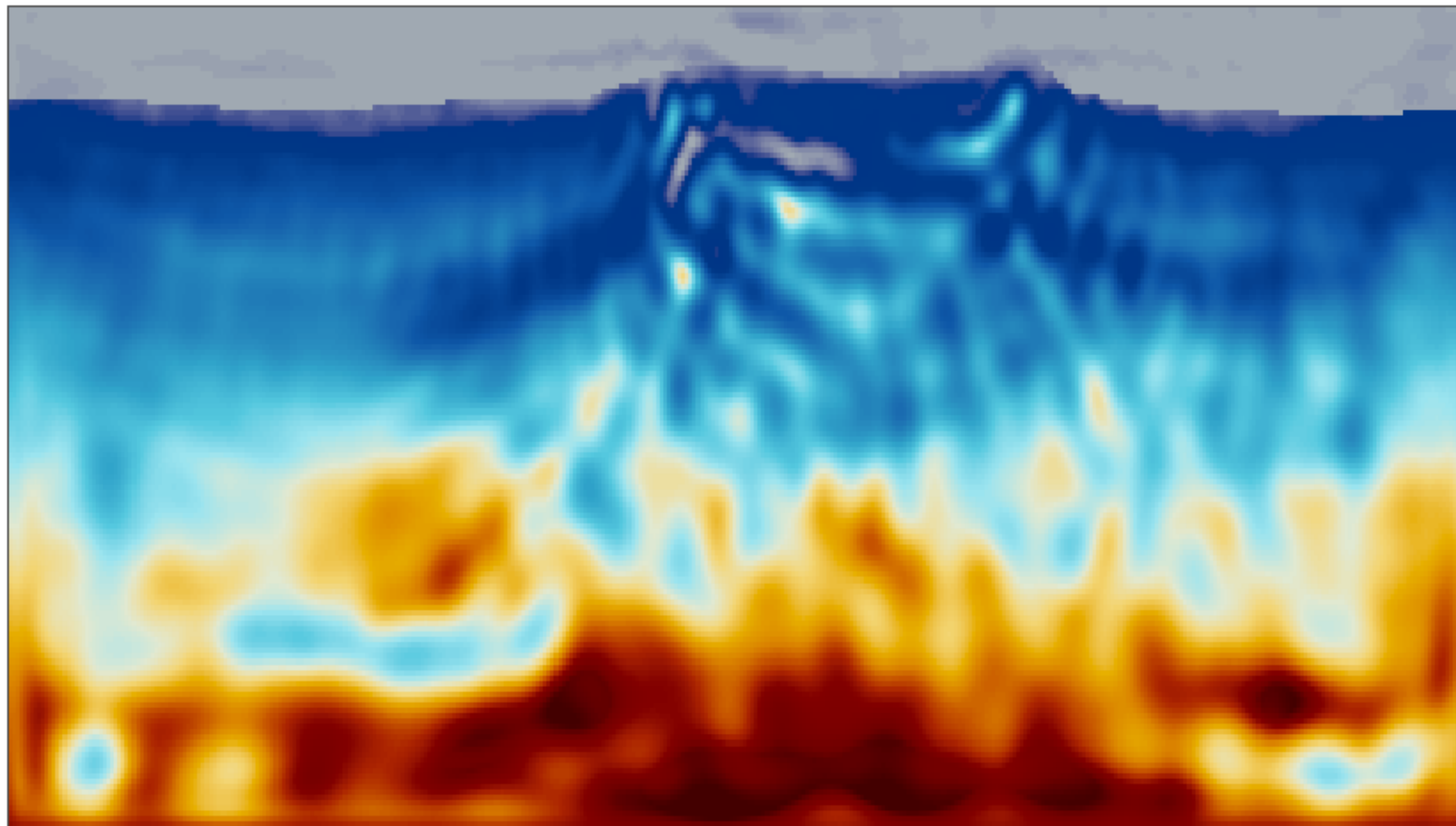
Adjoint-state – w/o constraints



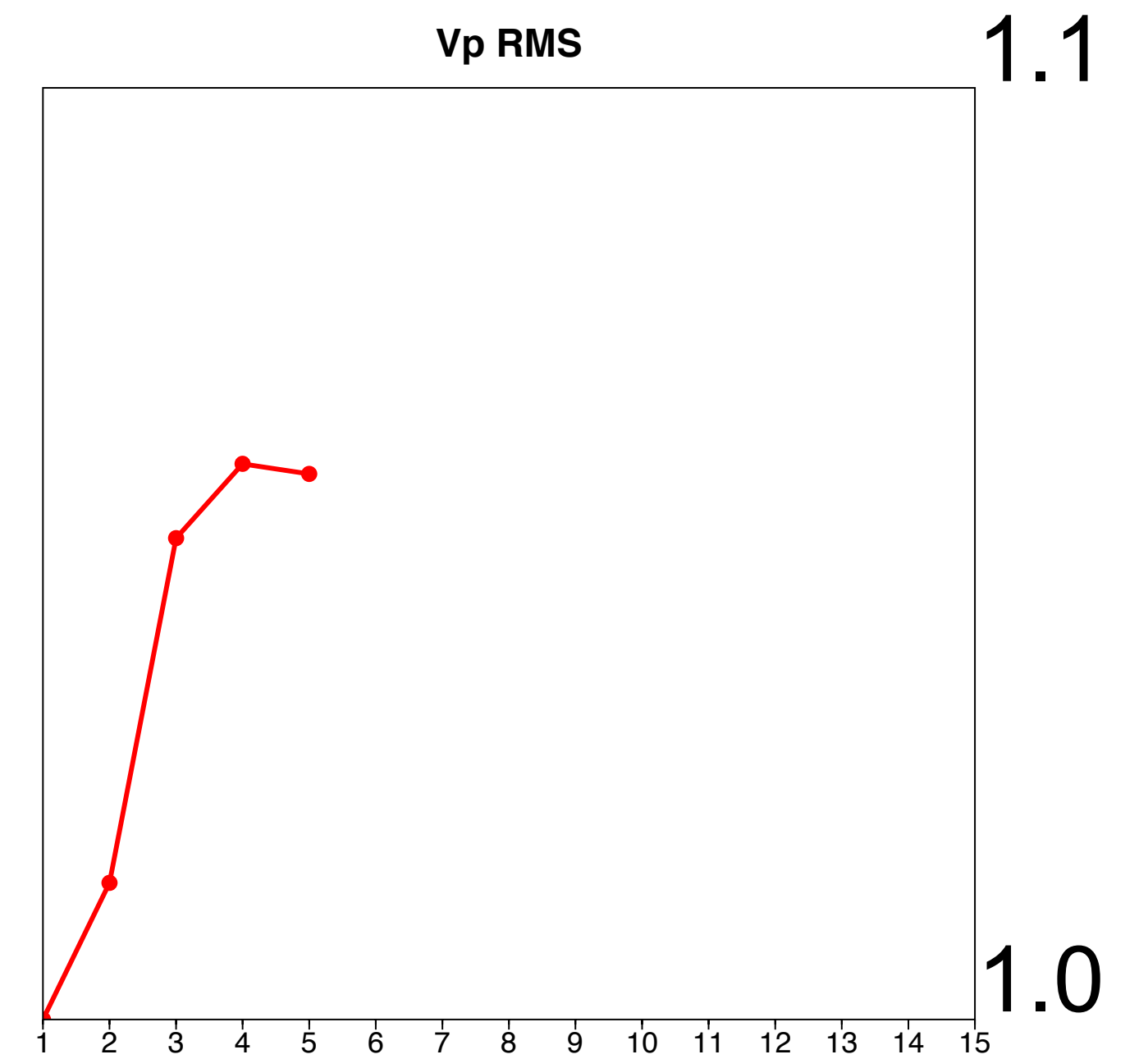
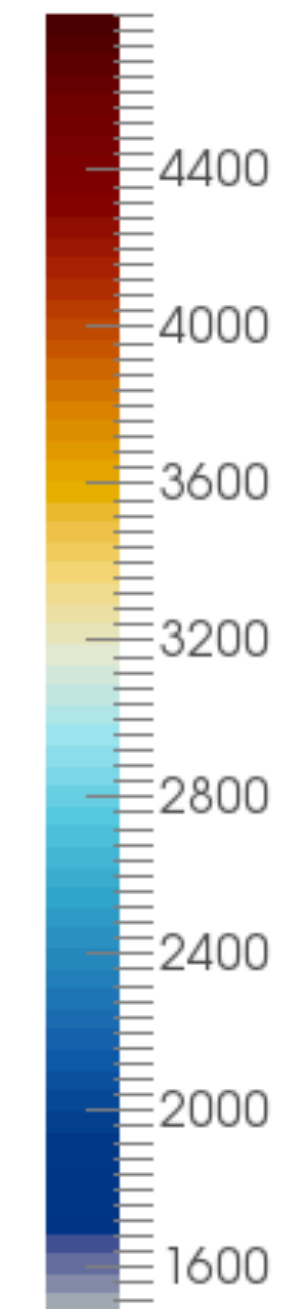
V_p (m/s)



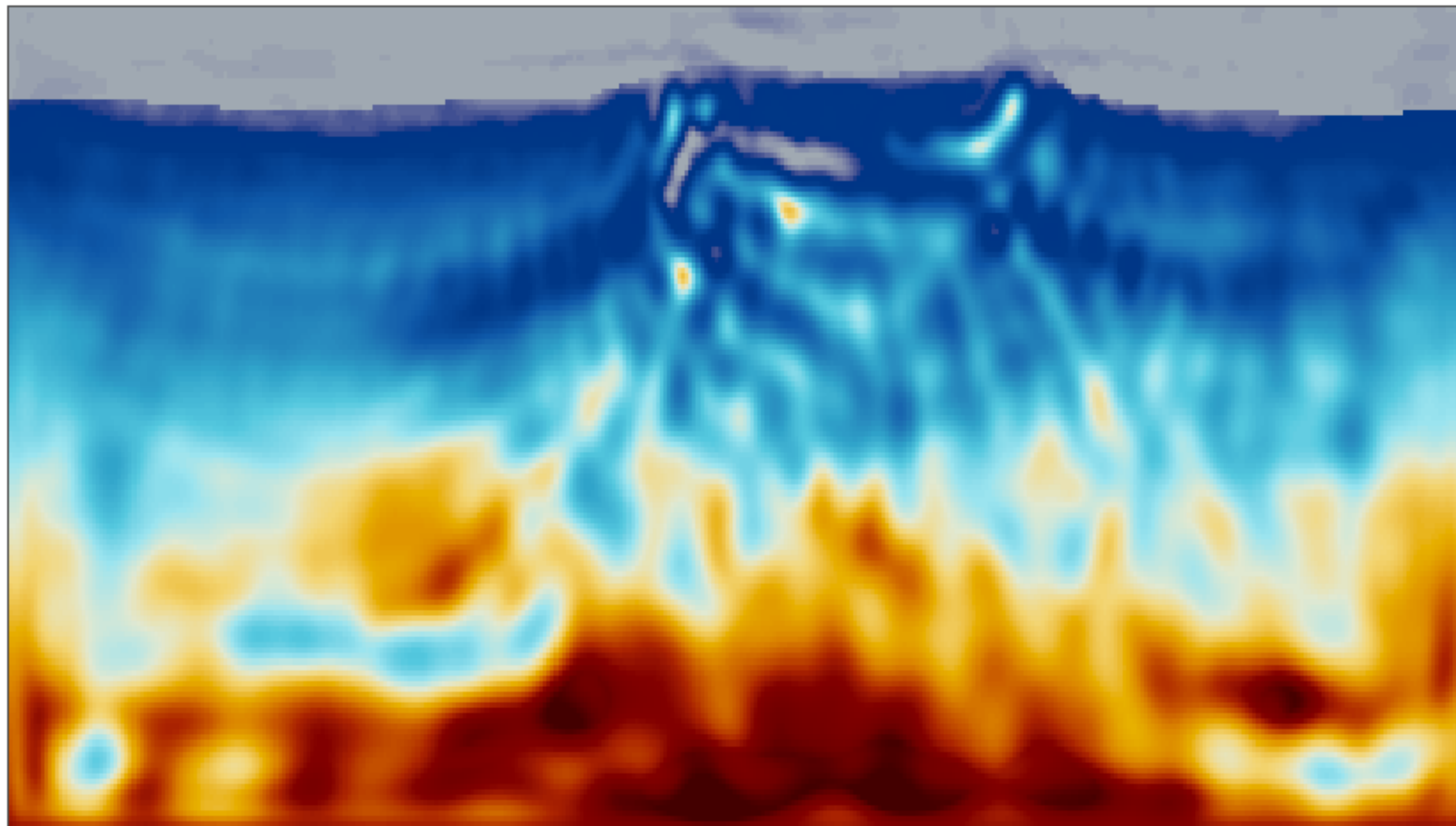
Adjoint-state – w/o constraints



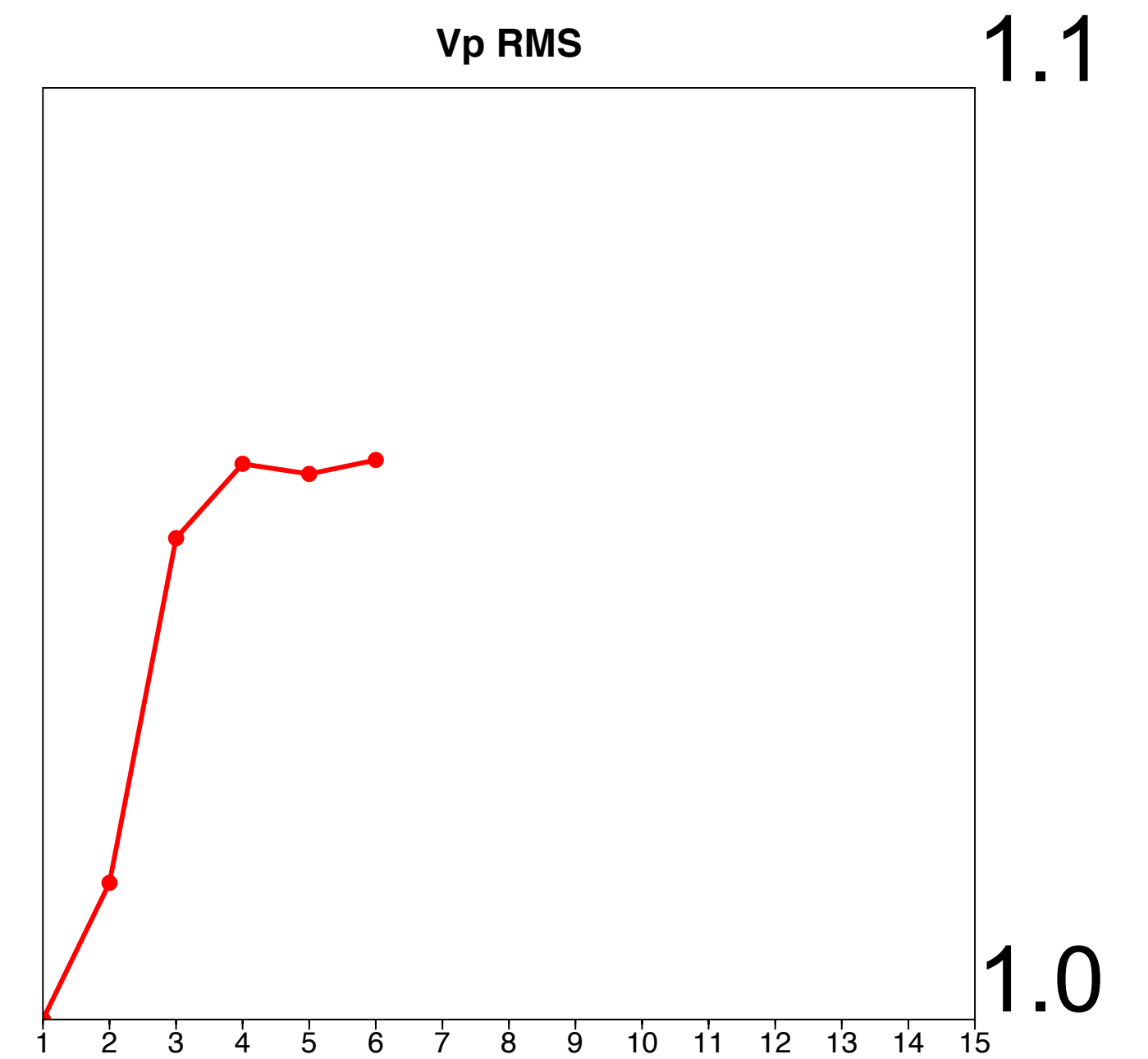
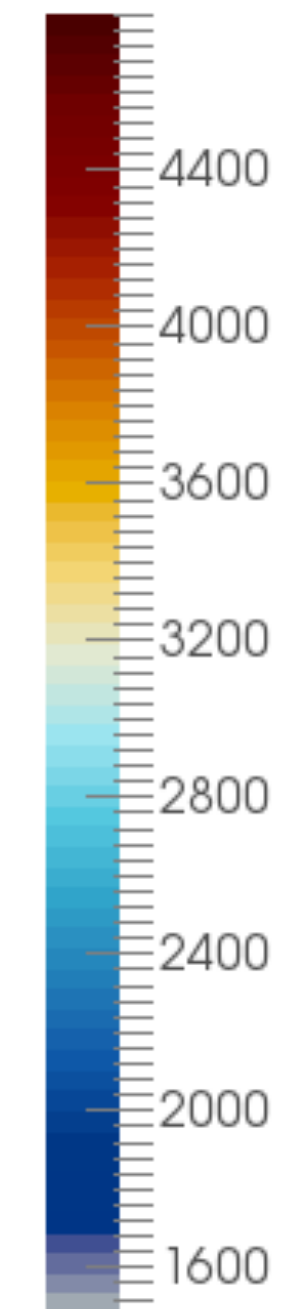
V_p (m/s)



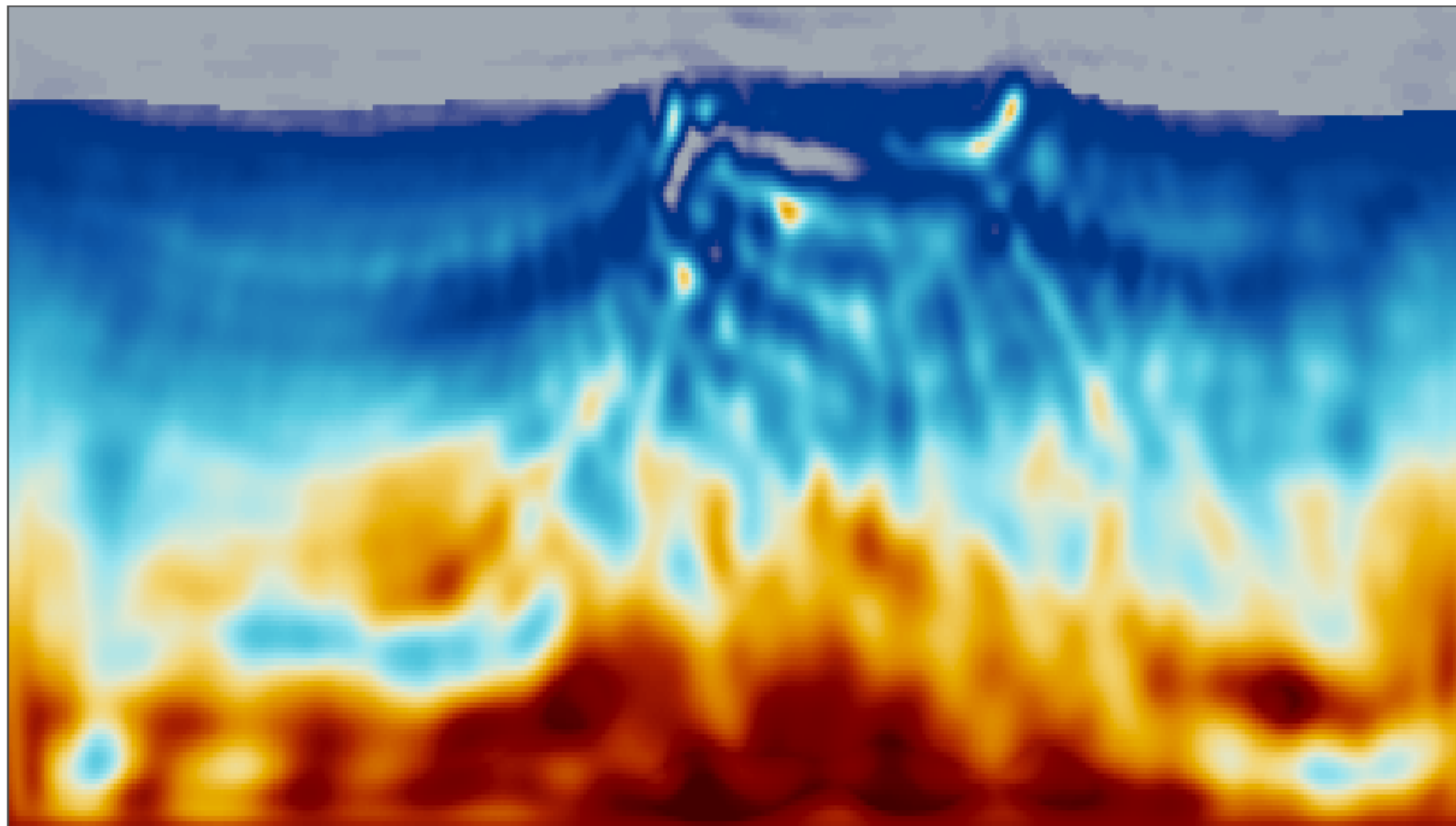
Adjoint-state – w/o constraints



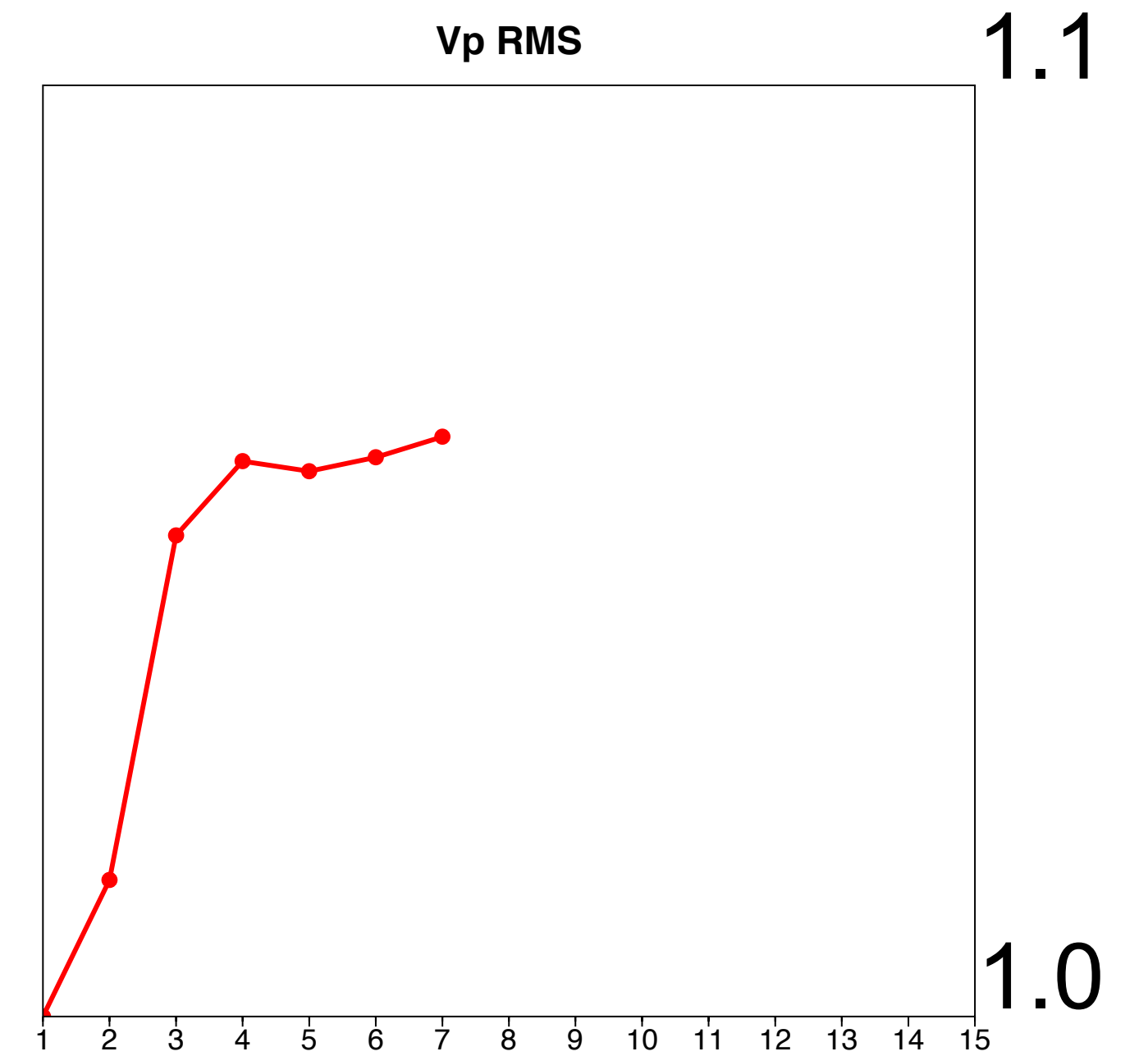
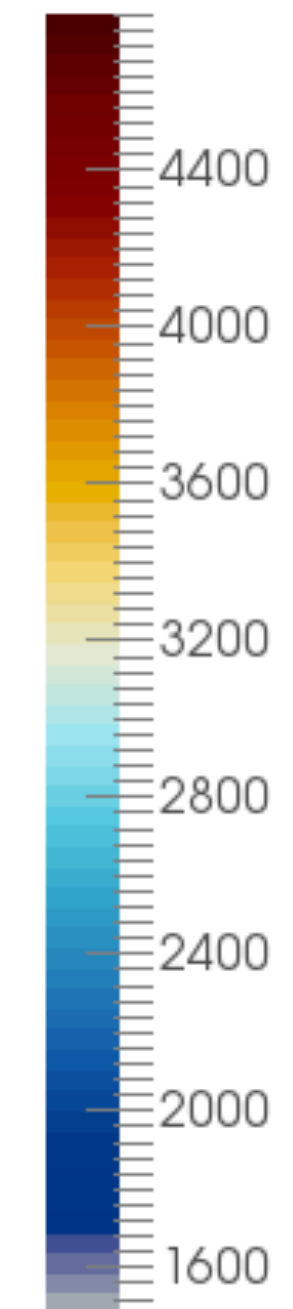
V_p (m/s)



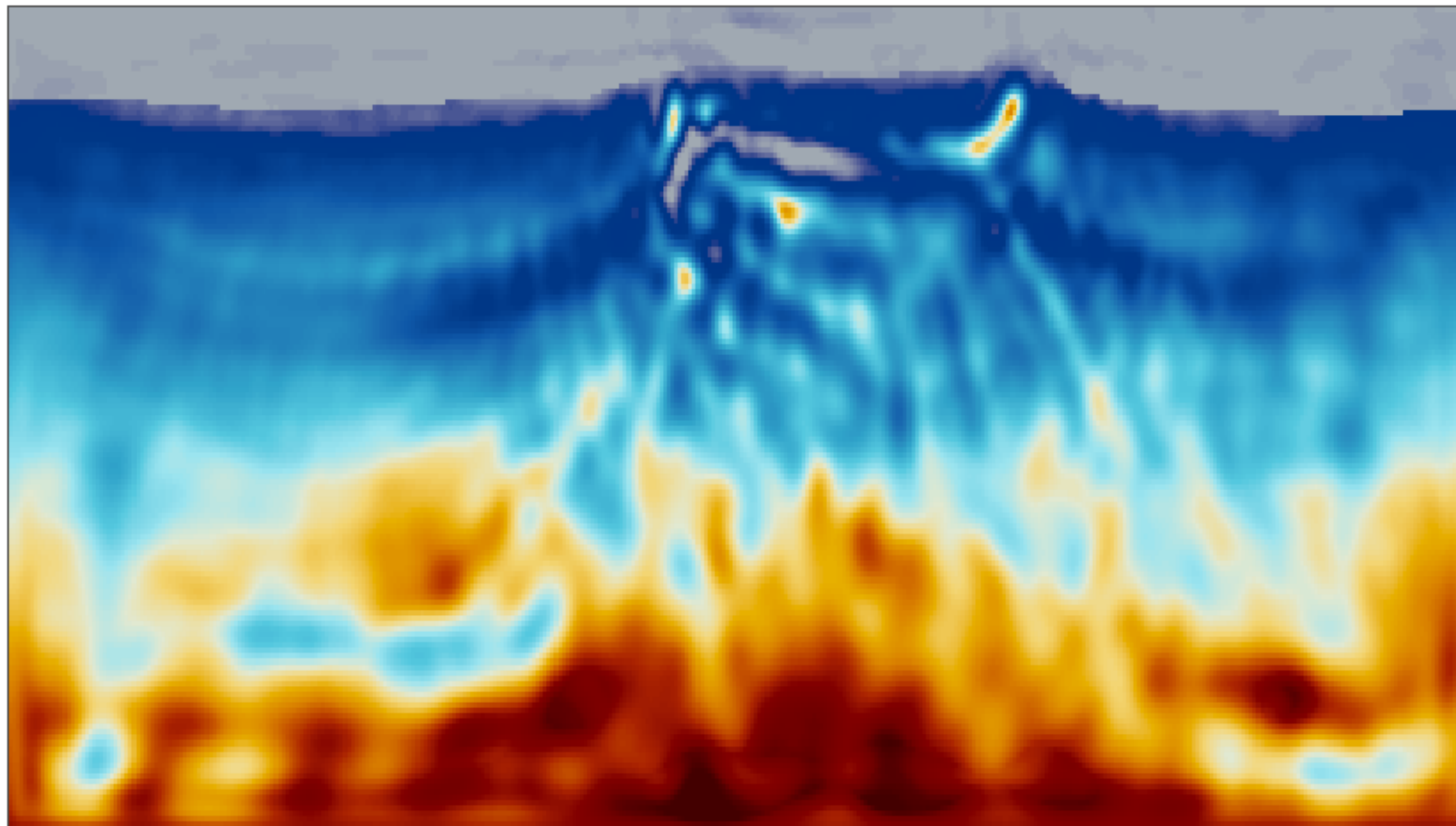
Adjoint-state – w/o constraints



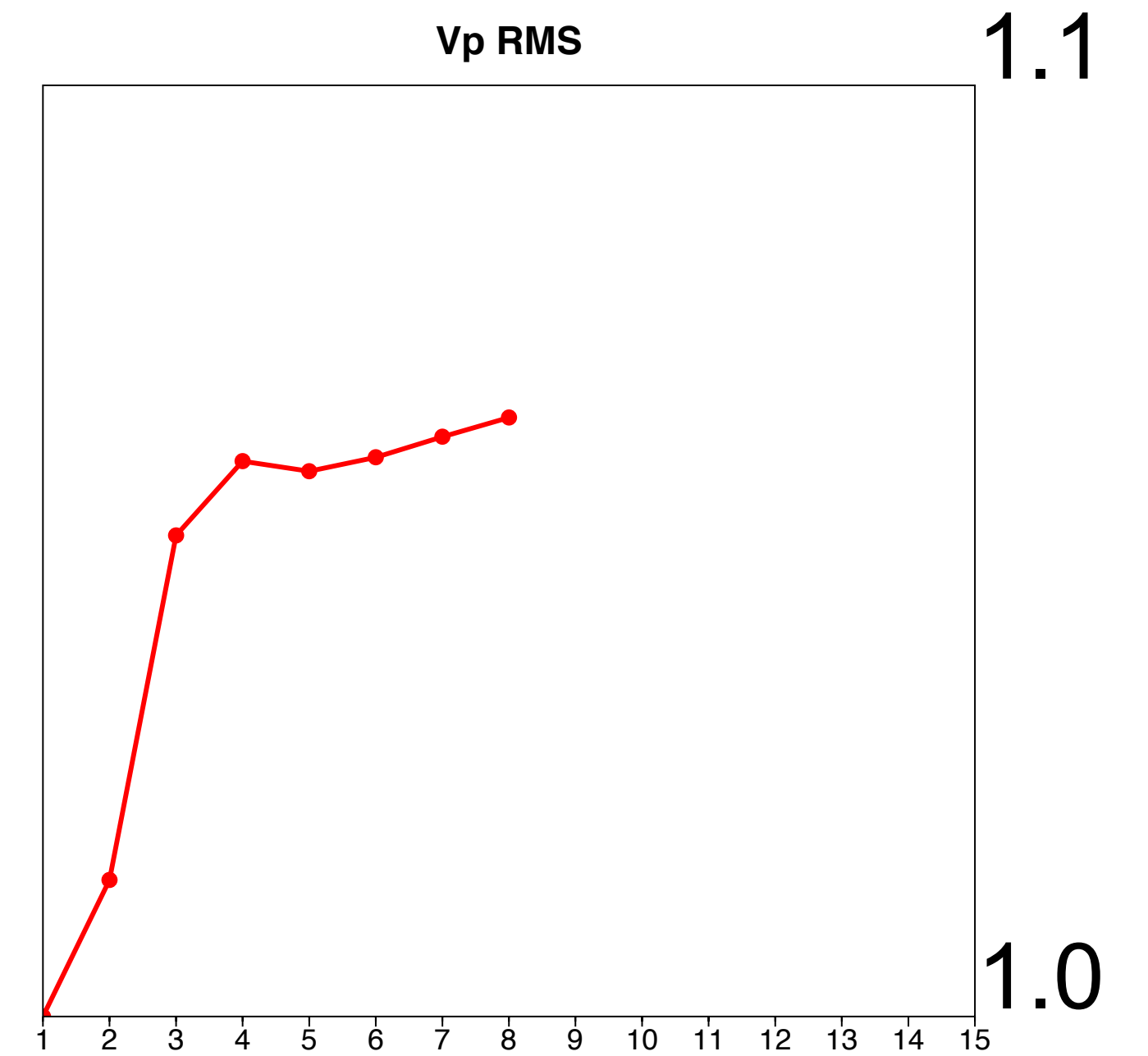
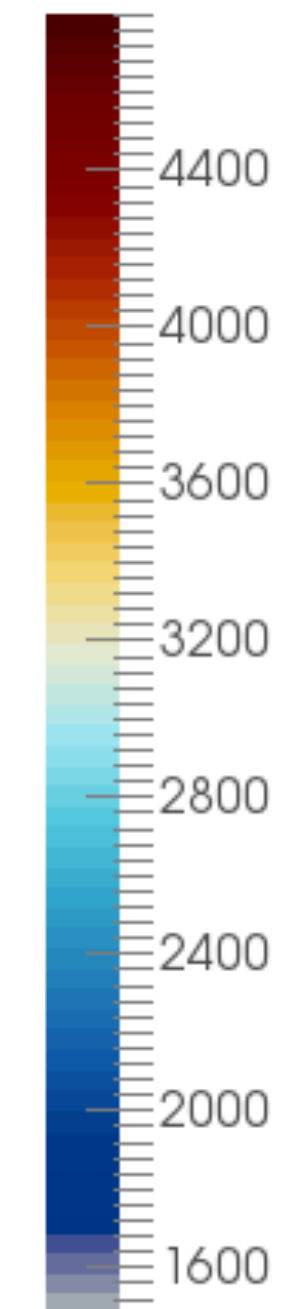
V_p (m/s)



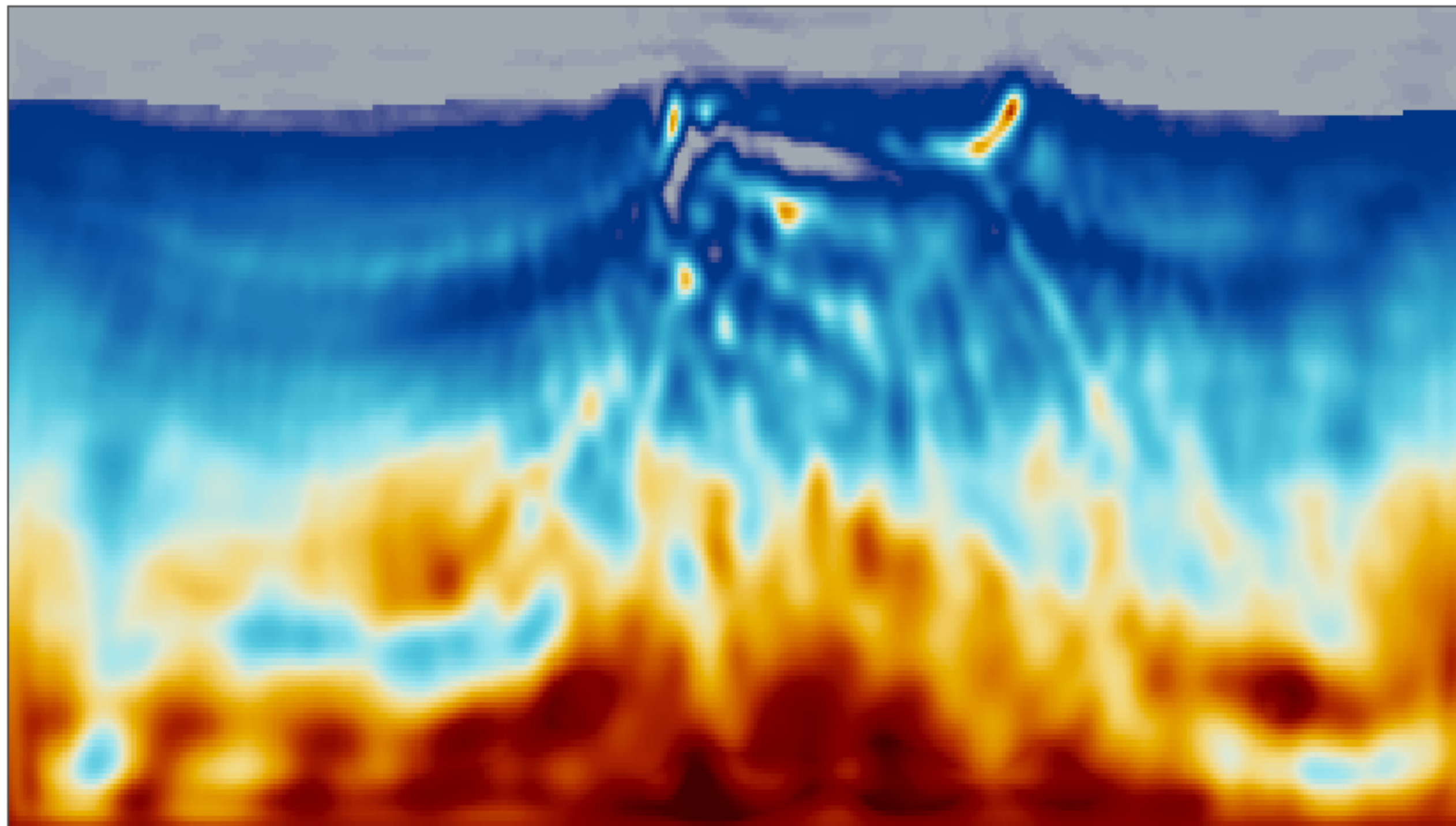
Adjoint-state – w/o constraints



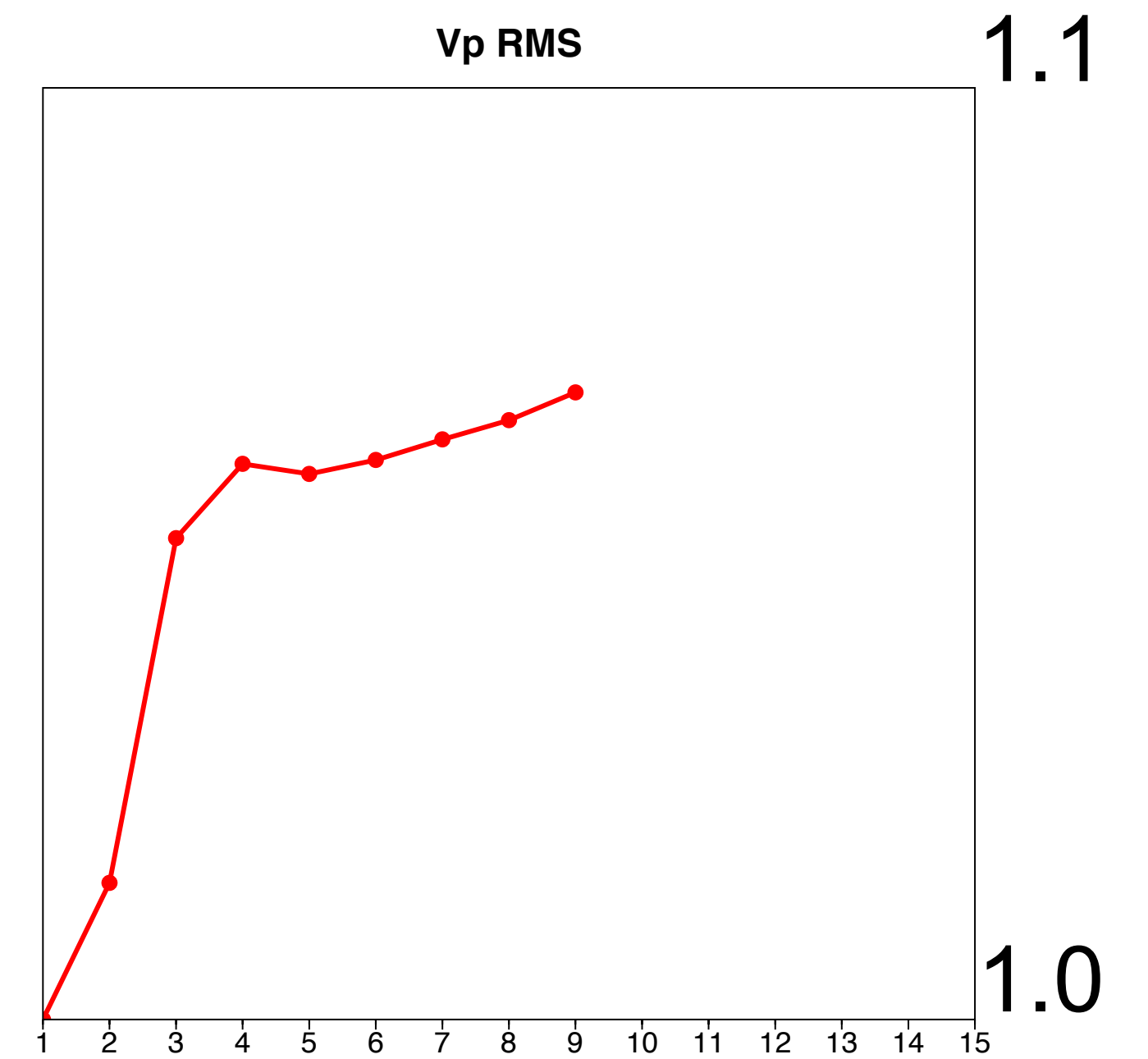
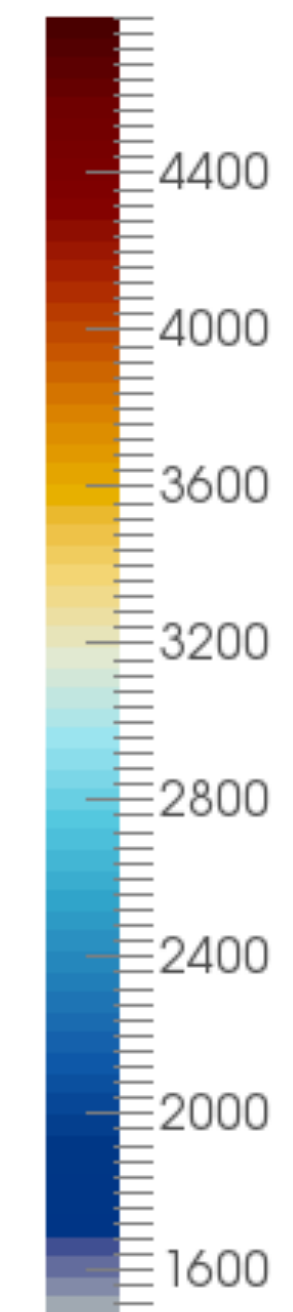
Vp (m/s)



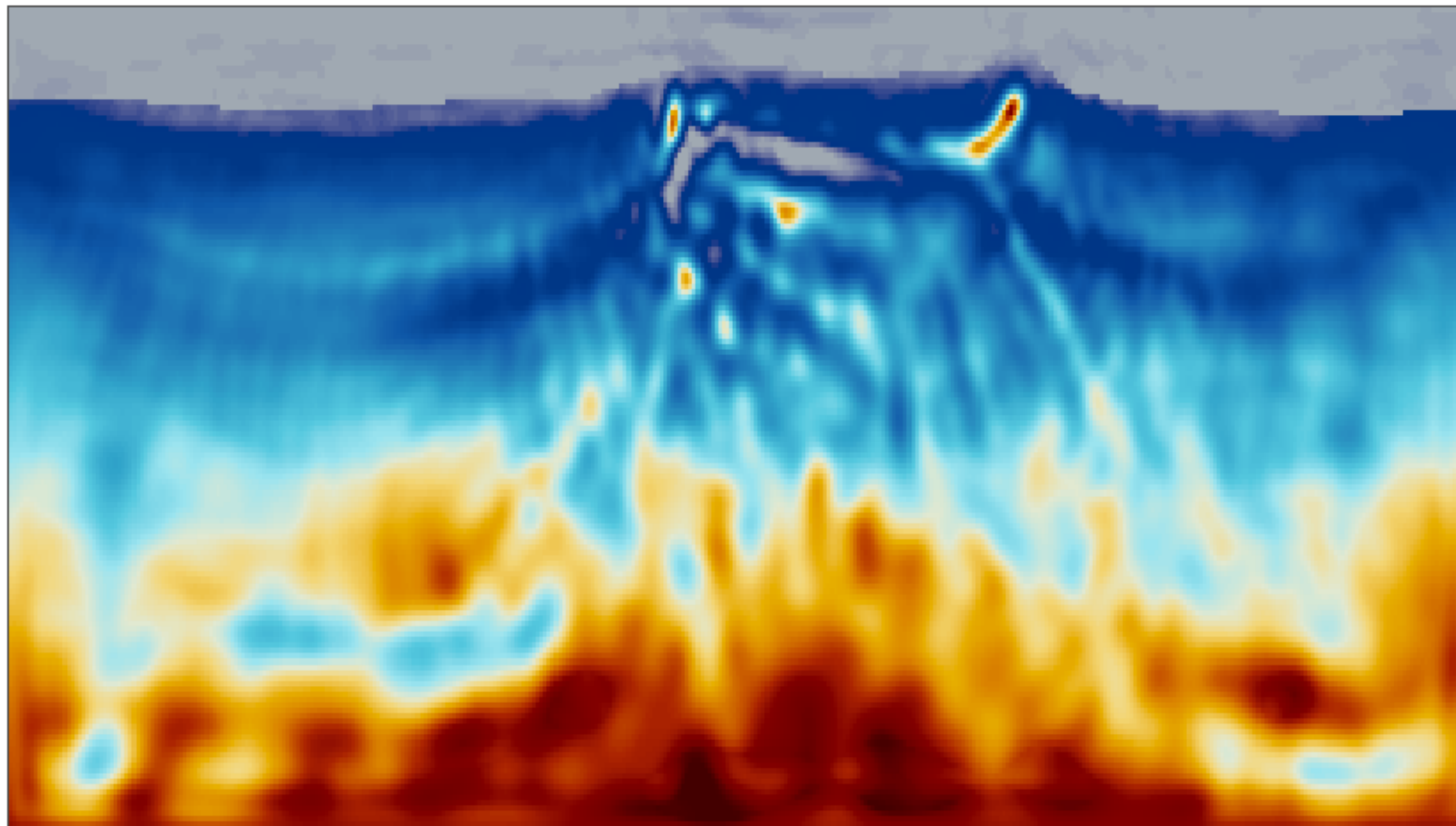
Adjoint-state – w/o constraints



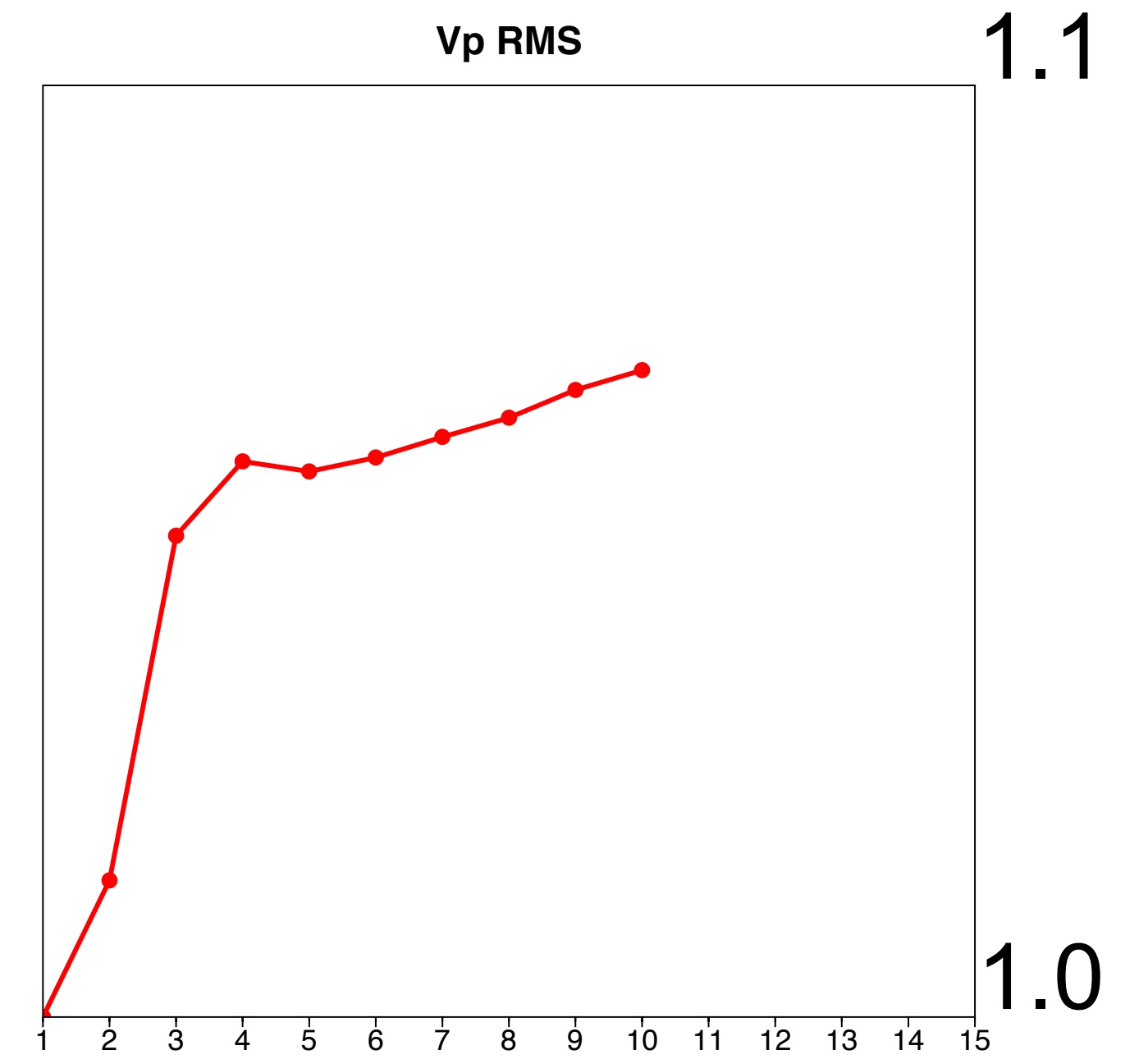
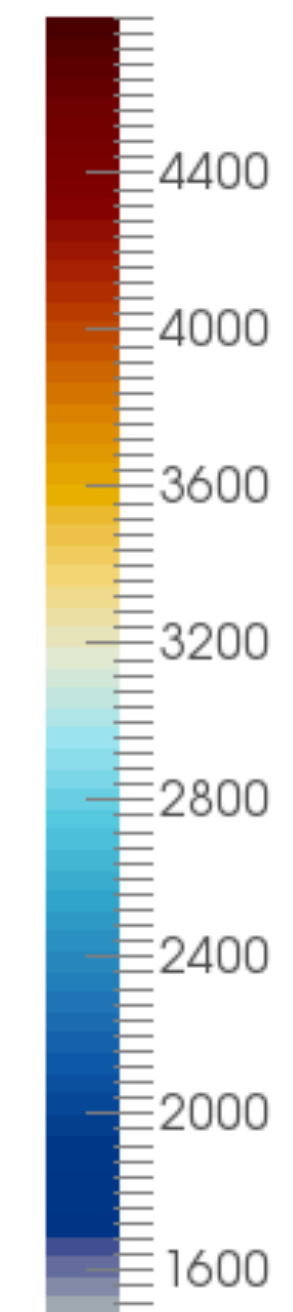
V_p (m/s)



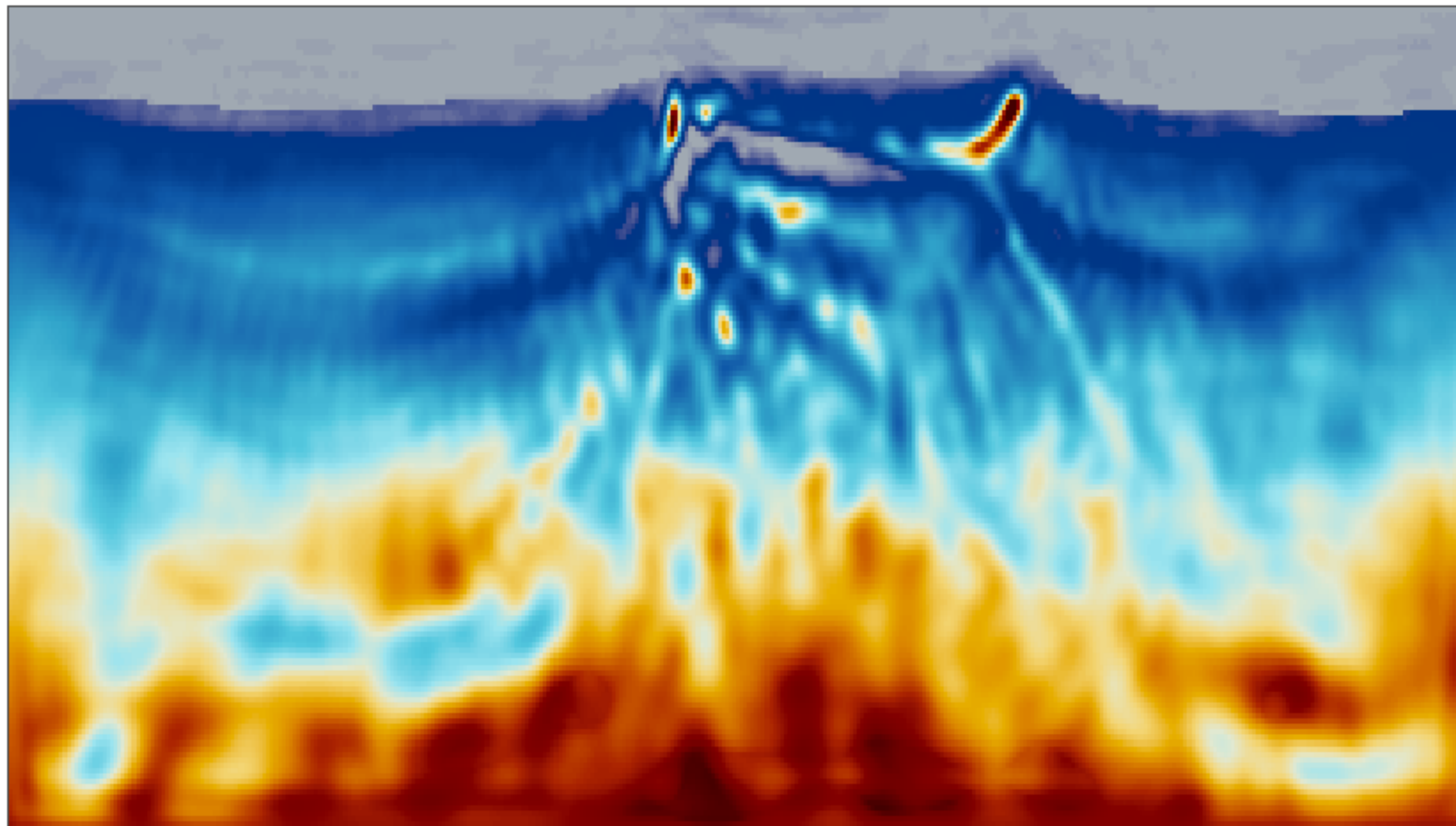
Adjoint-state – w/o constraints



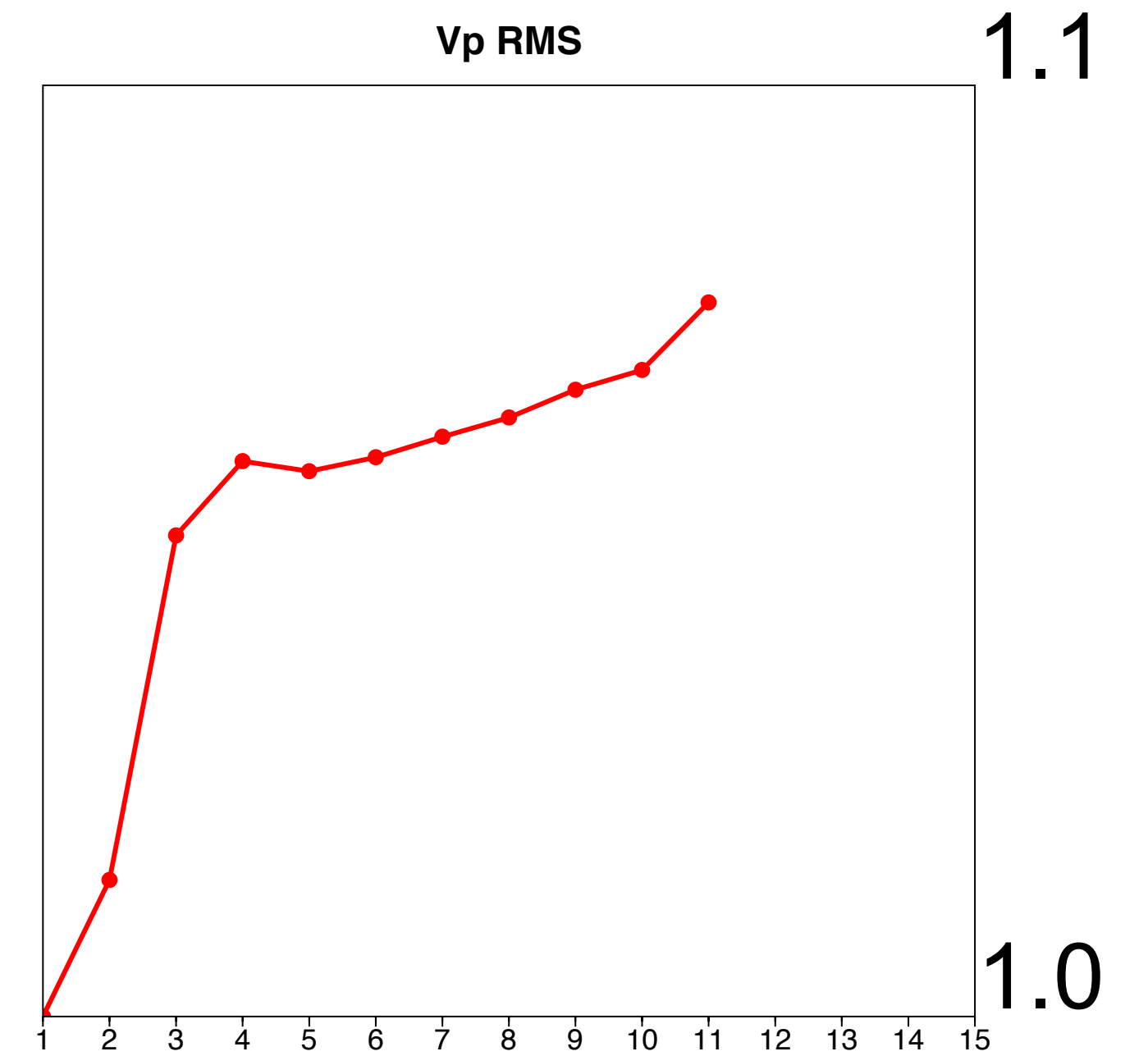
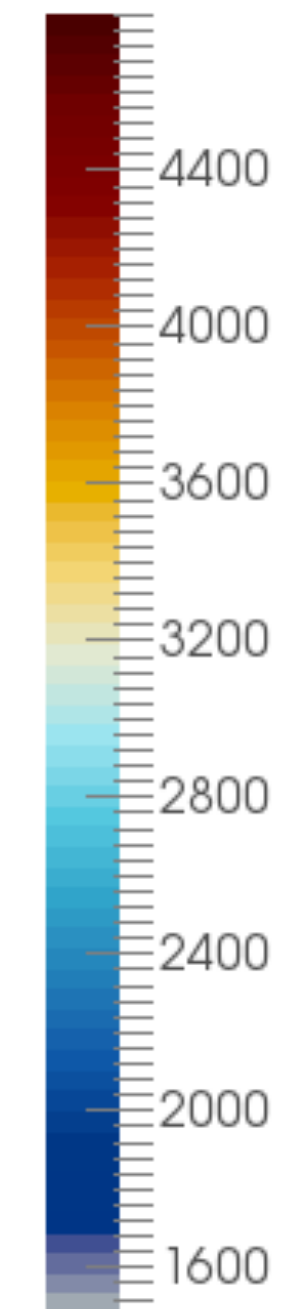
V_p (m/s)



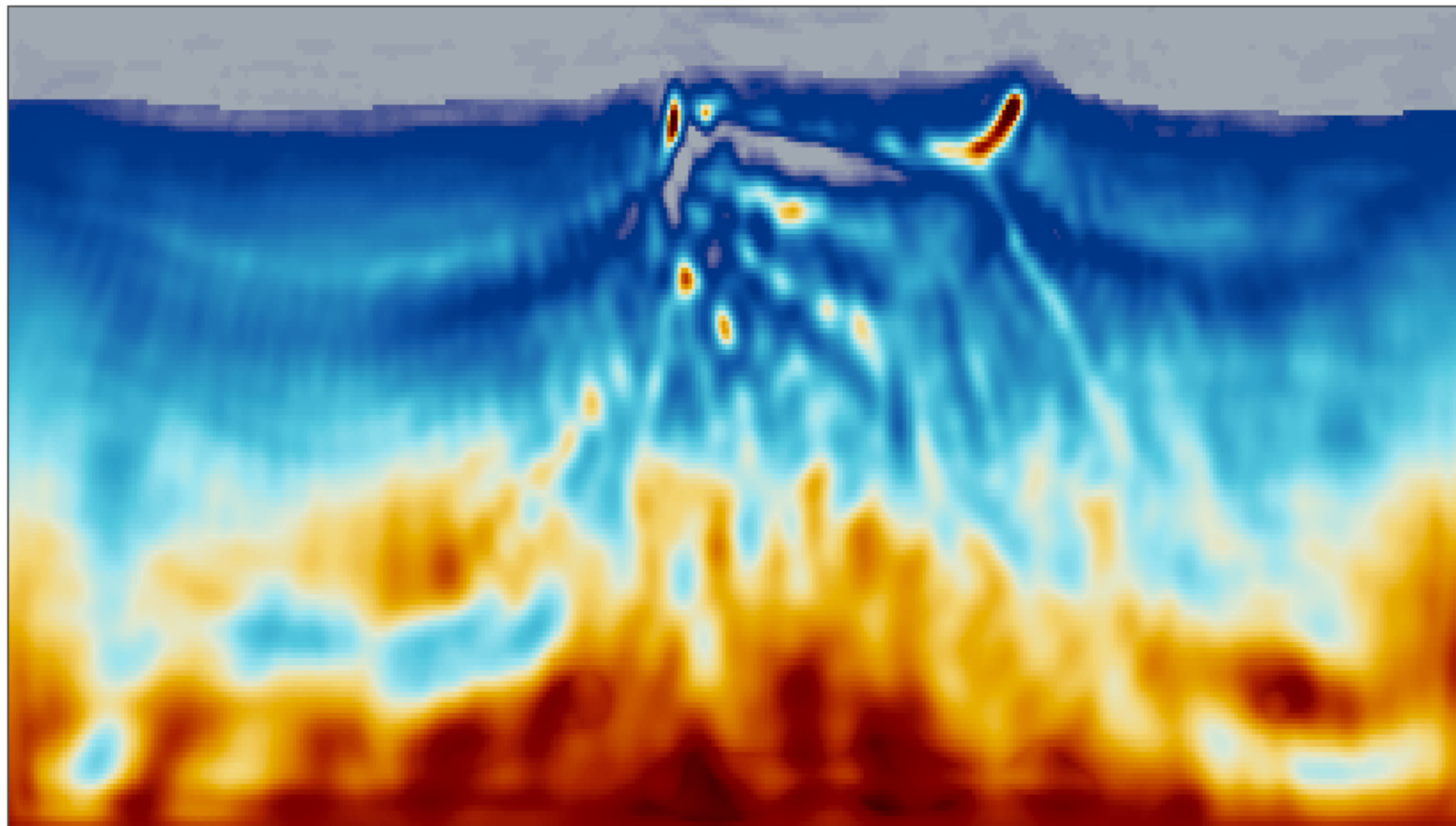
Adjoint-state – w/o constraints



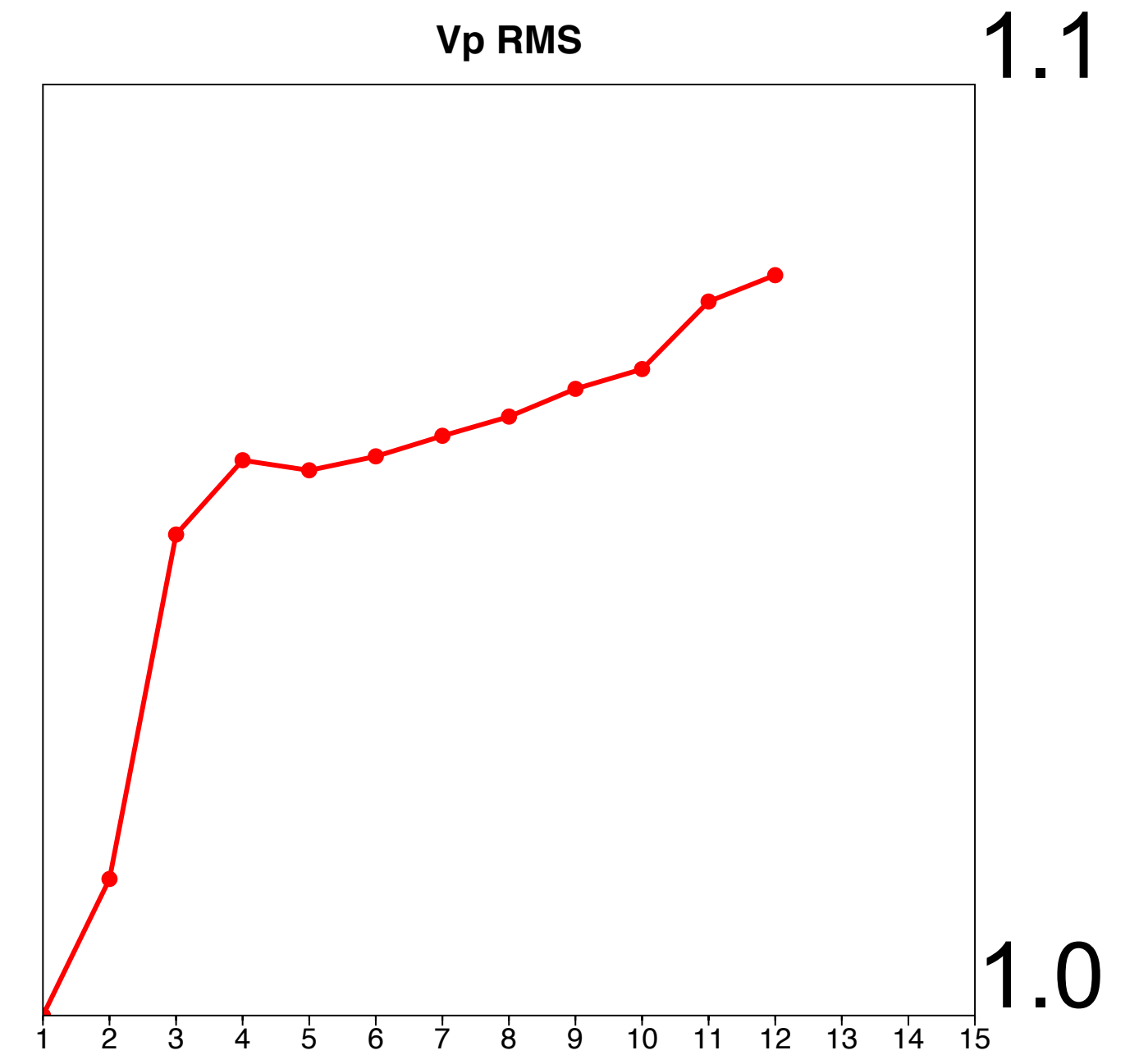
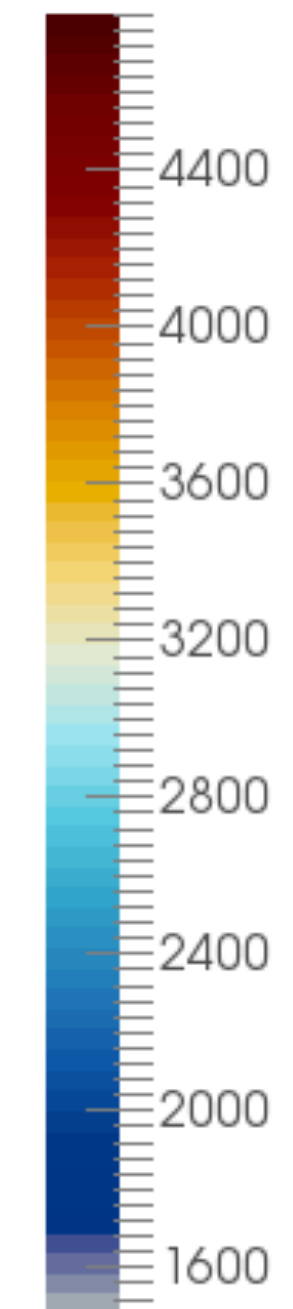
V_p (m/s)



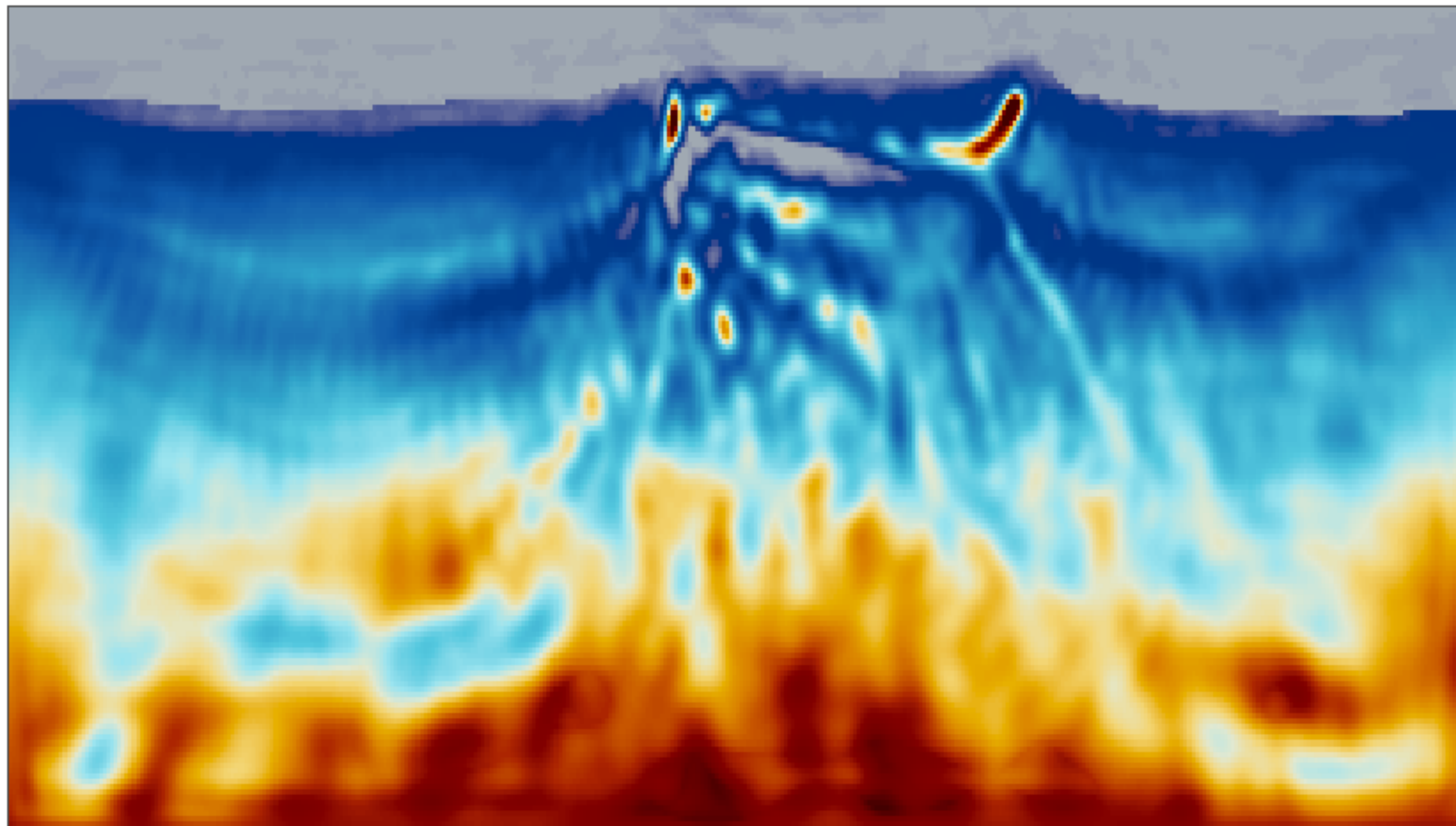
Adjoint-state – w/o constraints



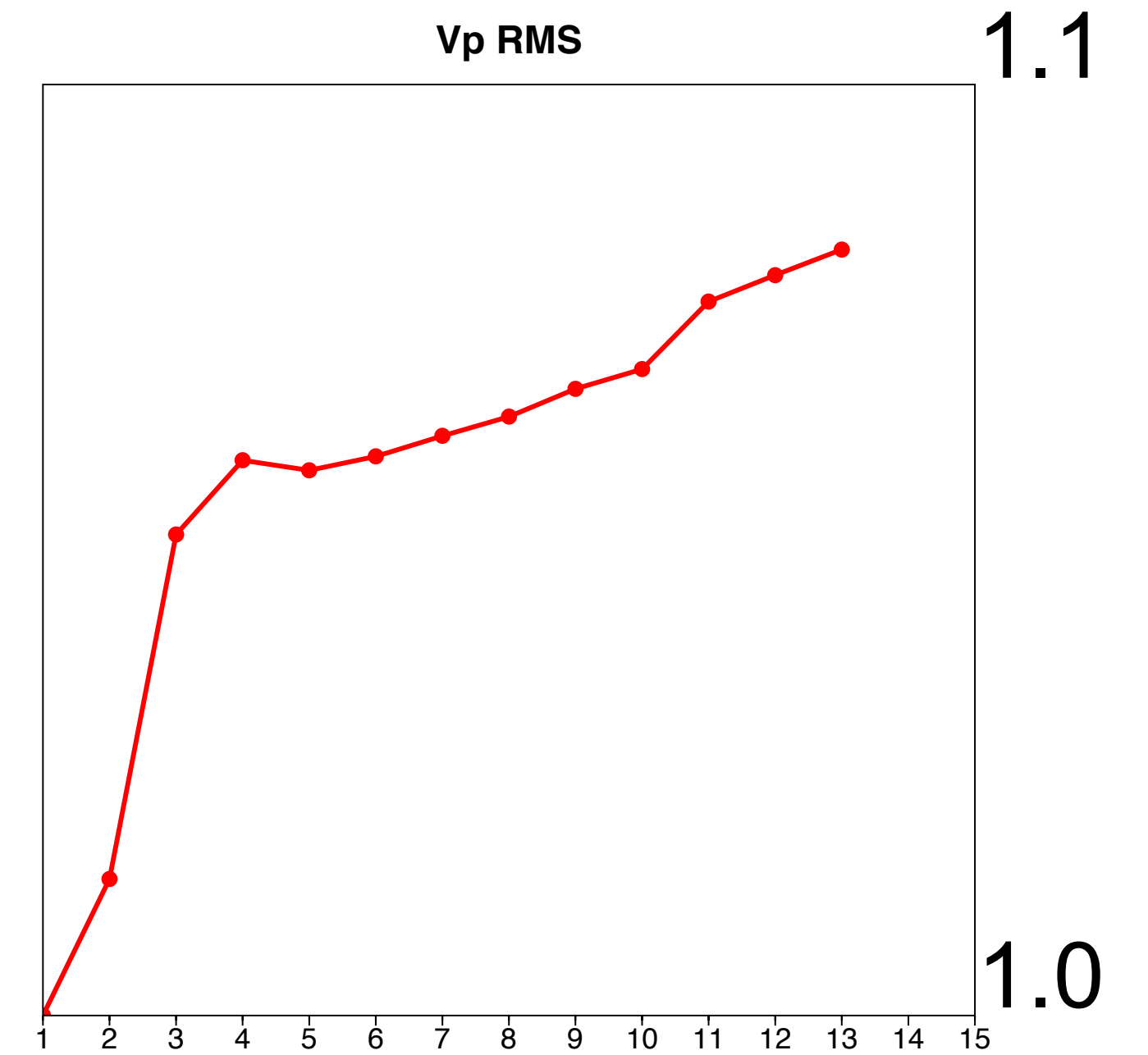
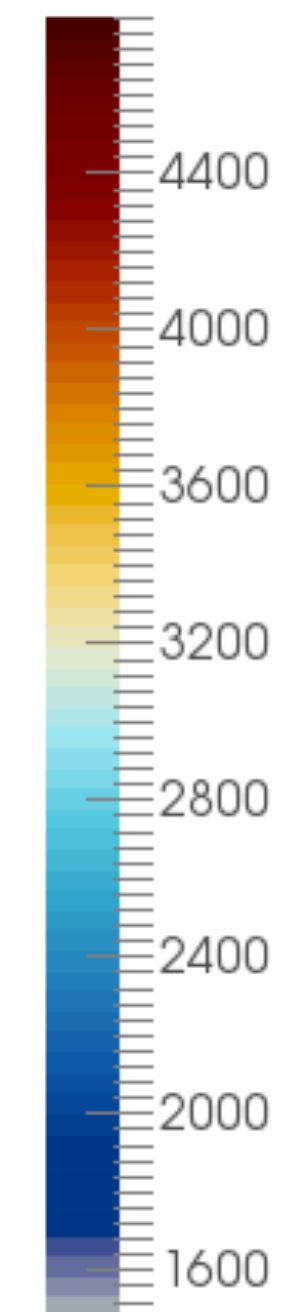
Vp (m/s)



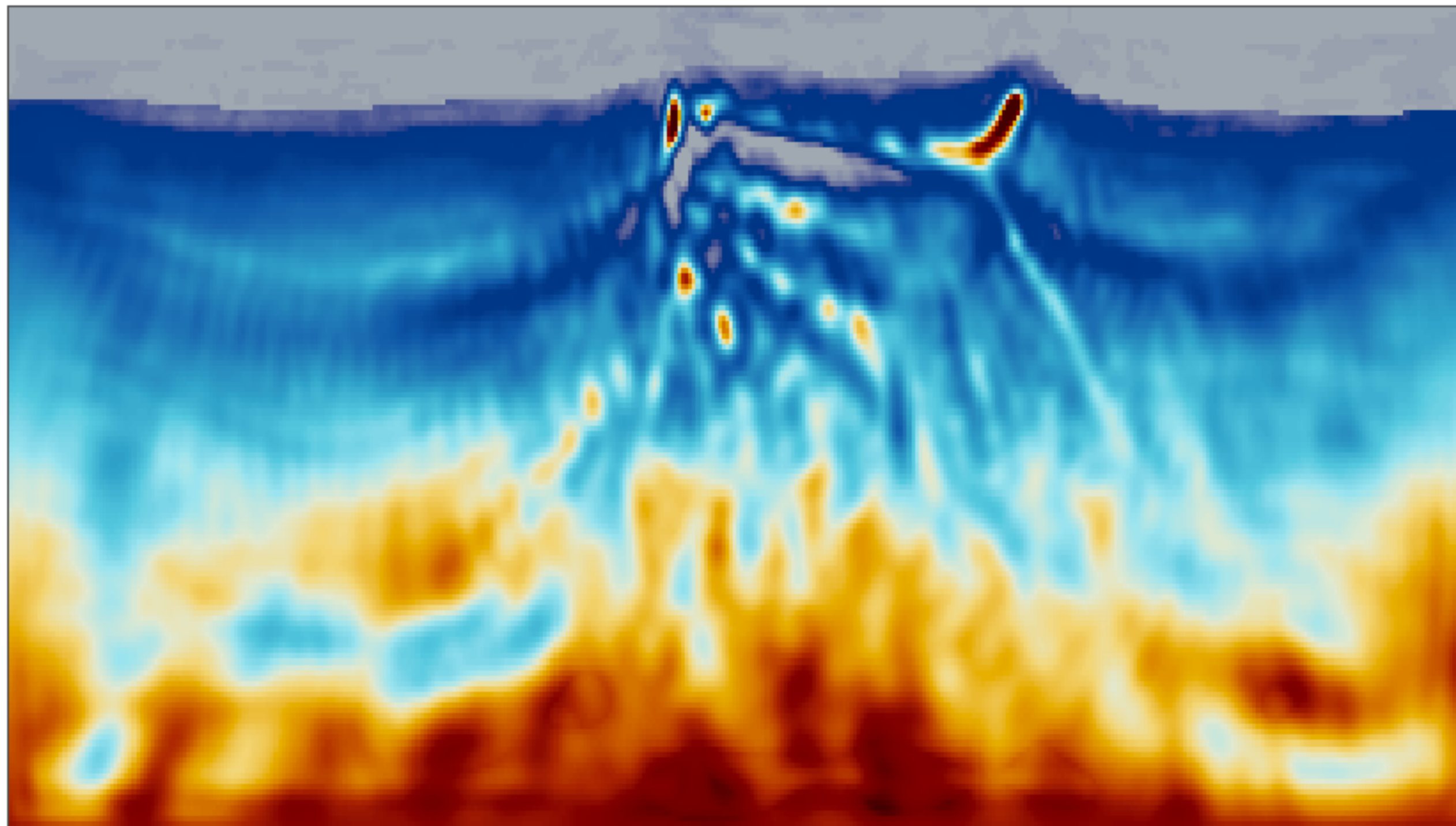
Adjoint-state – w/o constraints



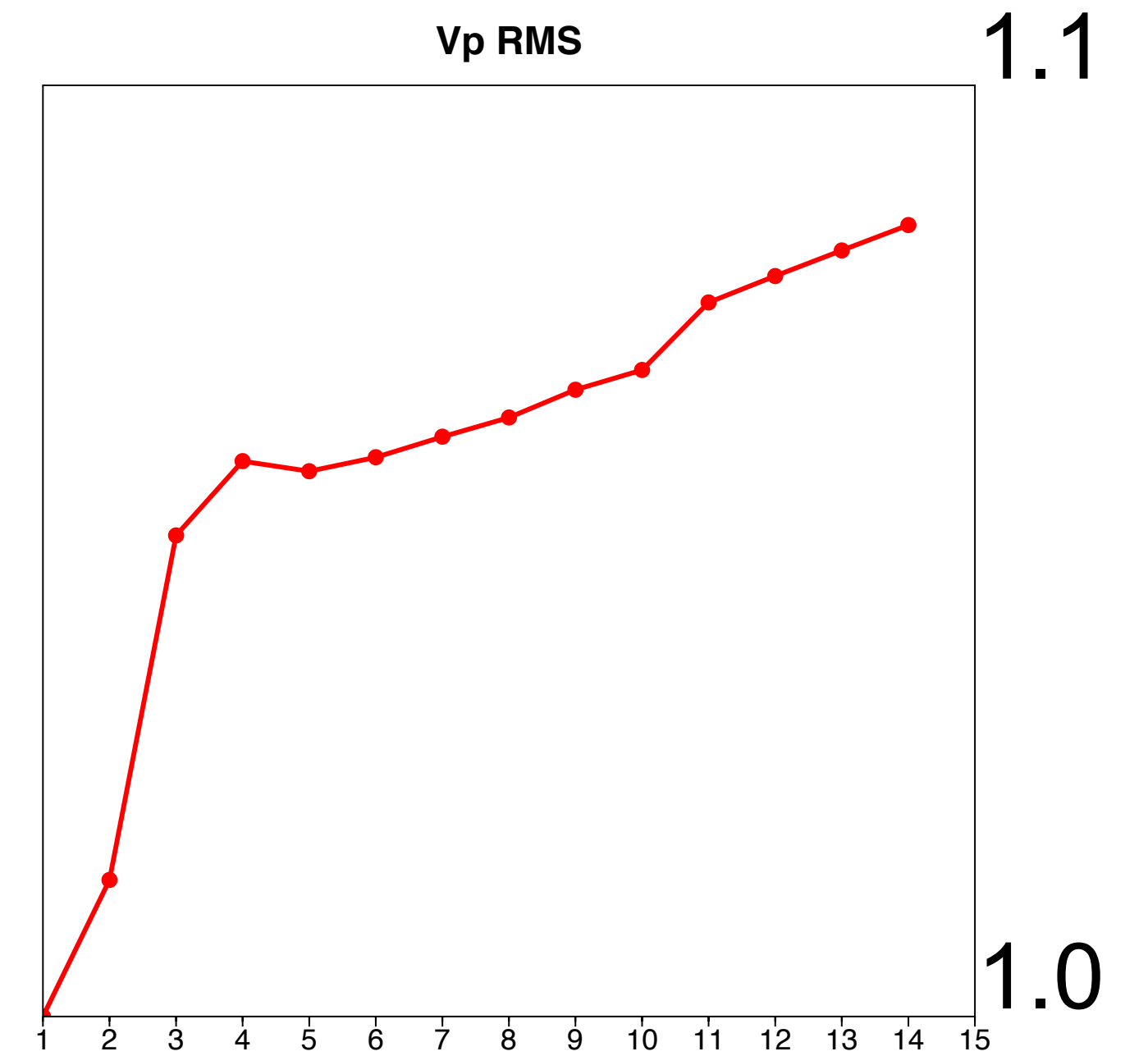
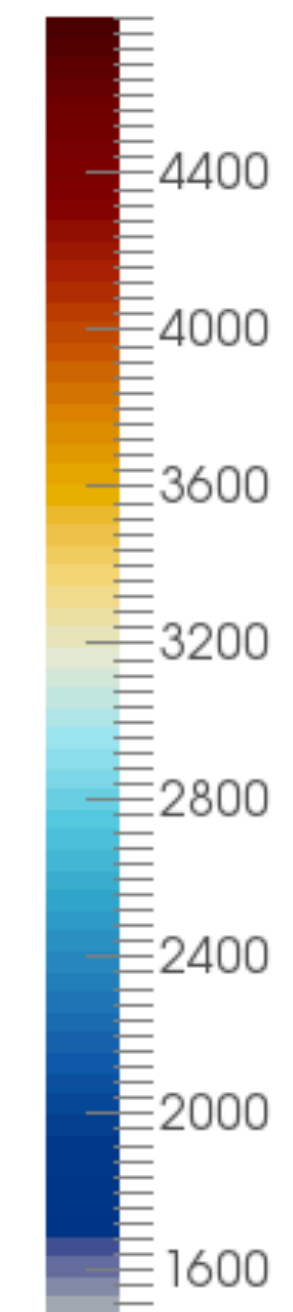
Vp (m/s)



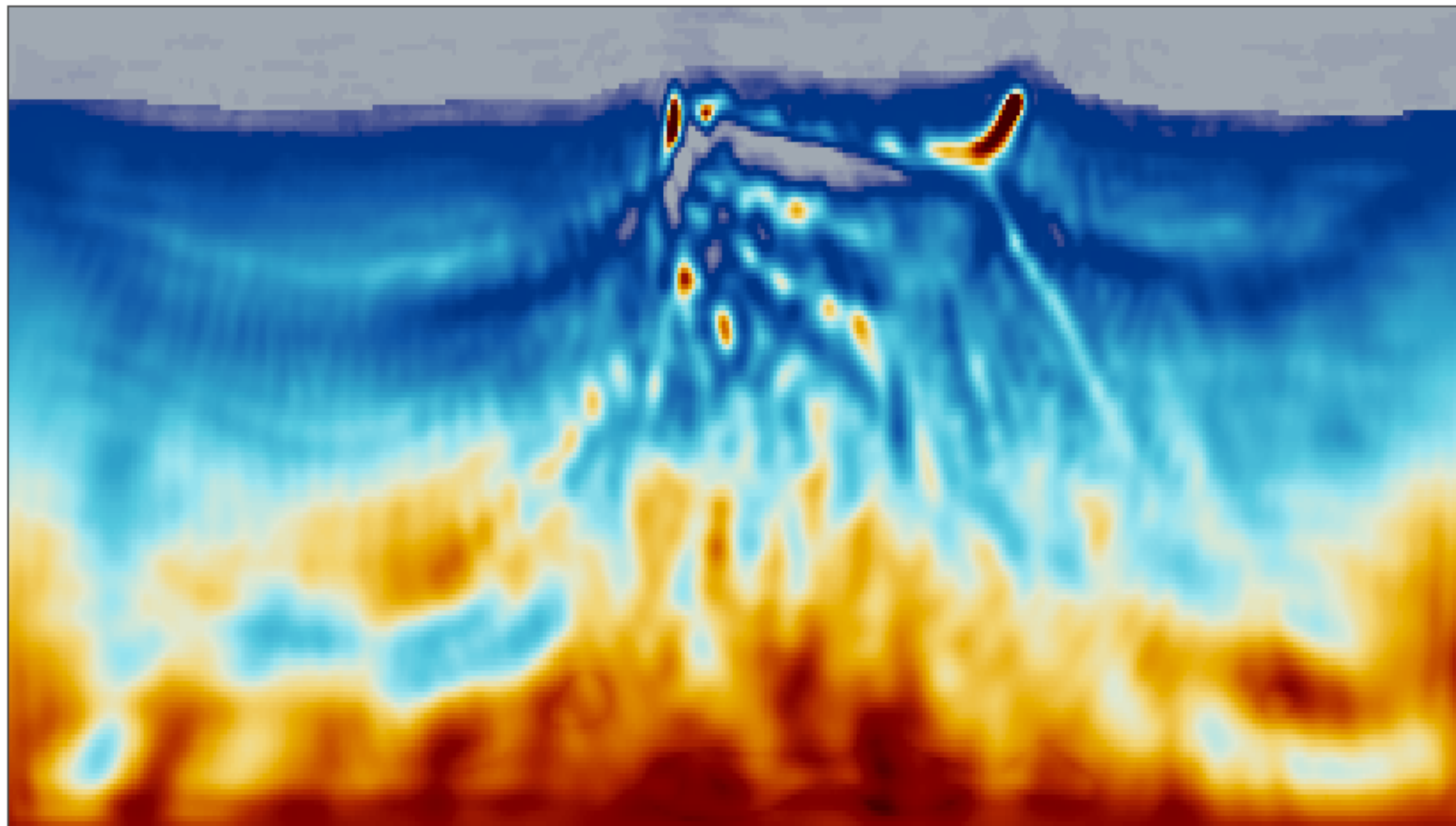
Adjoint-state – w/o constraints



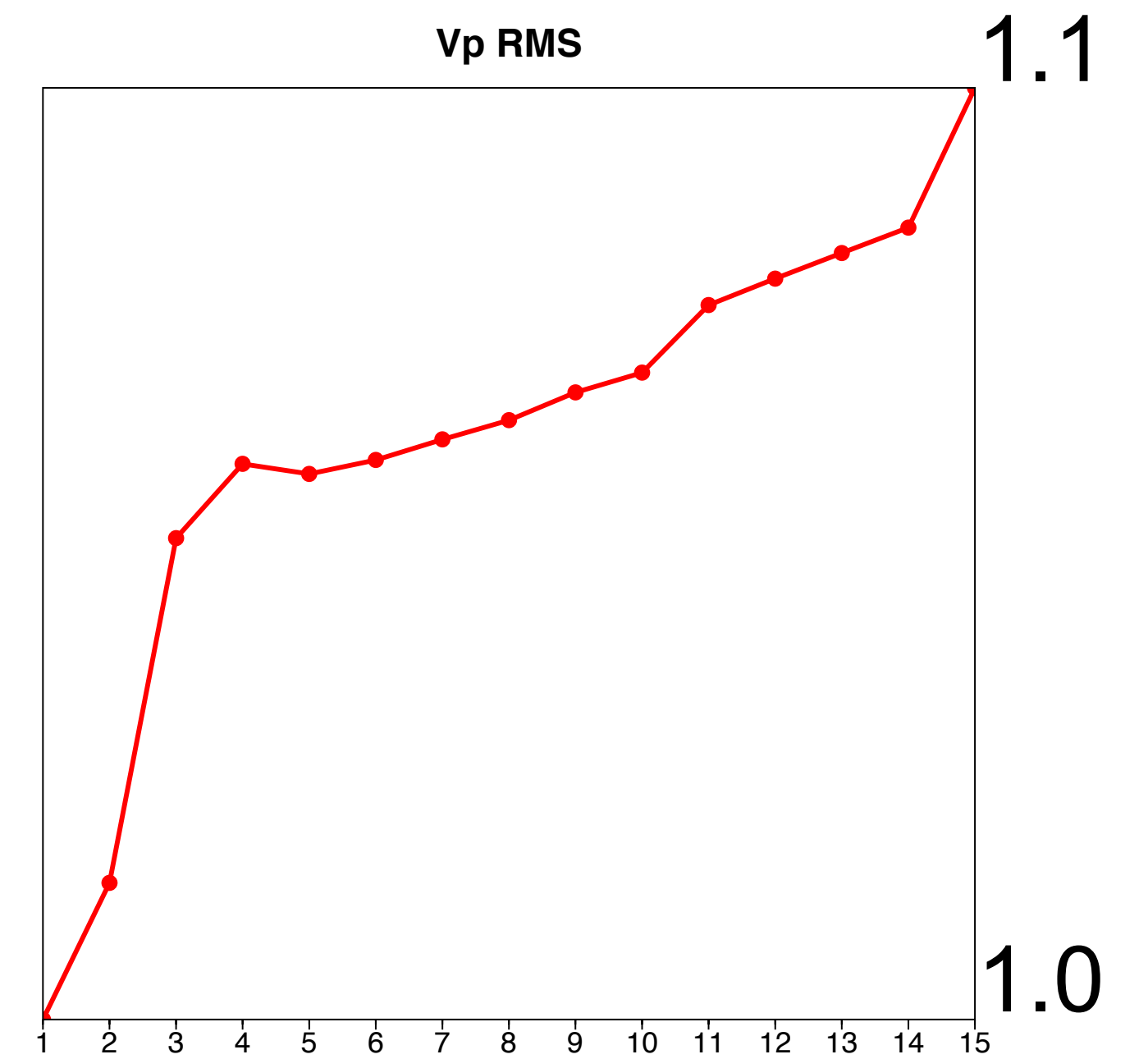
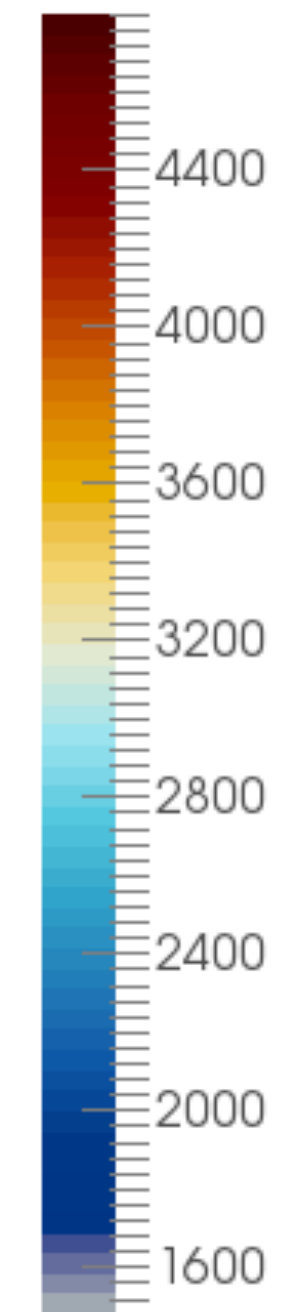
V_p (m/s)



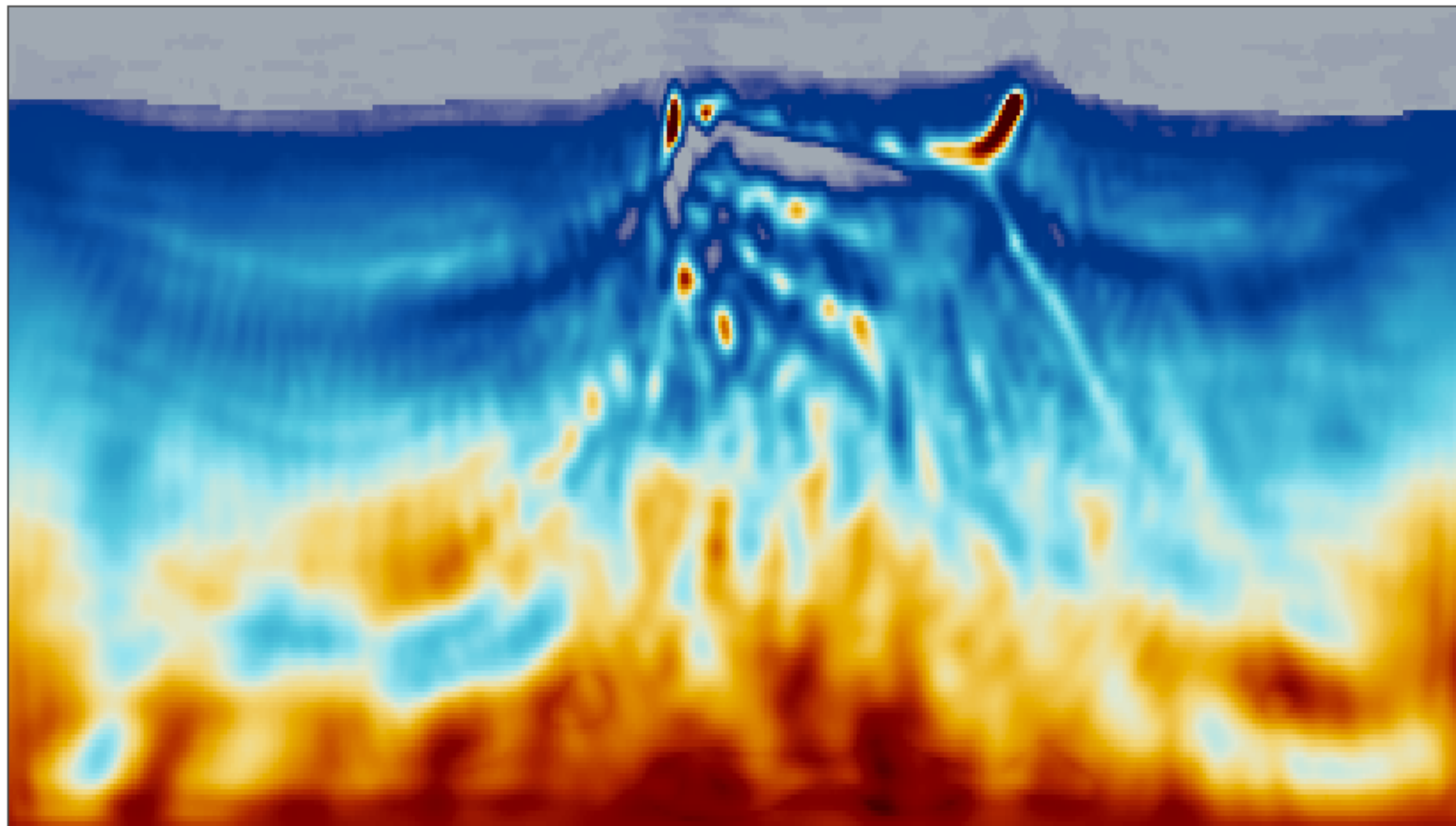
Adjoint-state – w/o constraints



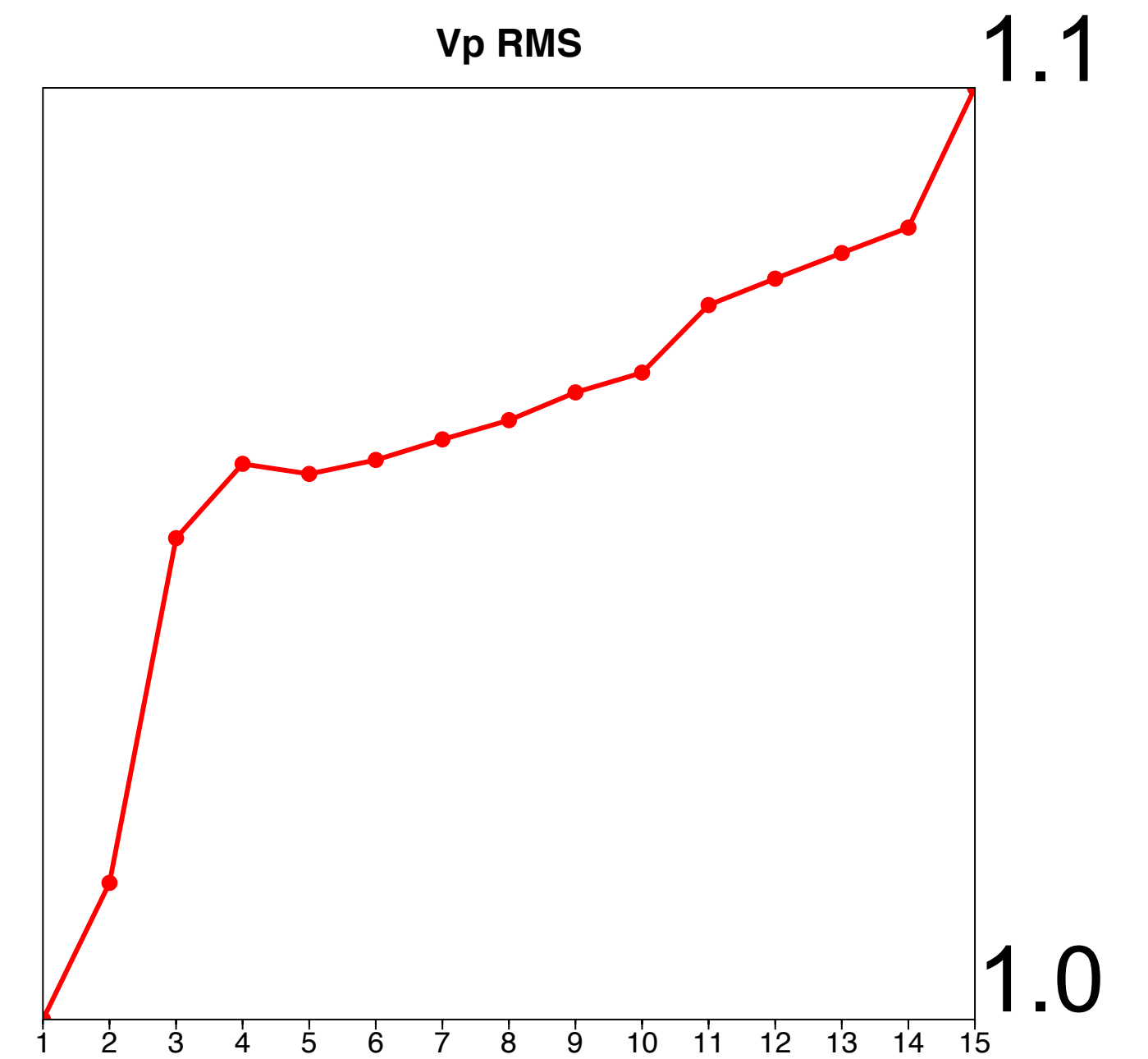
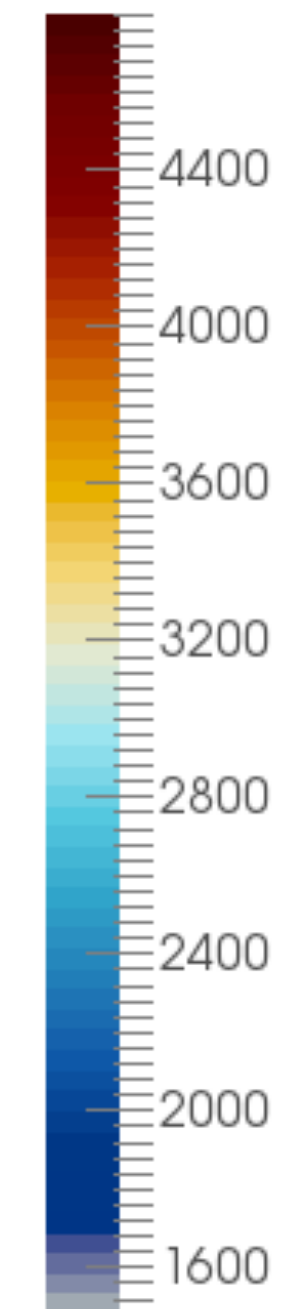
Vp (m/s)



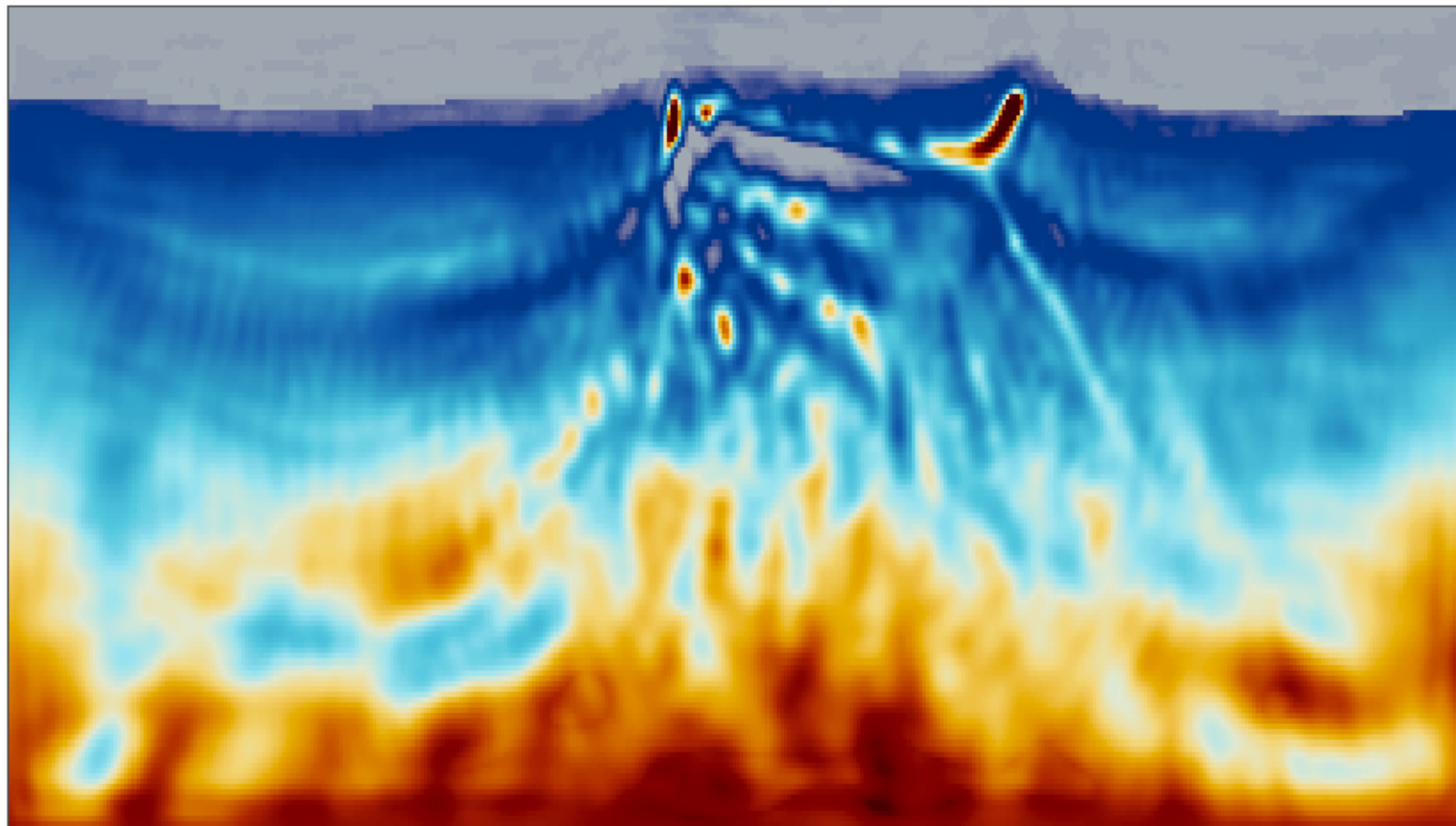
Adjoint-state – w/o constraints



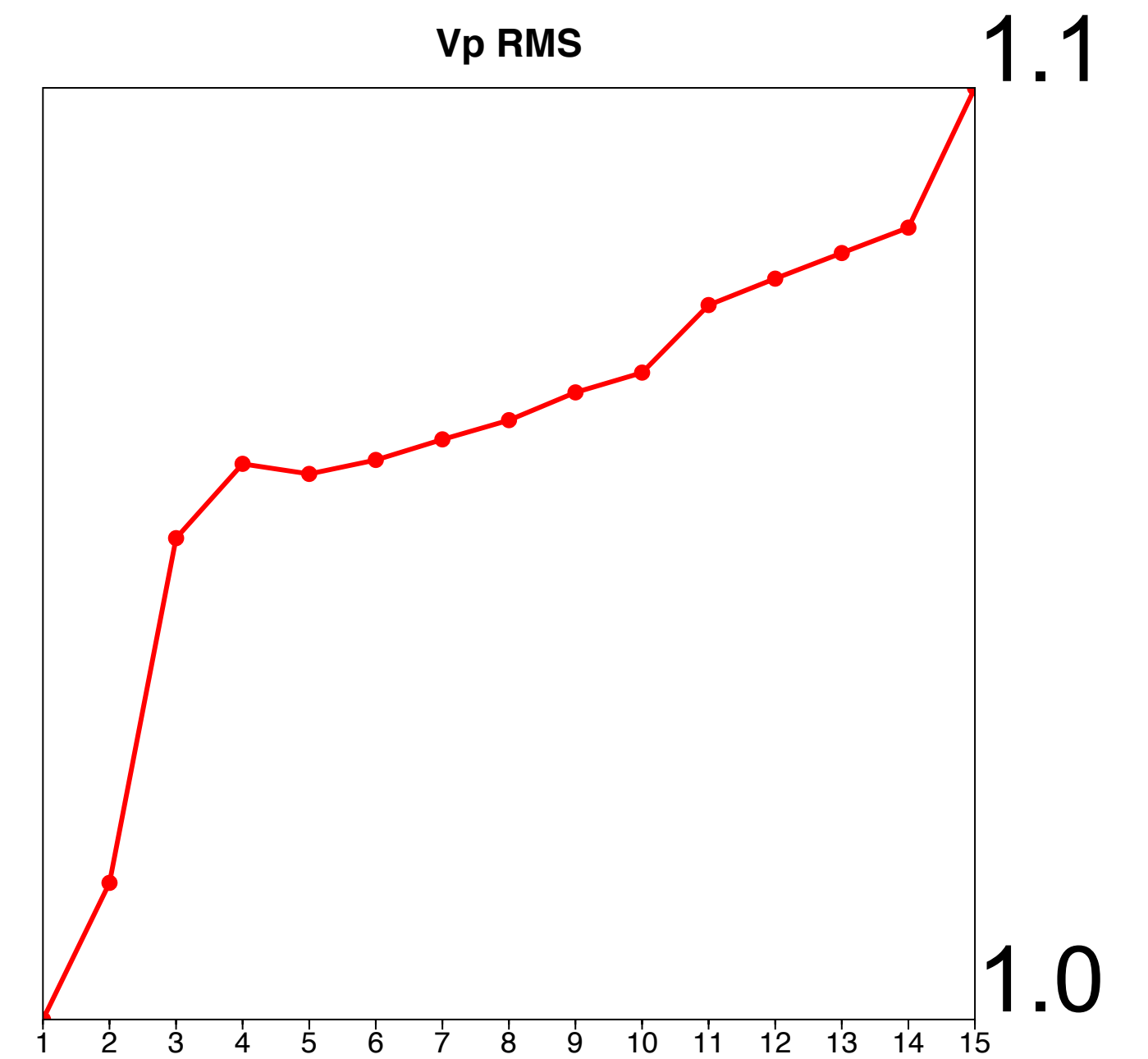
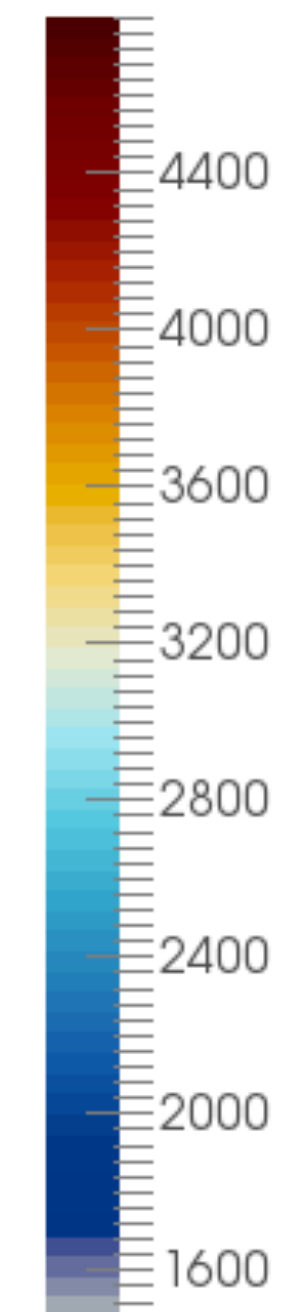
V_p (m/s)



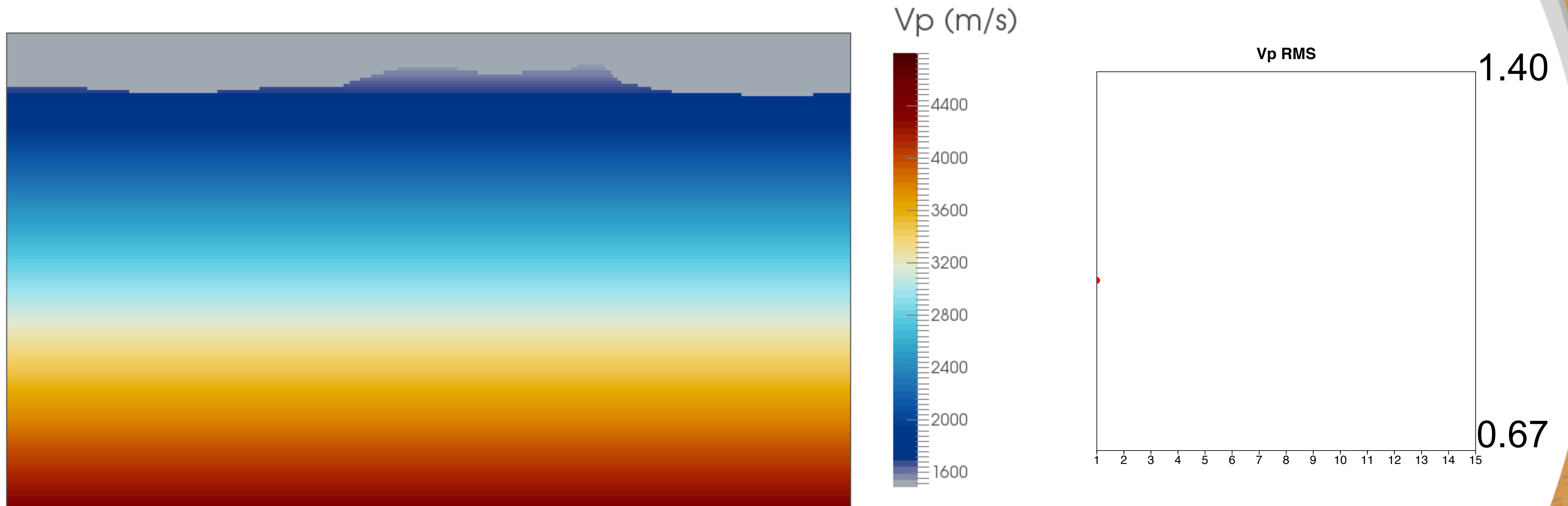
Adjoint-state – w/o constraints



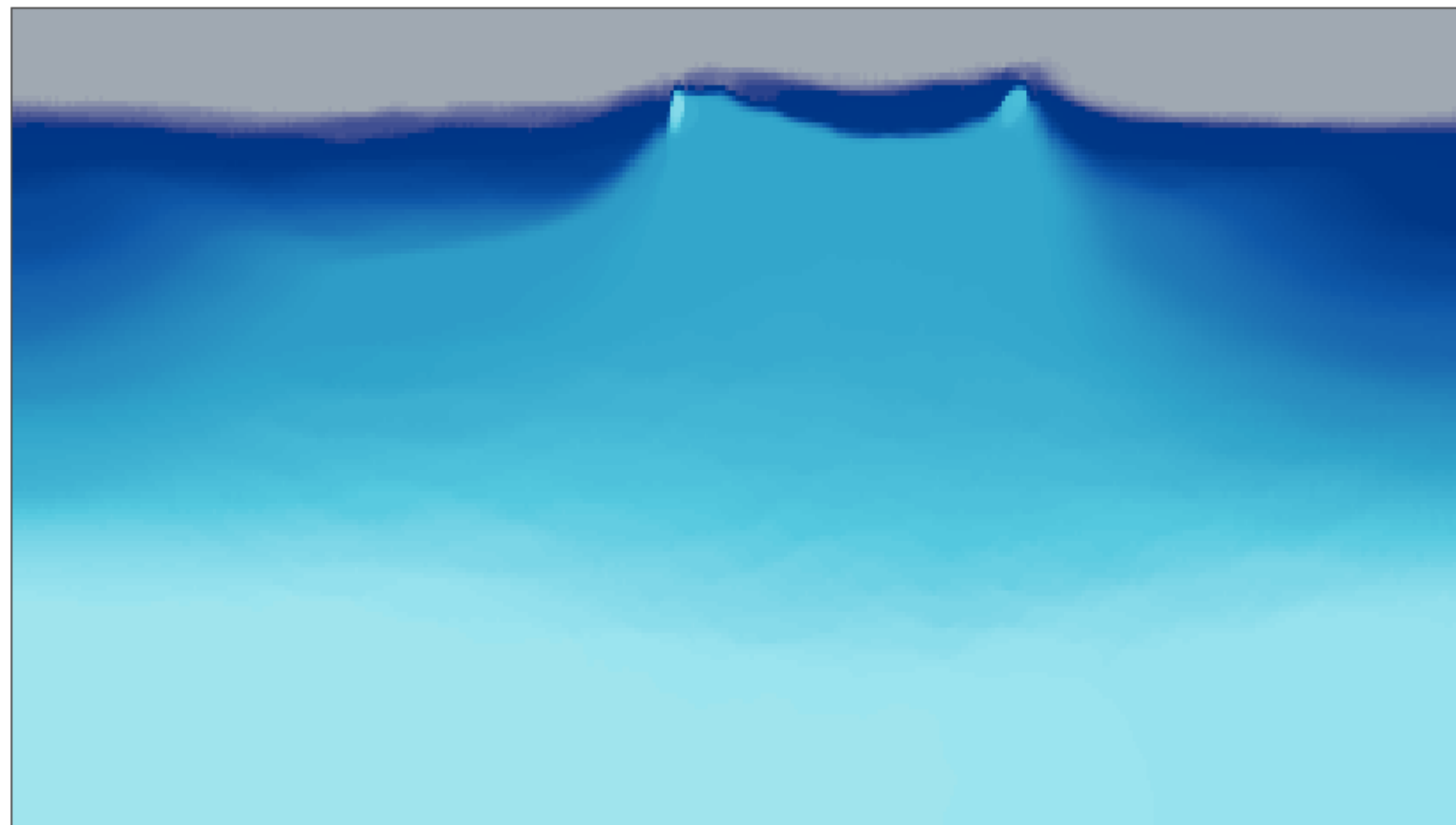
Vp (m/s)



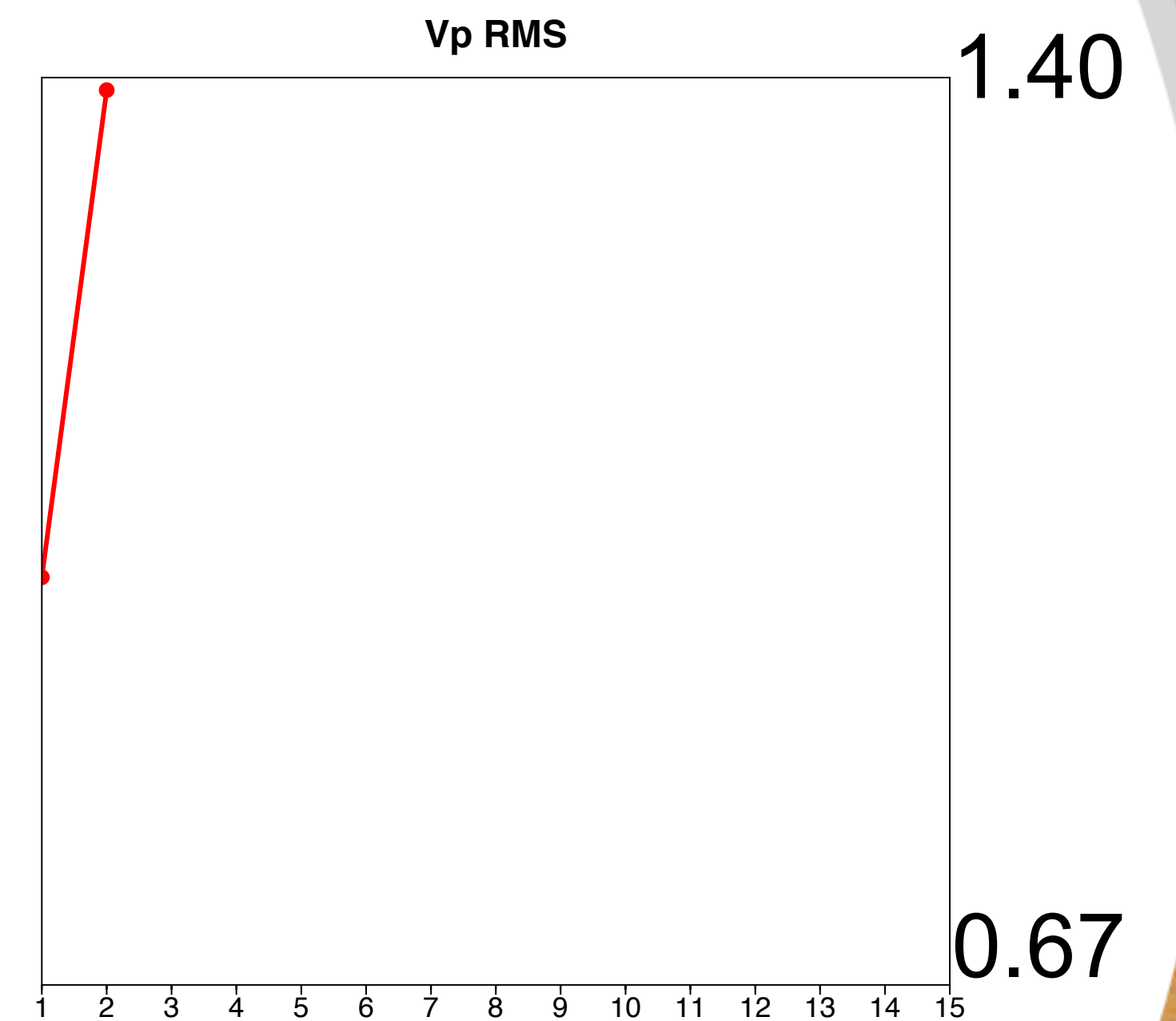
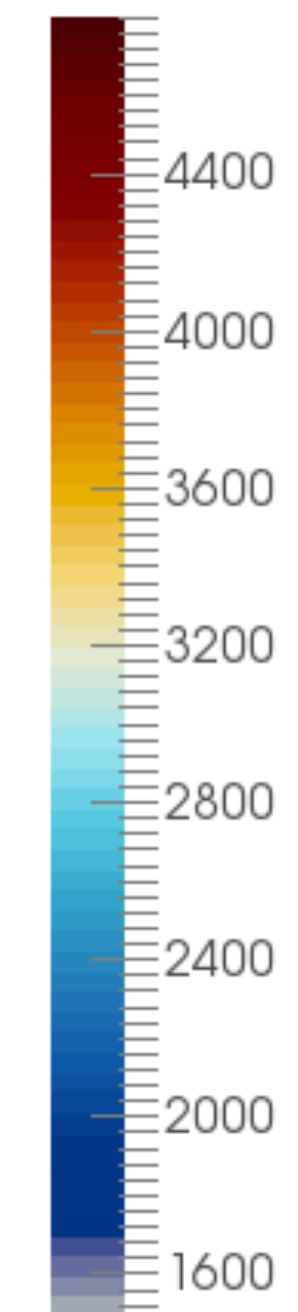
Adjoint-state – w/ TV-norm & hinge-loss projections



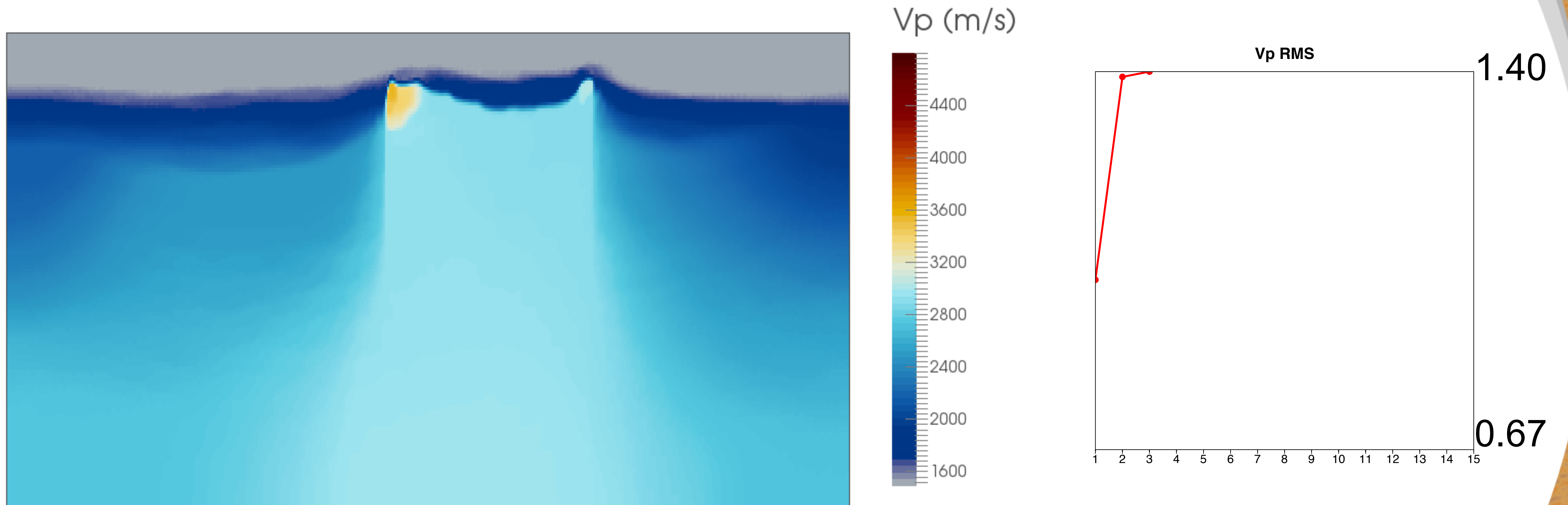
Adjoint-state – w/ TV-norm & hinge-loss projections



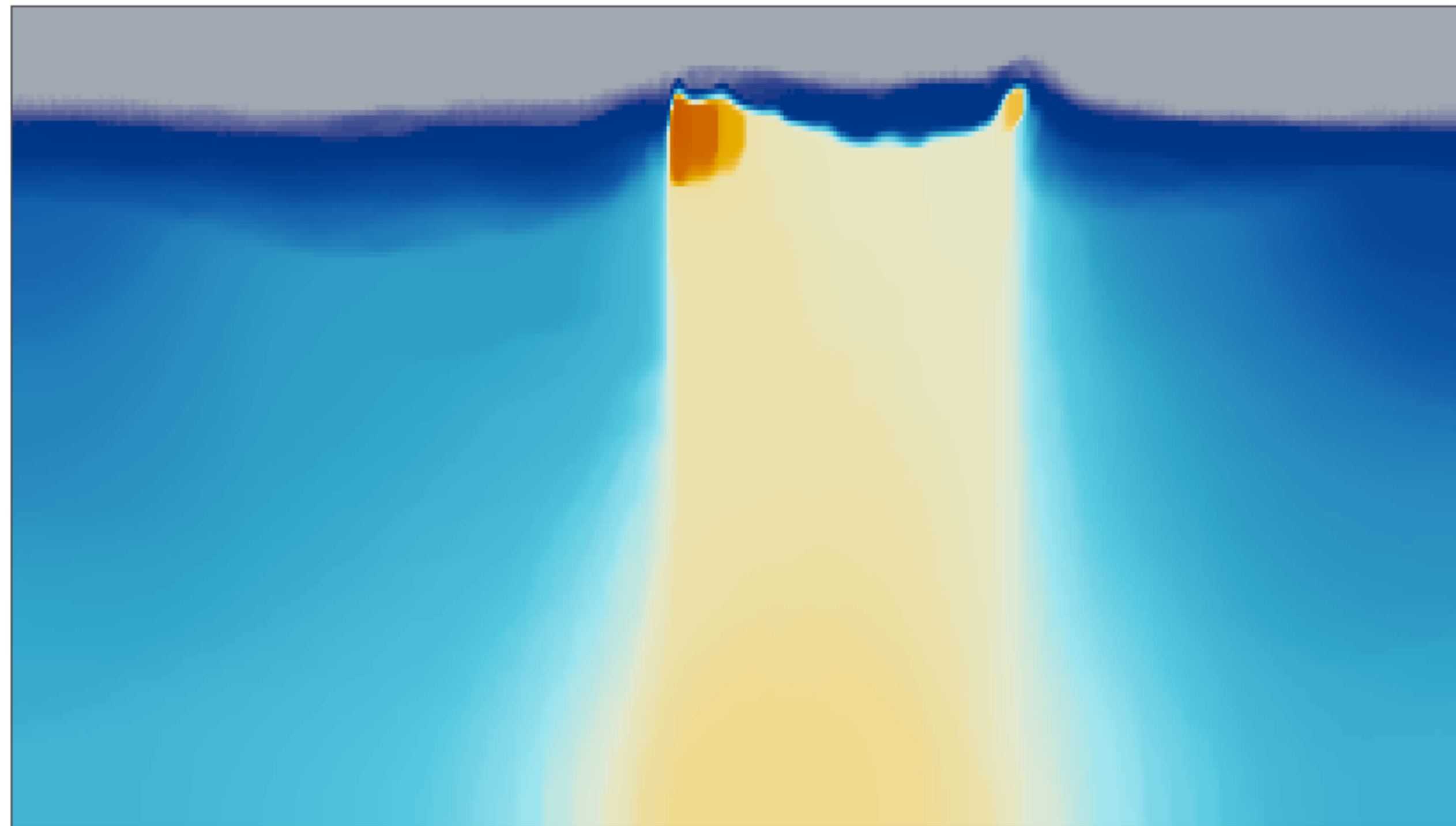
Vp (m/s)



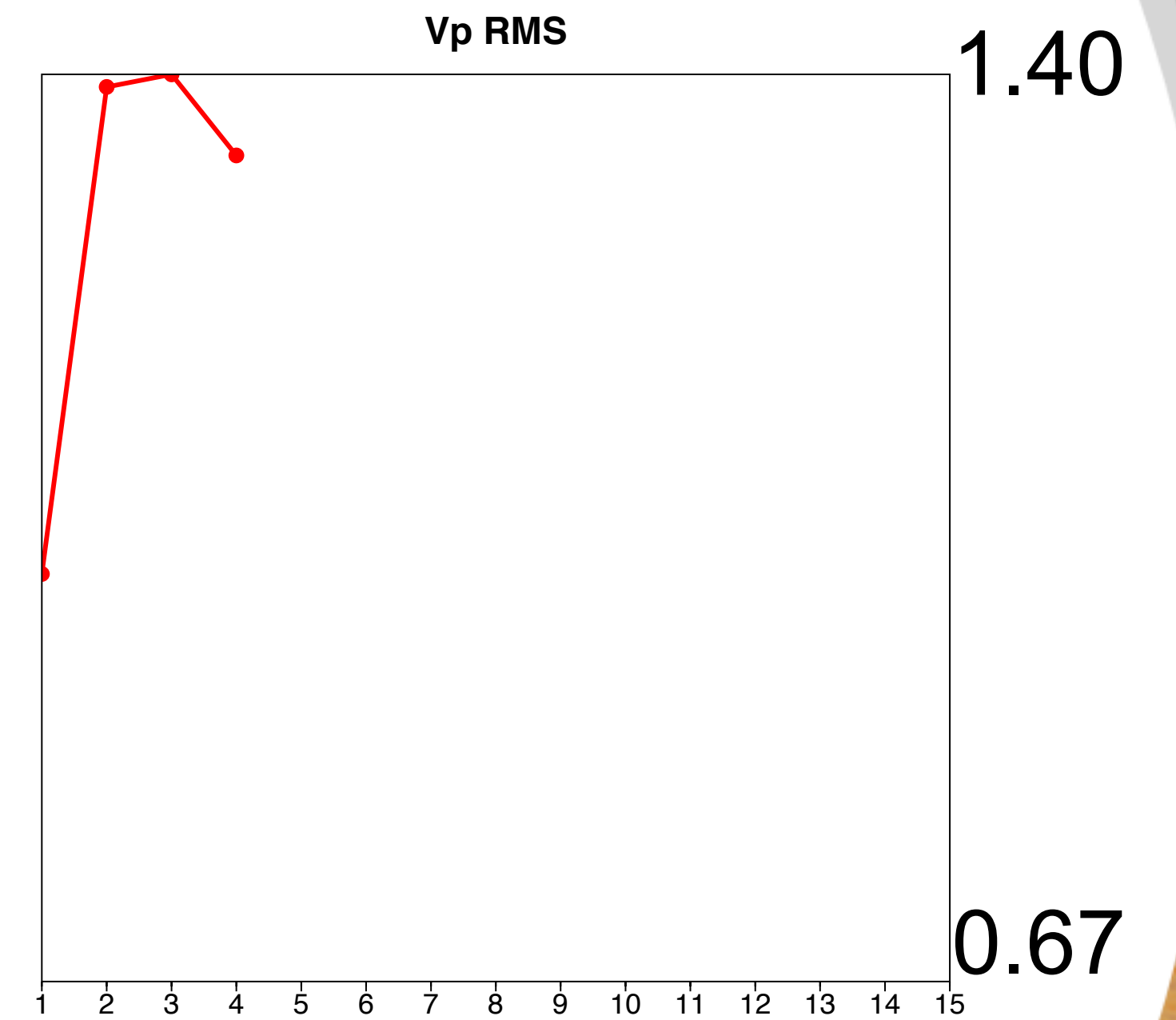
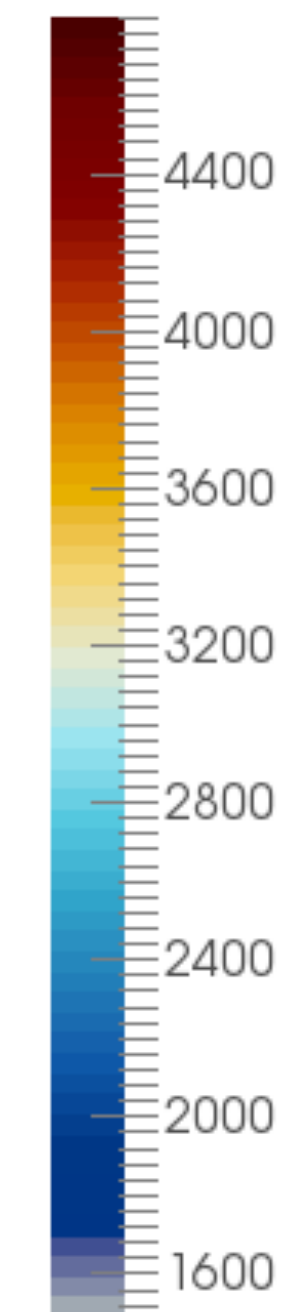
Adjoint-state – w/ TV-norm & hinge-loss projections



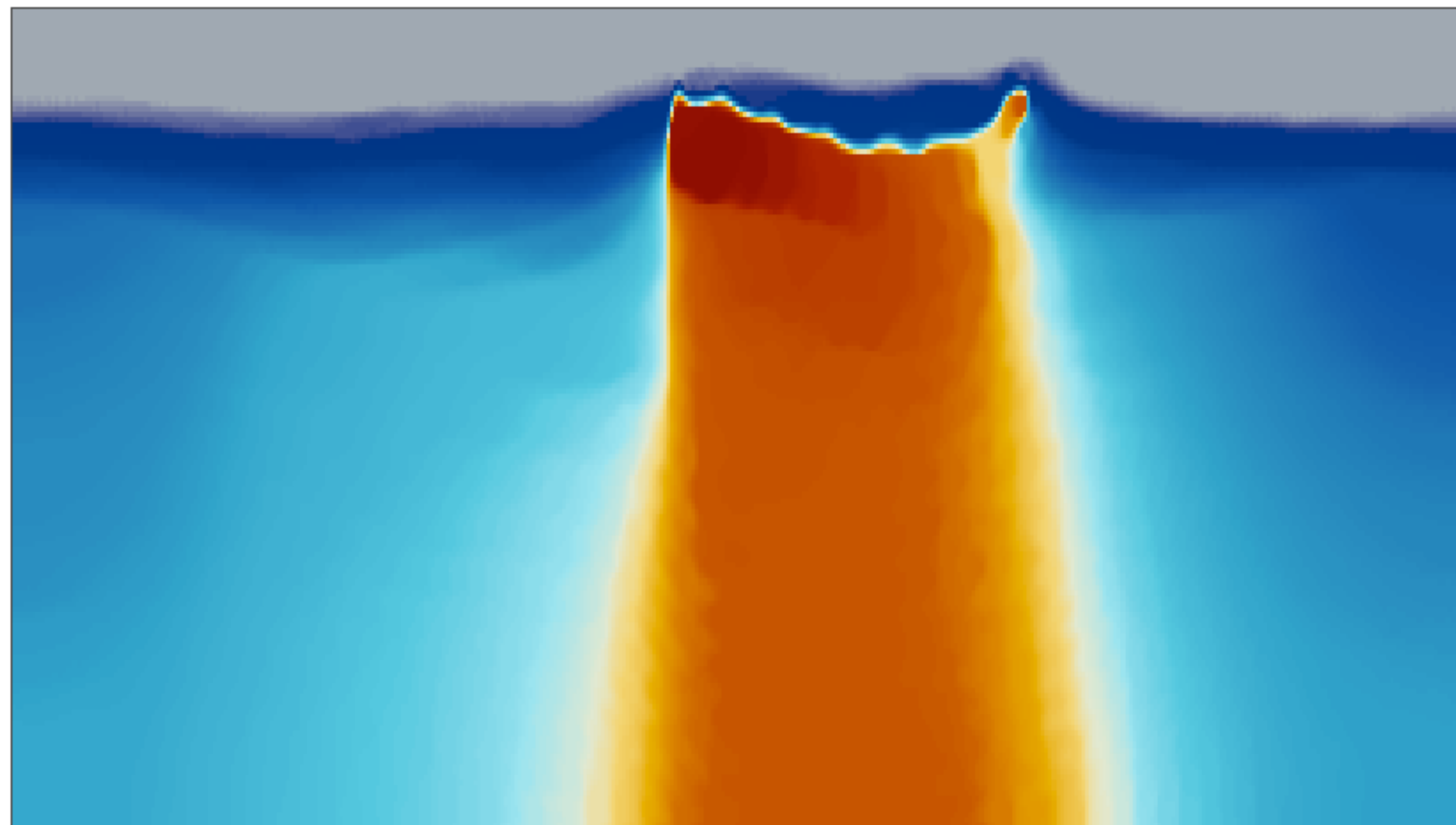
Adjoint-state – w/ TV-norm & hinge-loss projections



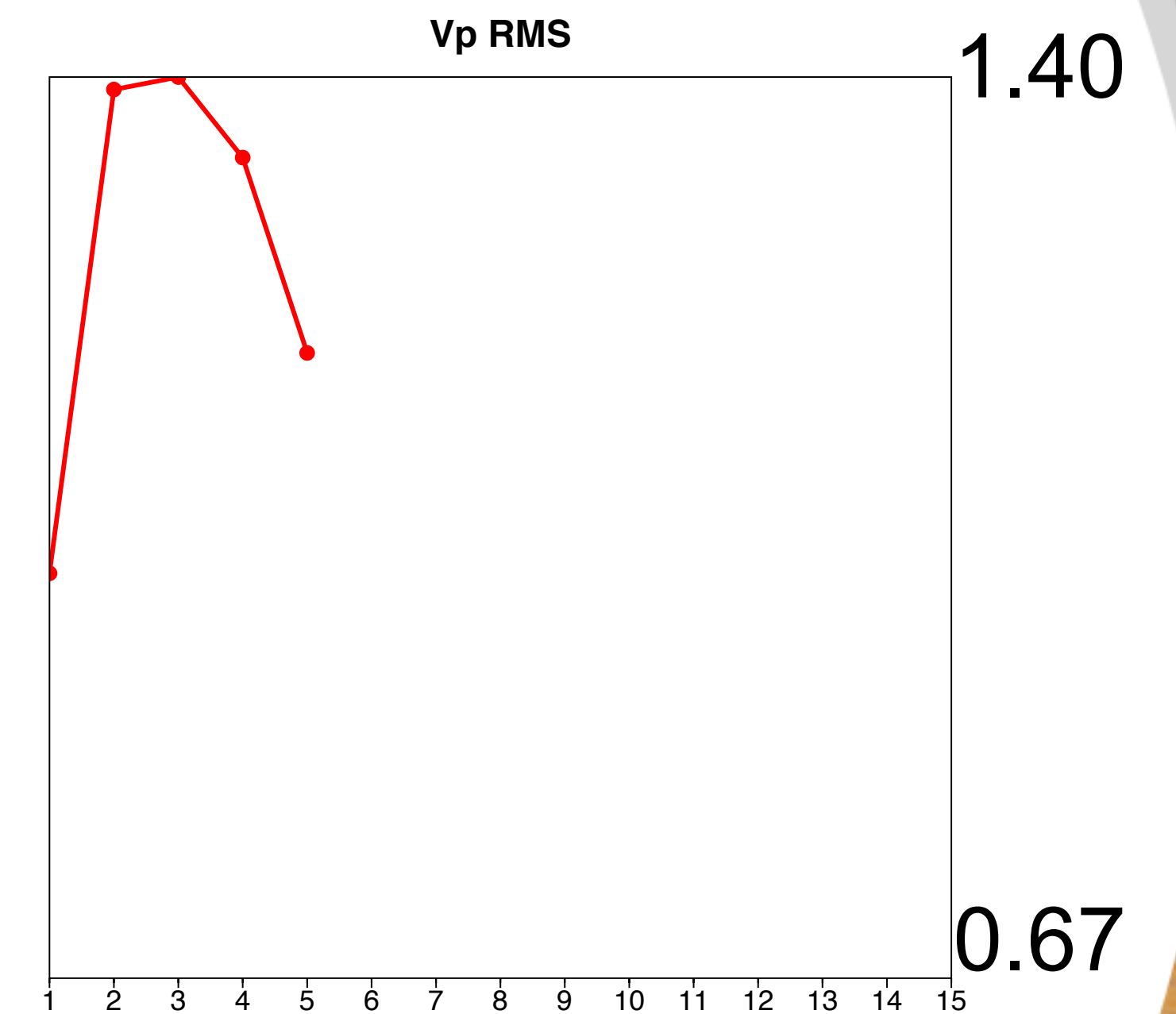
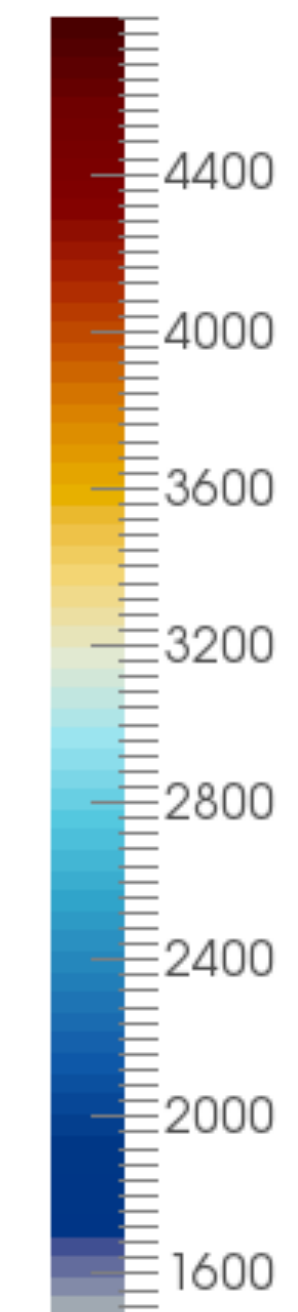
Vp (m/s)



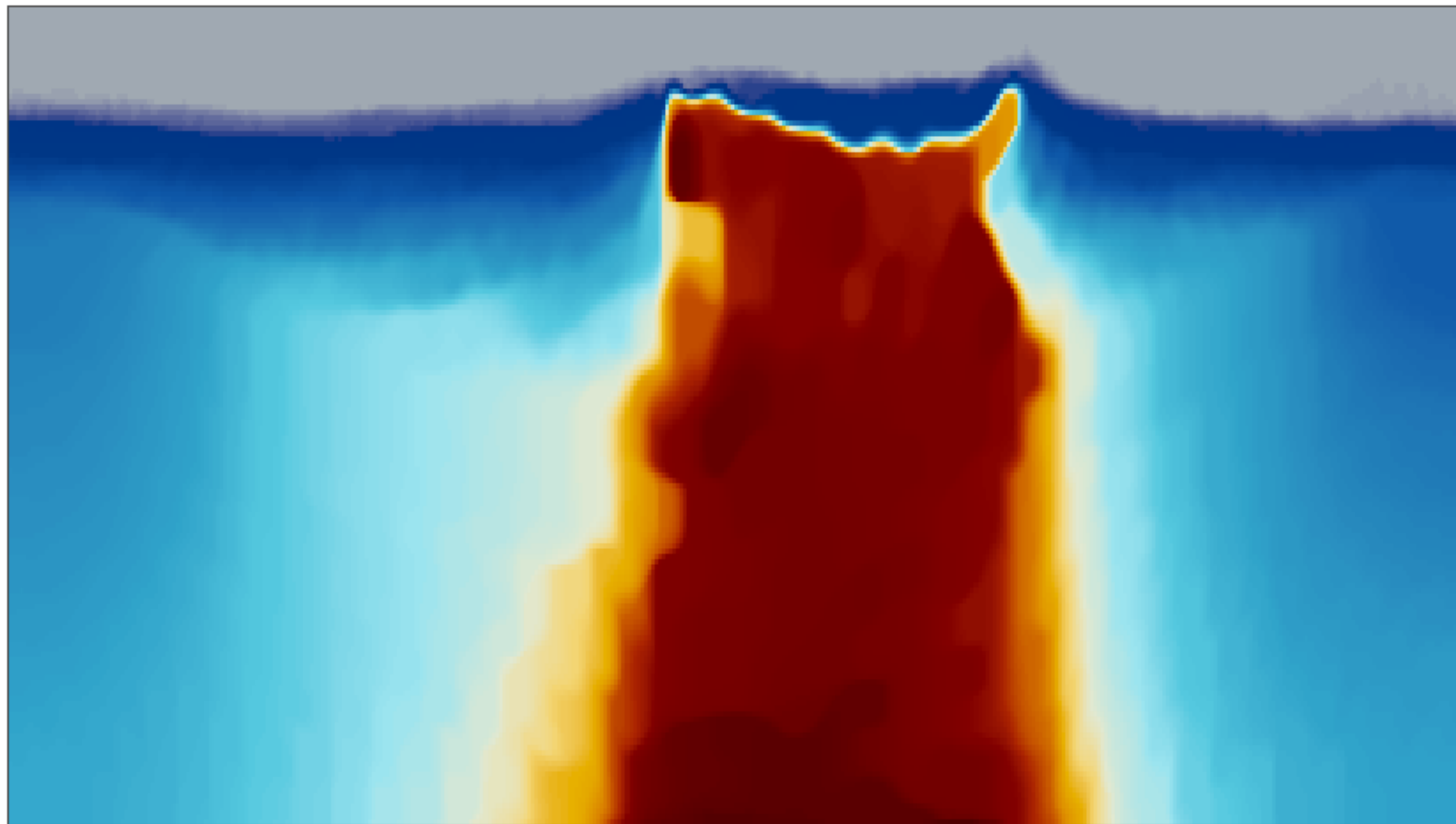
Adjoint-state – w/ TV-norm & hinge-loss projections



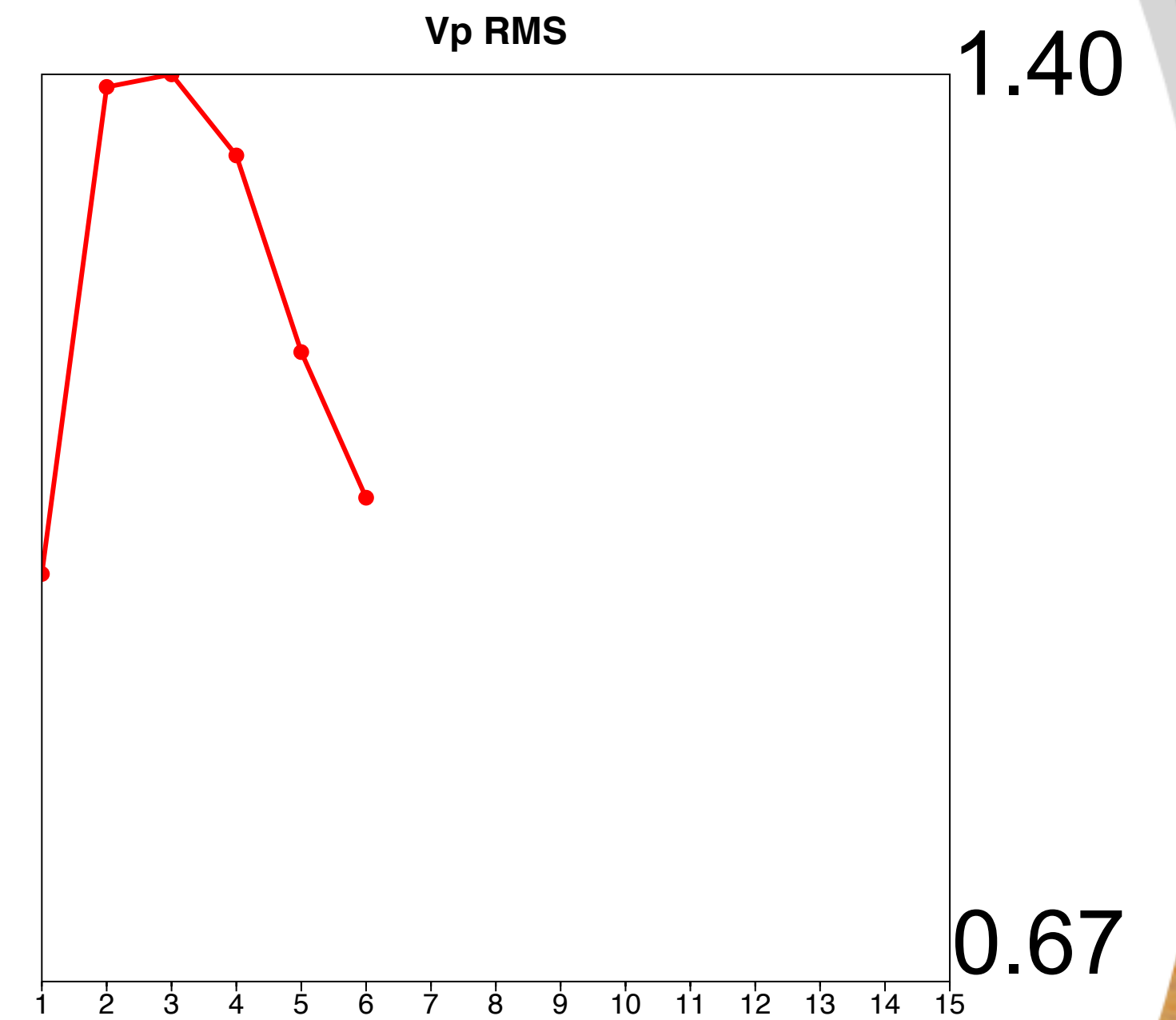
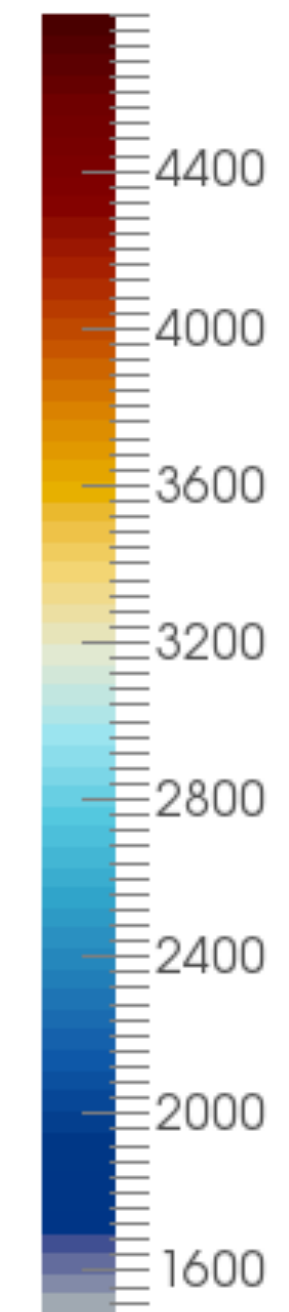
Vp (m/s)



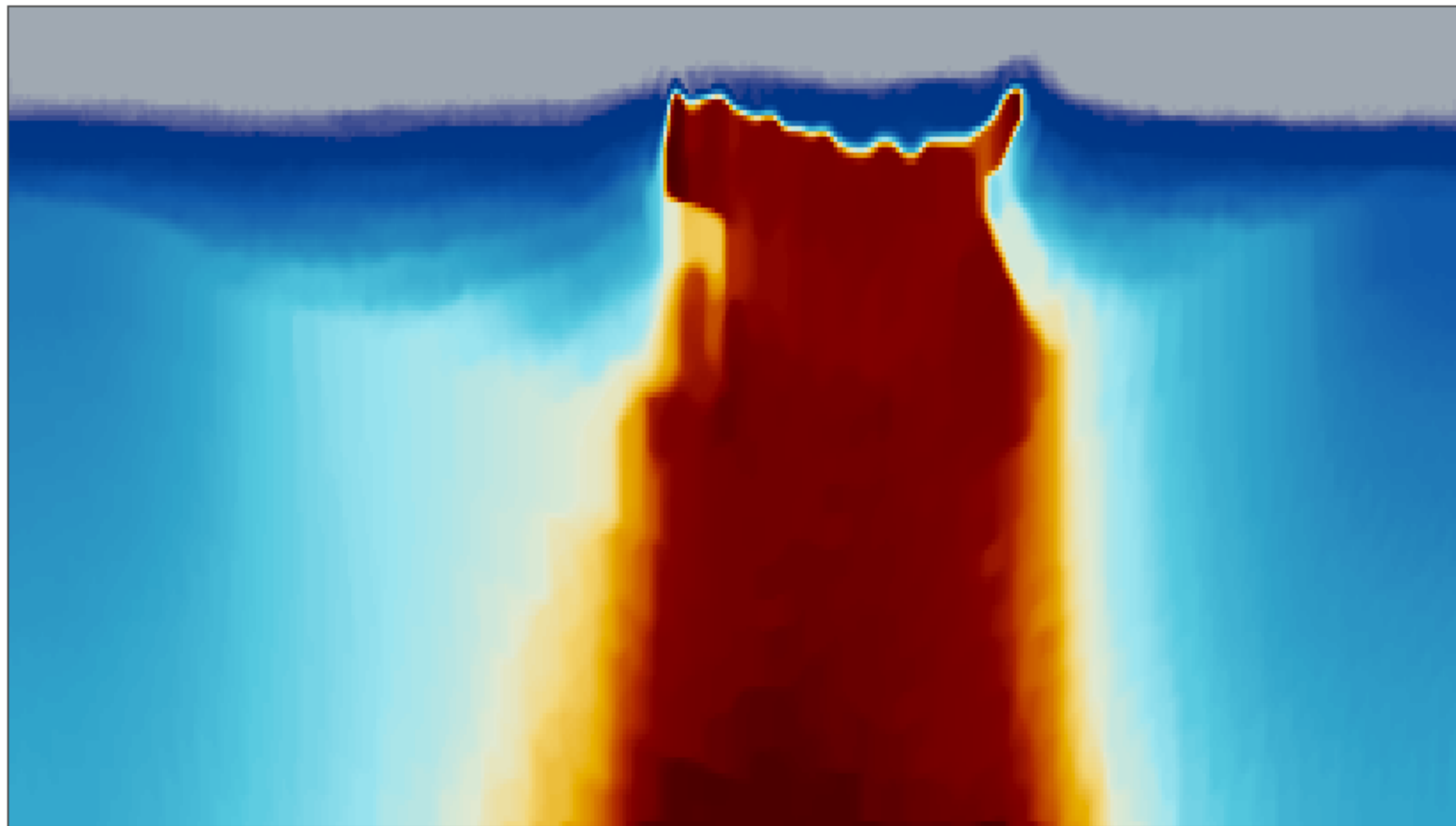
Adjoint-state – w/ TV-norm & hinge-loss projections



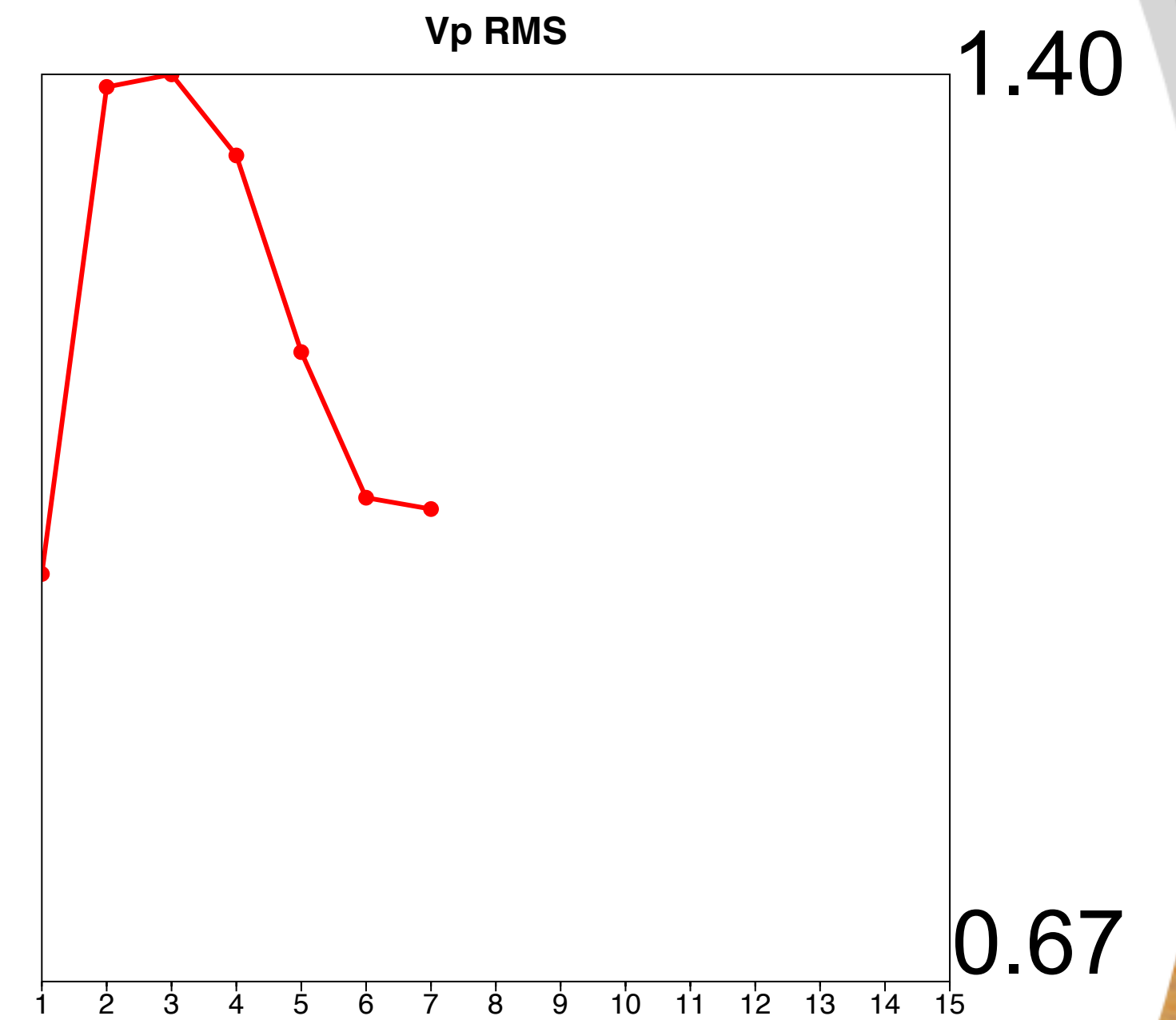
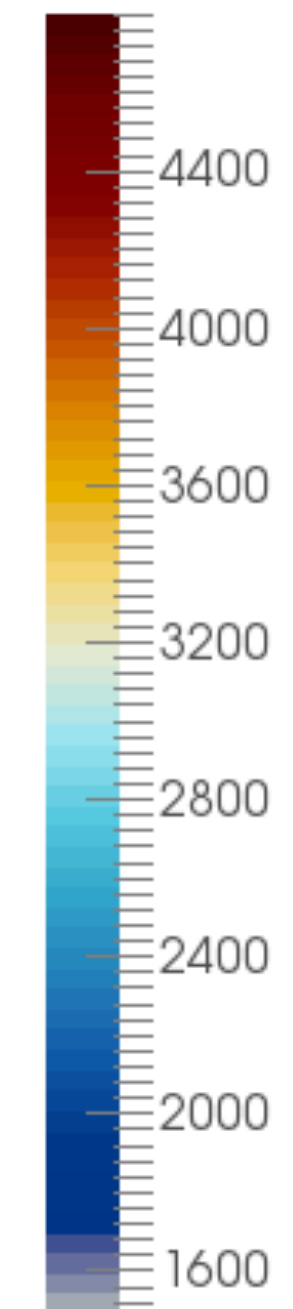
Vp (m/s)



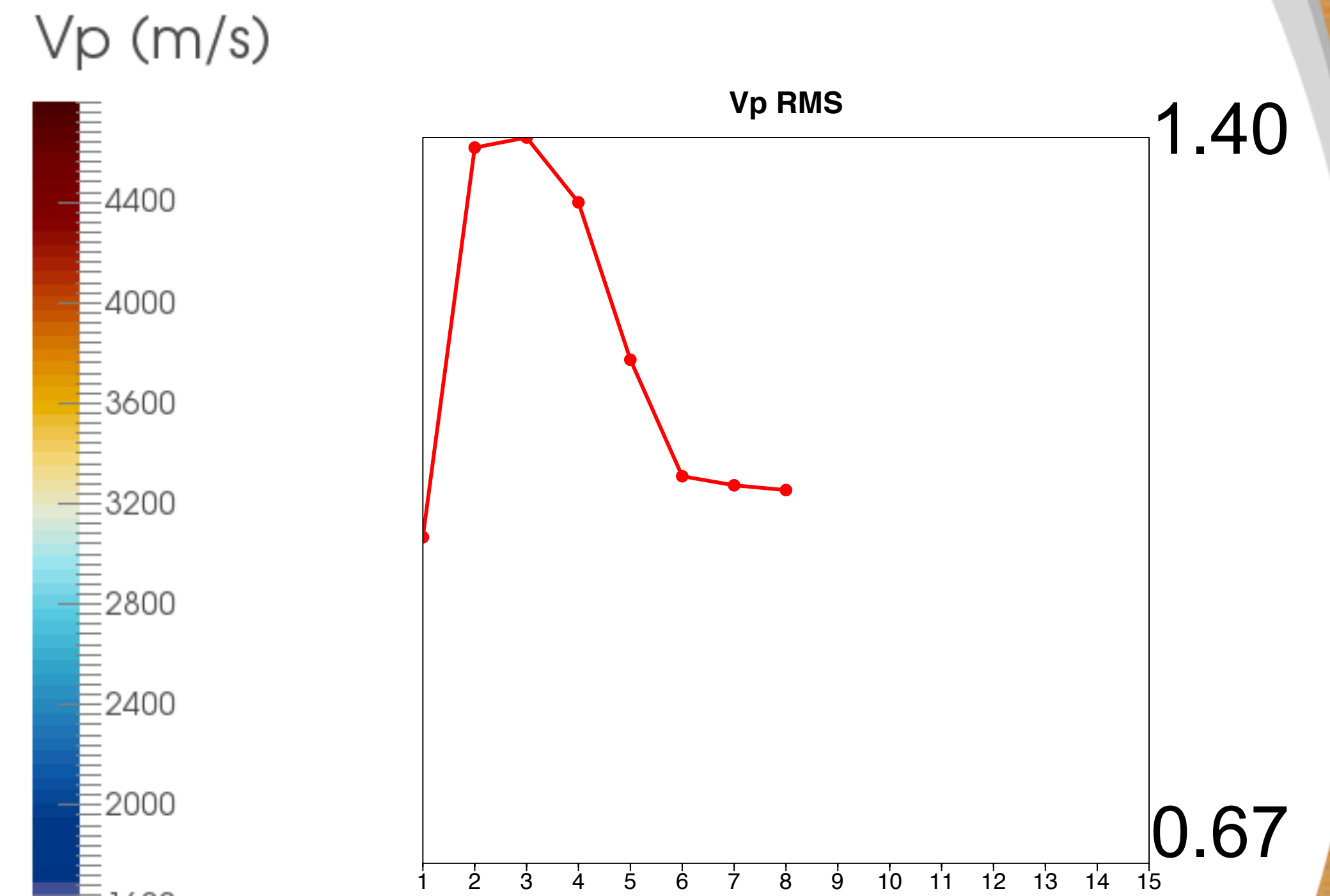
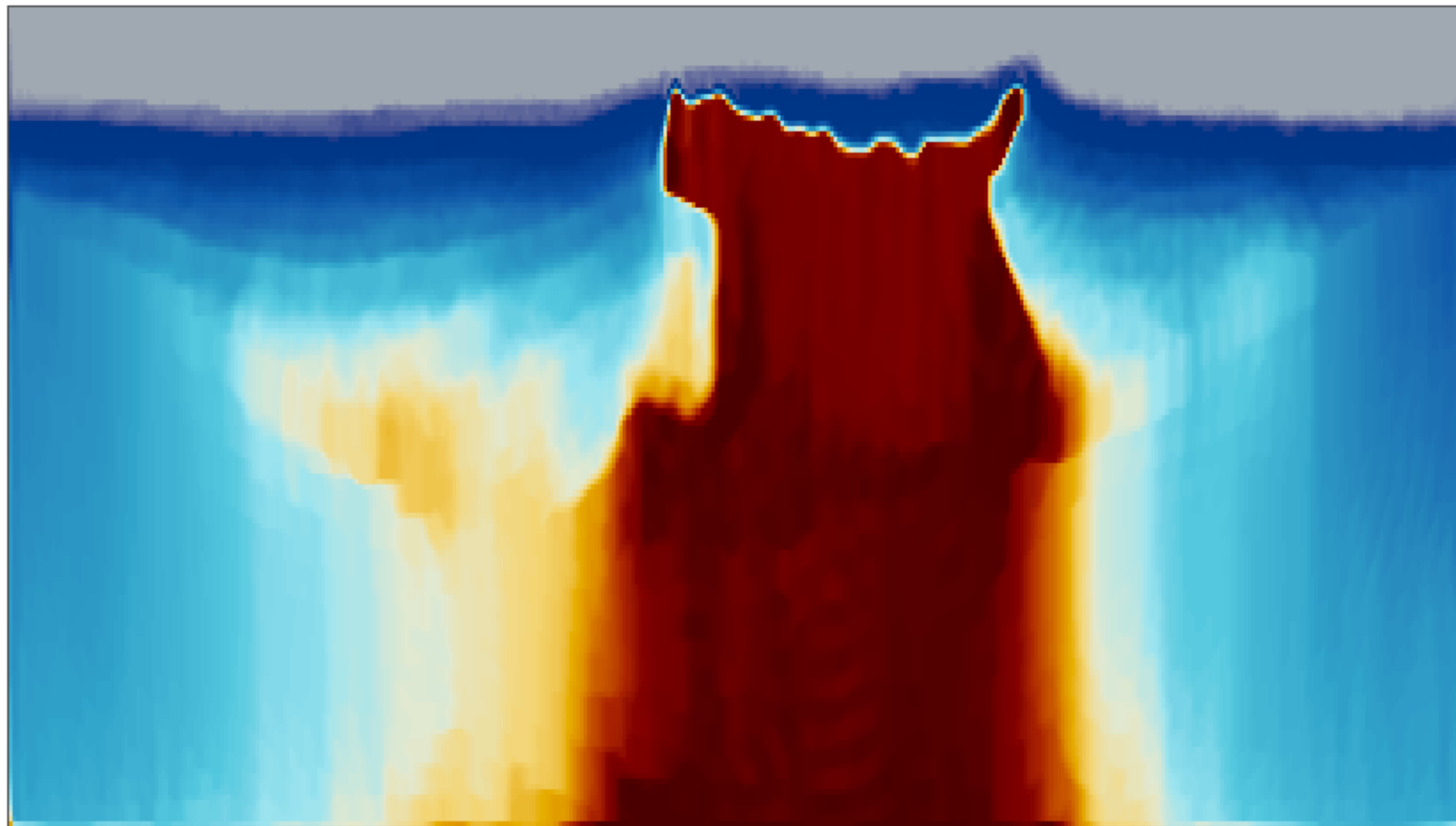
Adjoint-state – w/ TV-norm & hinge-loss projections



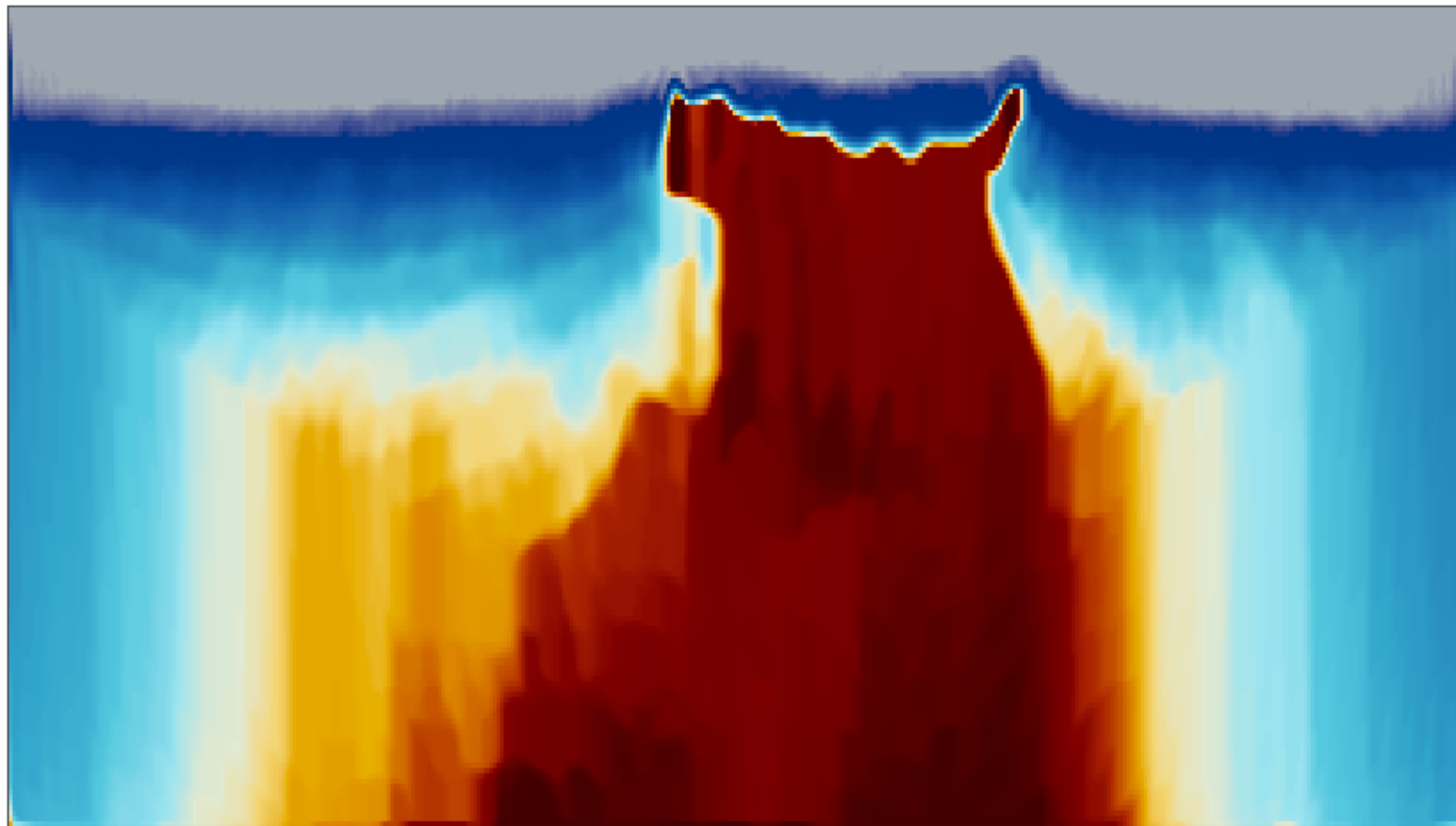
Vp (m/s)



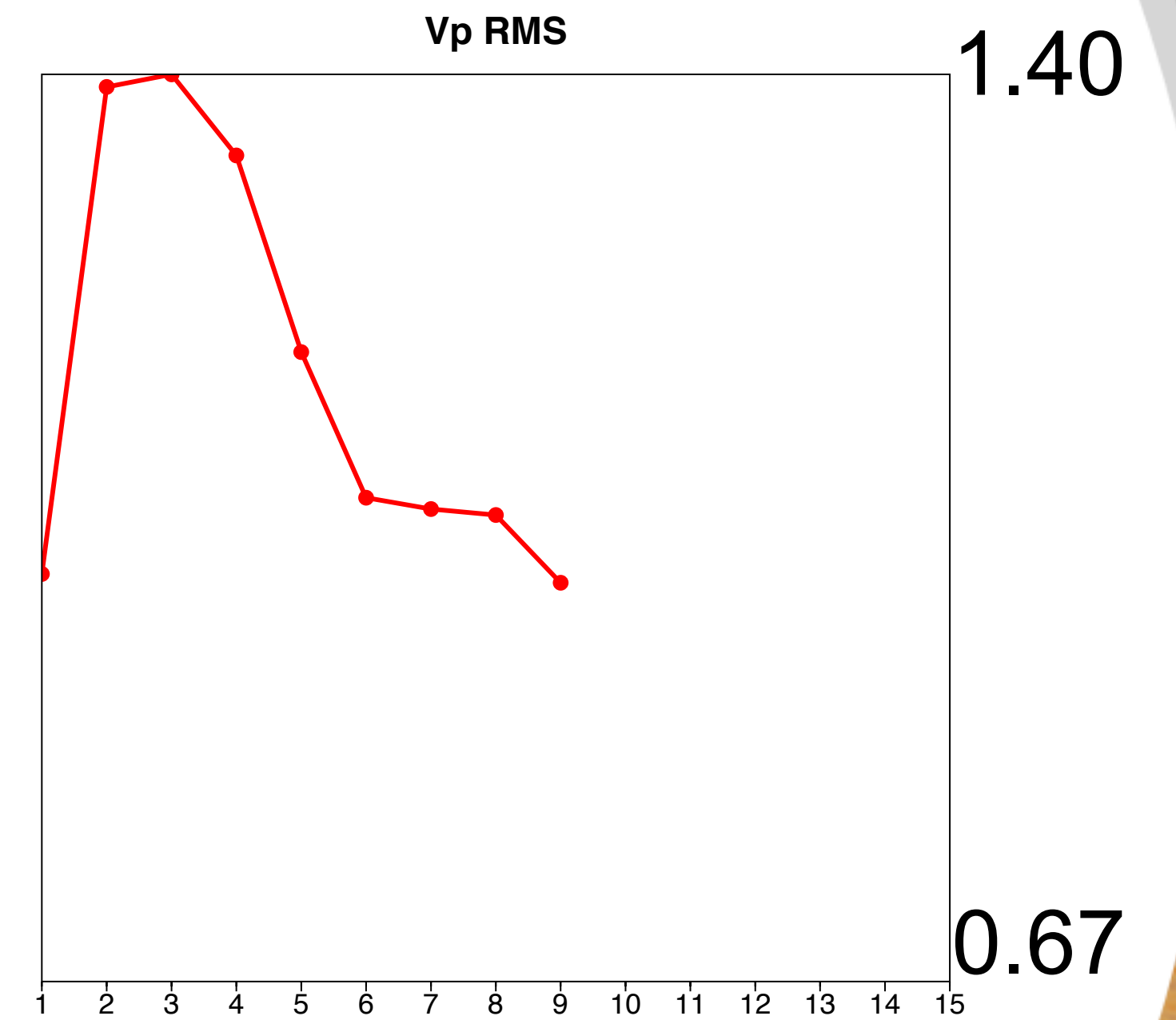
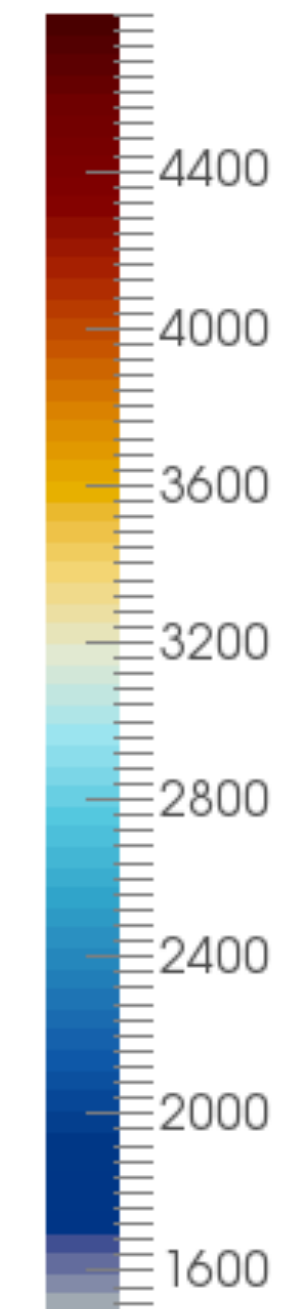
Adjoint-state – w/ TV-norm & hinge-loss projections



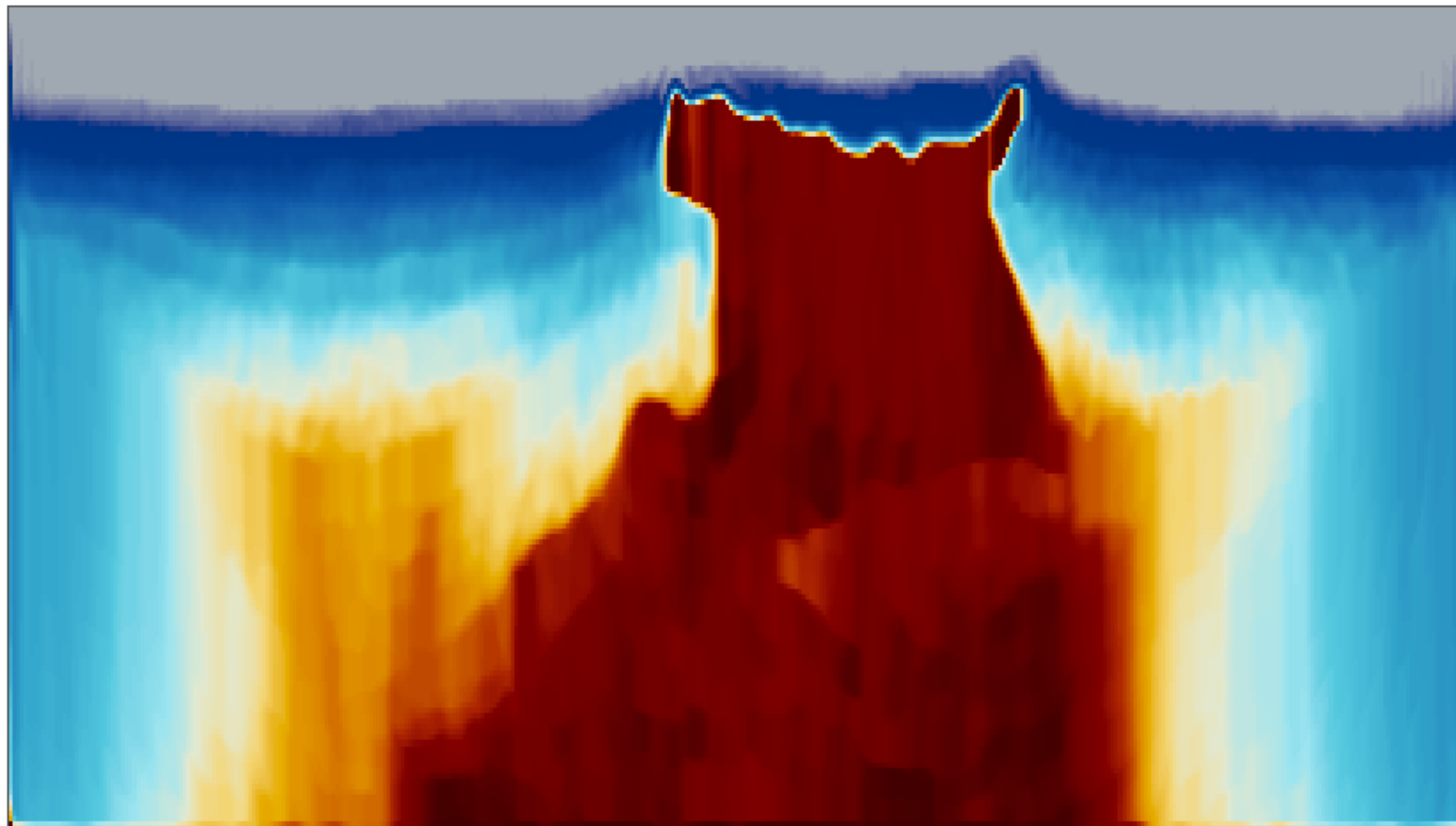
Adjoint-state – w/ TV-norm & hinge-loss projections



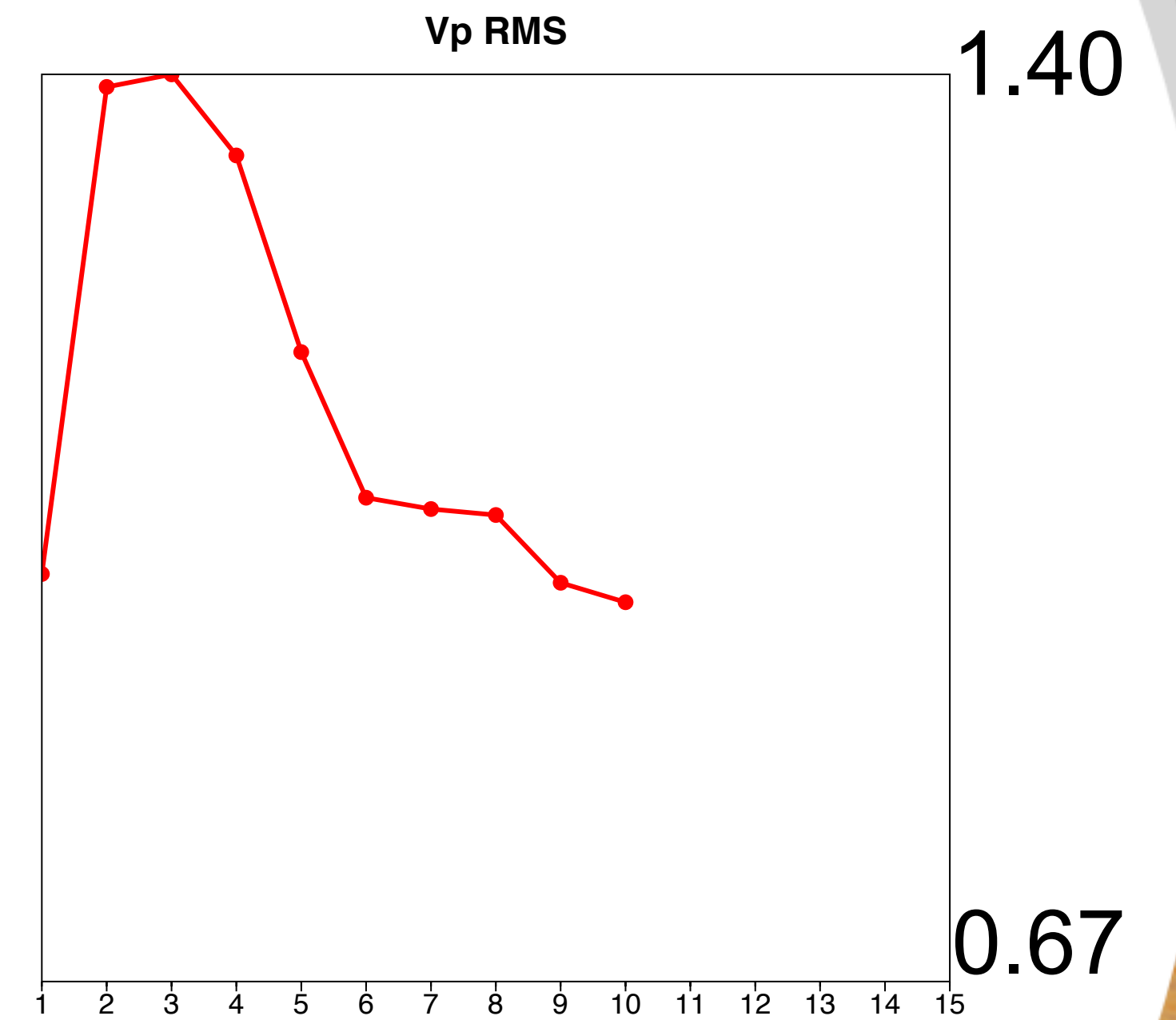
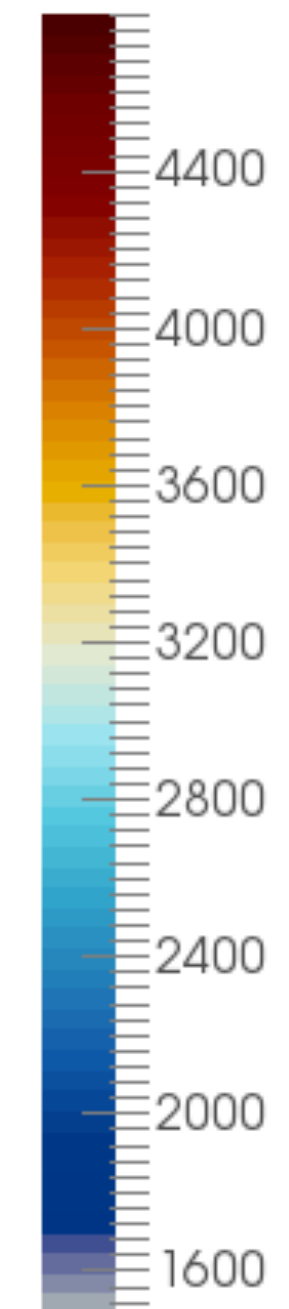
V_p (m/s)



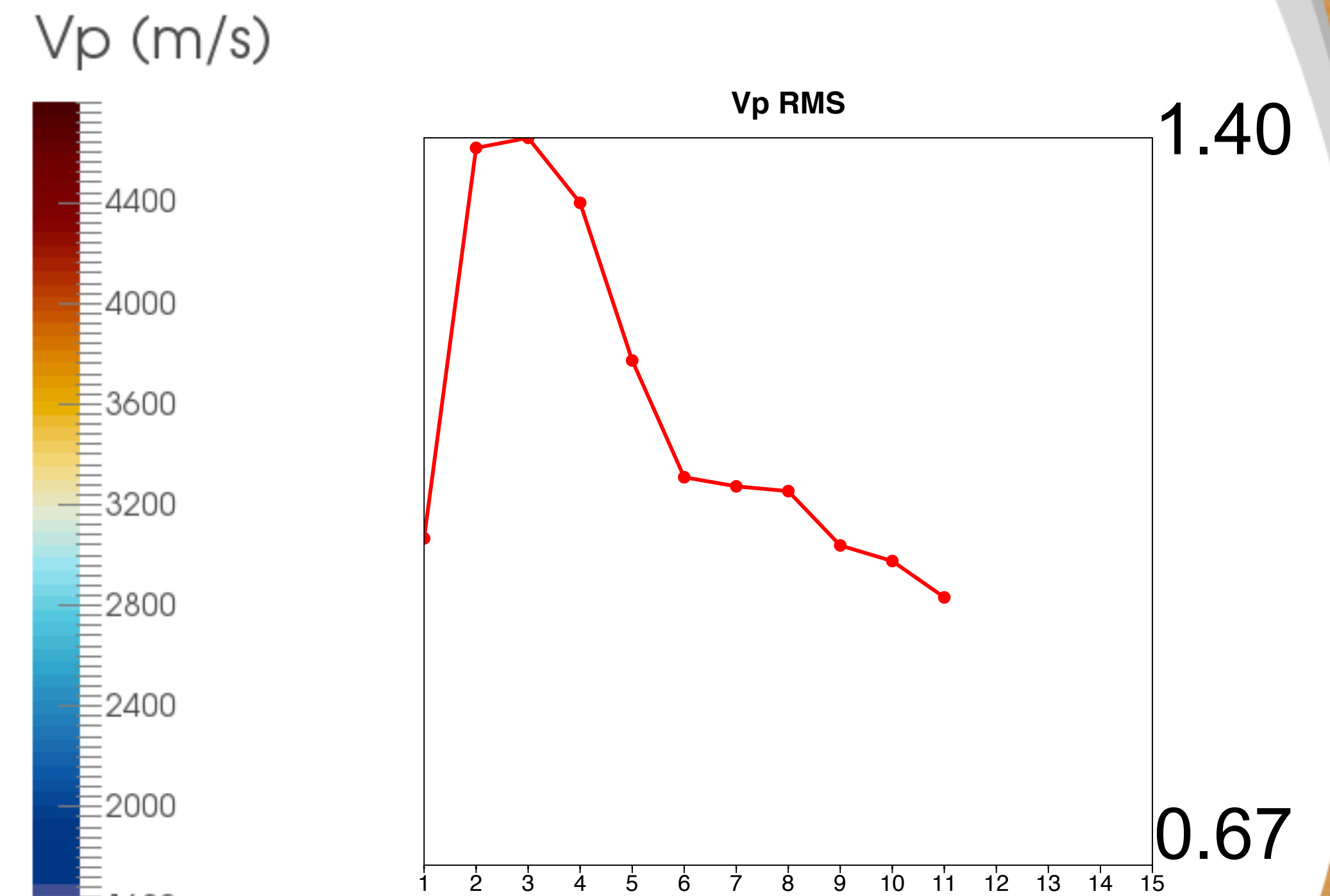
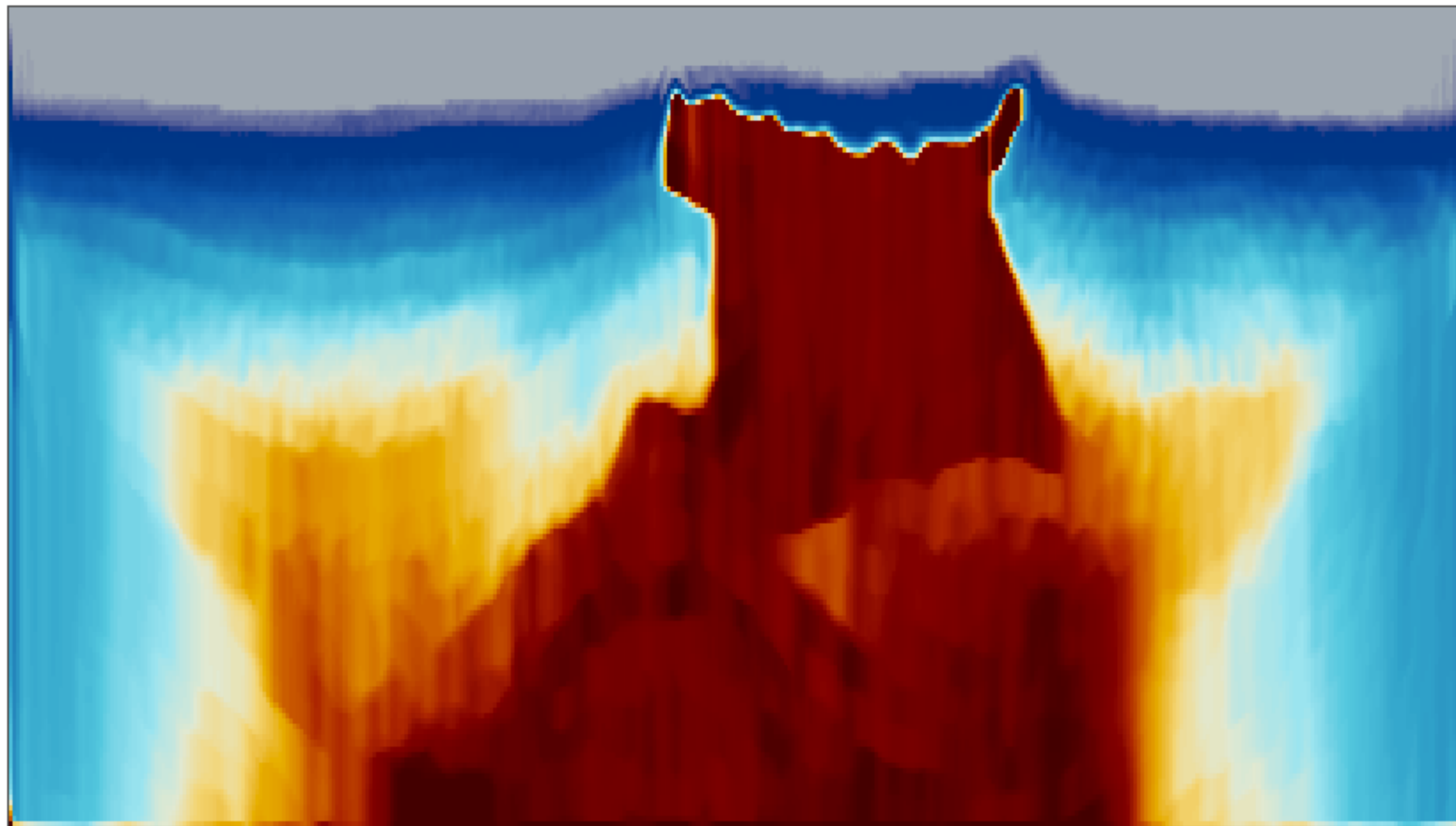
Adjoint-state – w/ TV-norm & hinge-loss projections



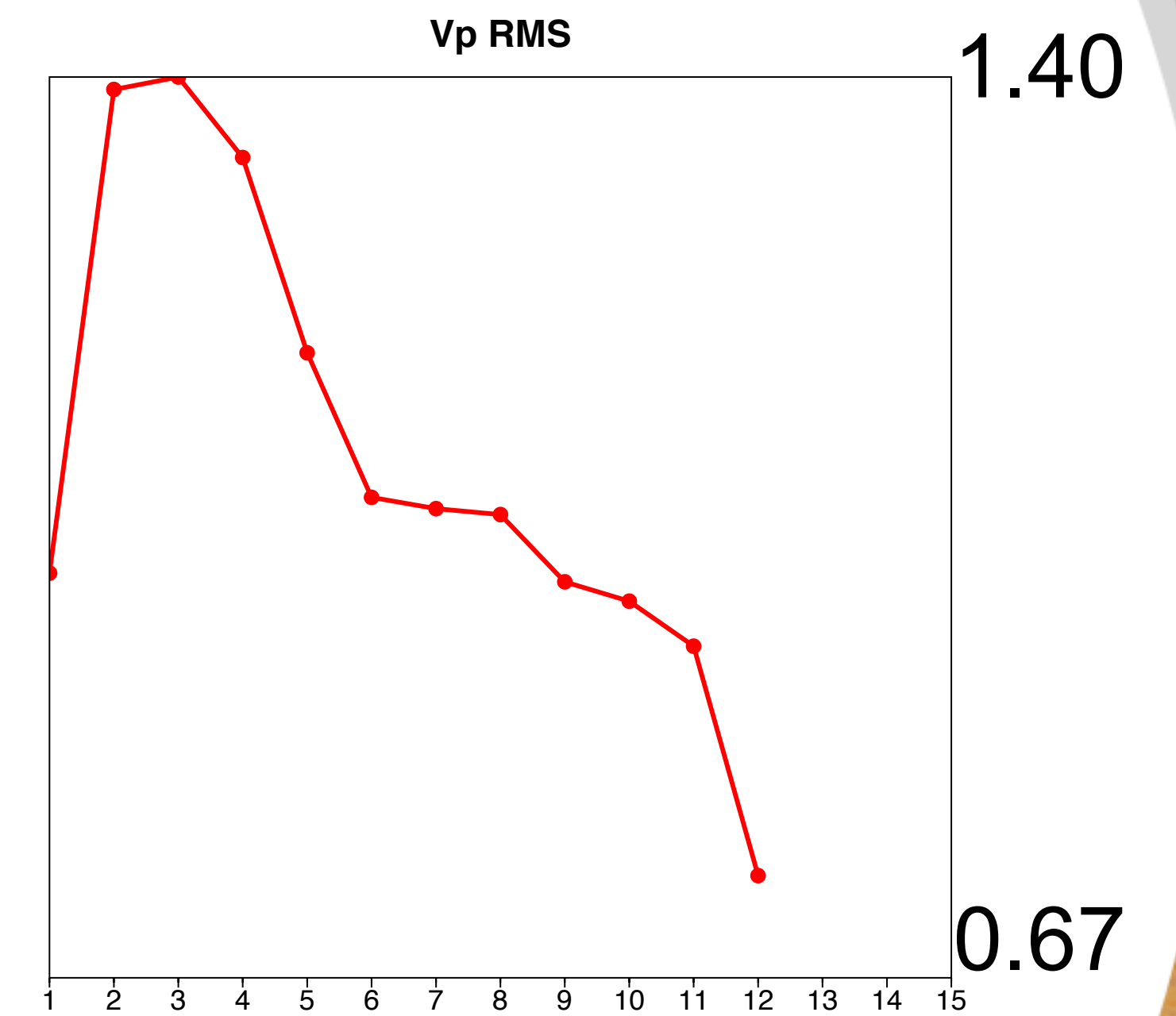
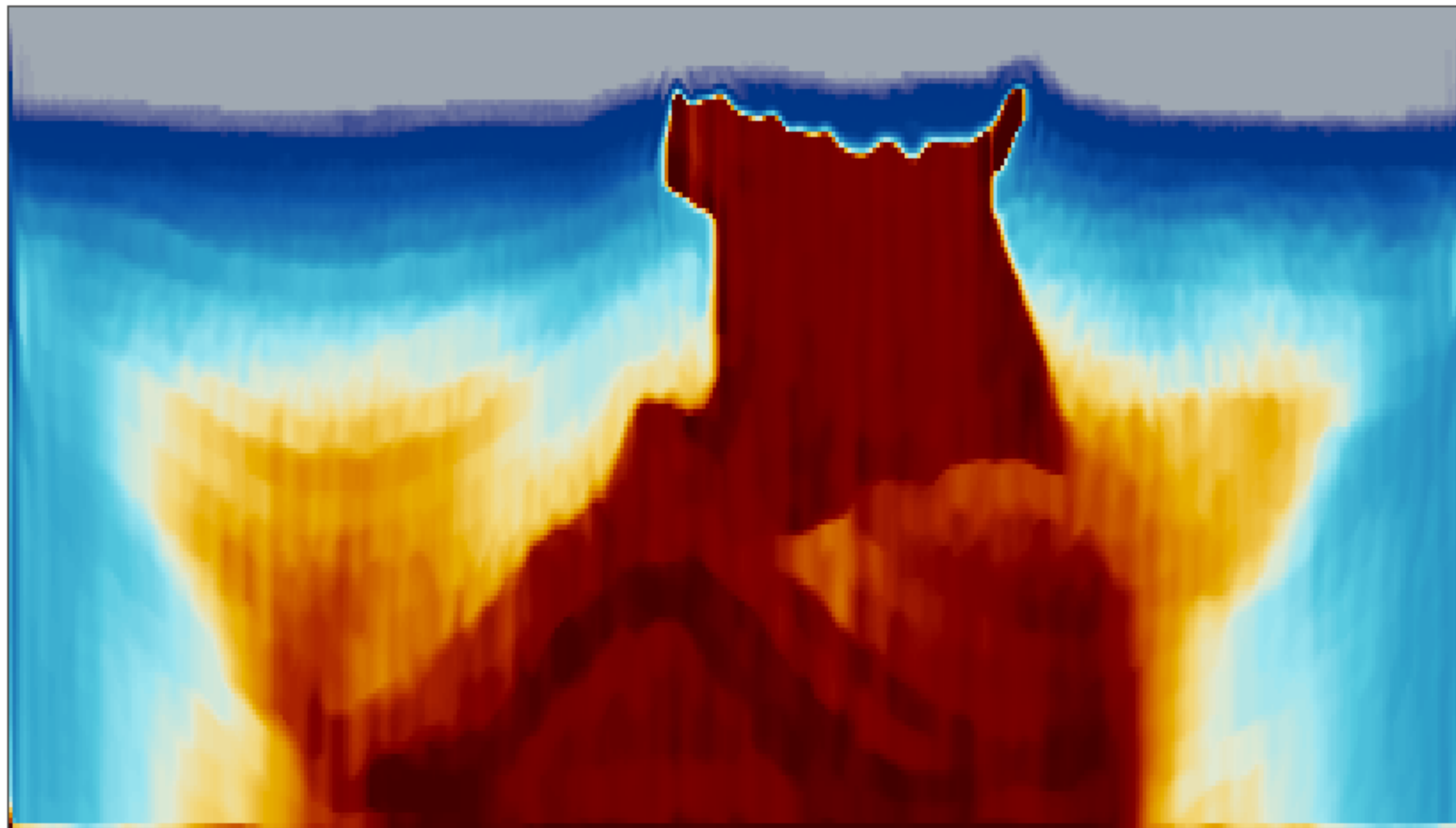
Vp (m/s)



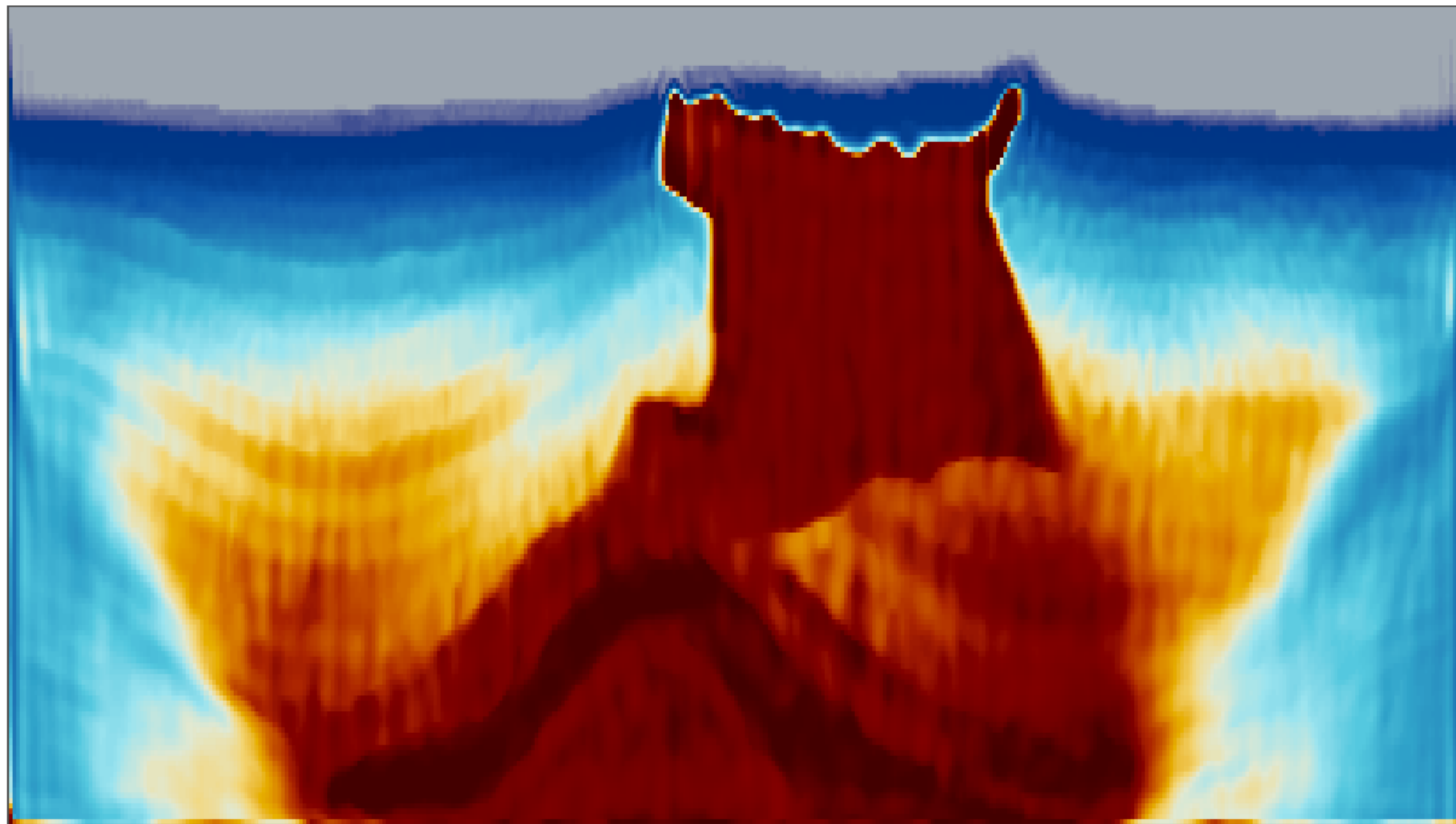
Adjoint-state – w/ TV-norm & hinge-loss projections



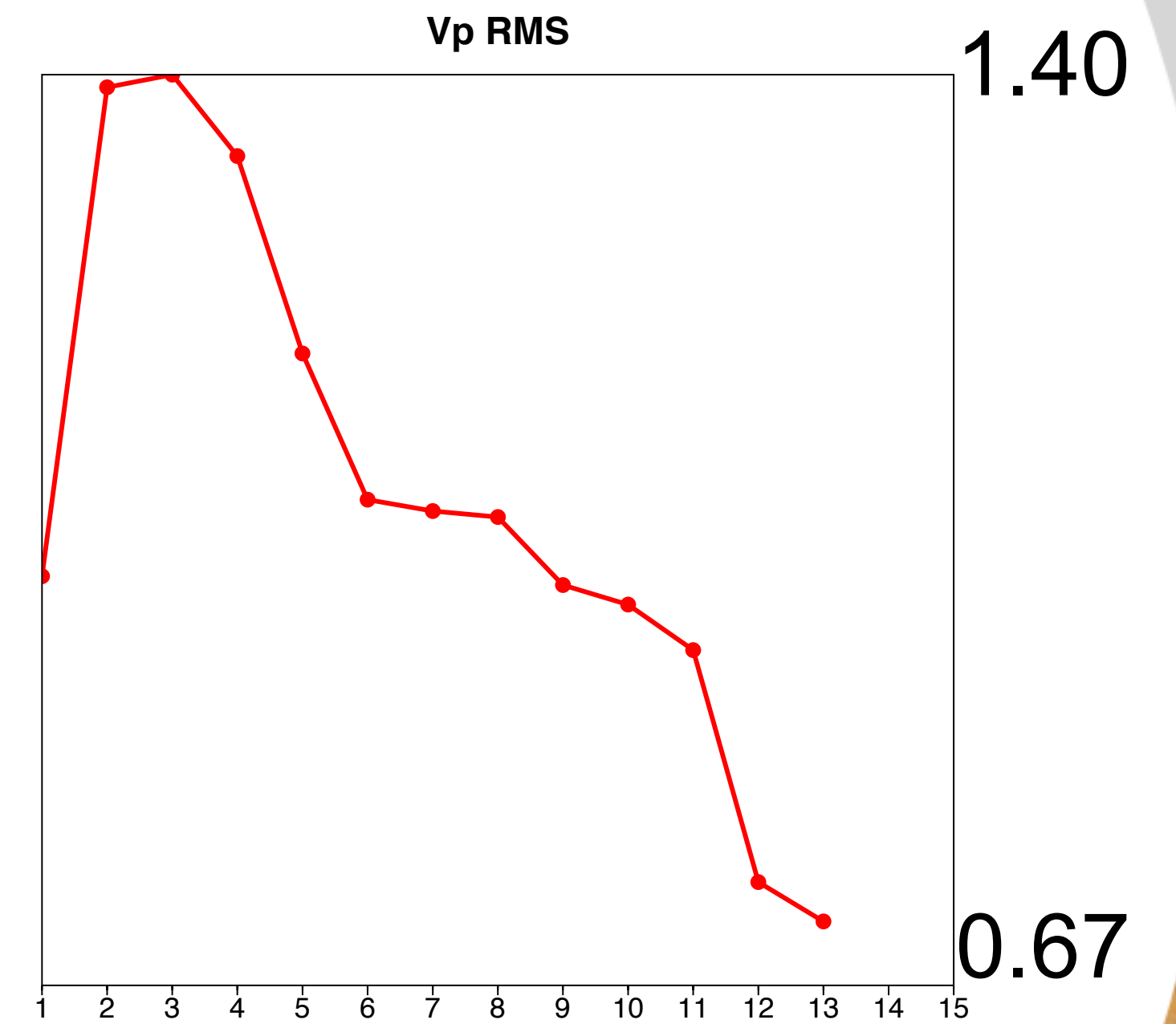
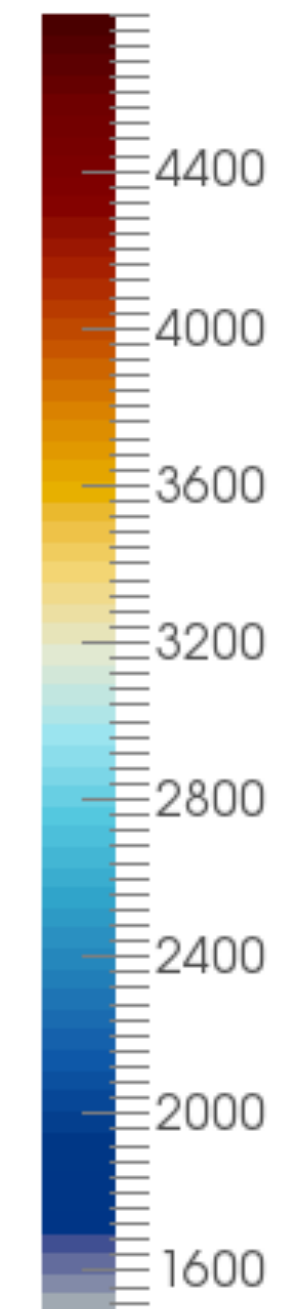
Adjoint-state – w/ TV-norm & hinge-loss projections



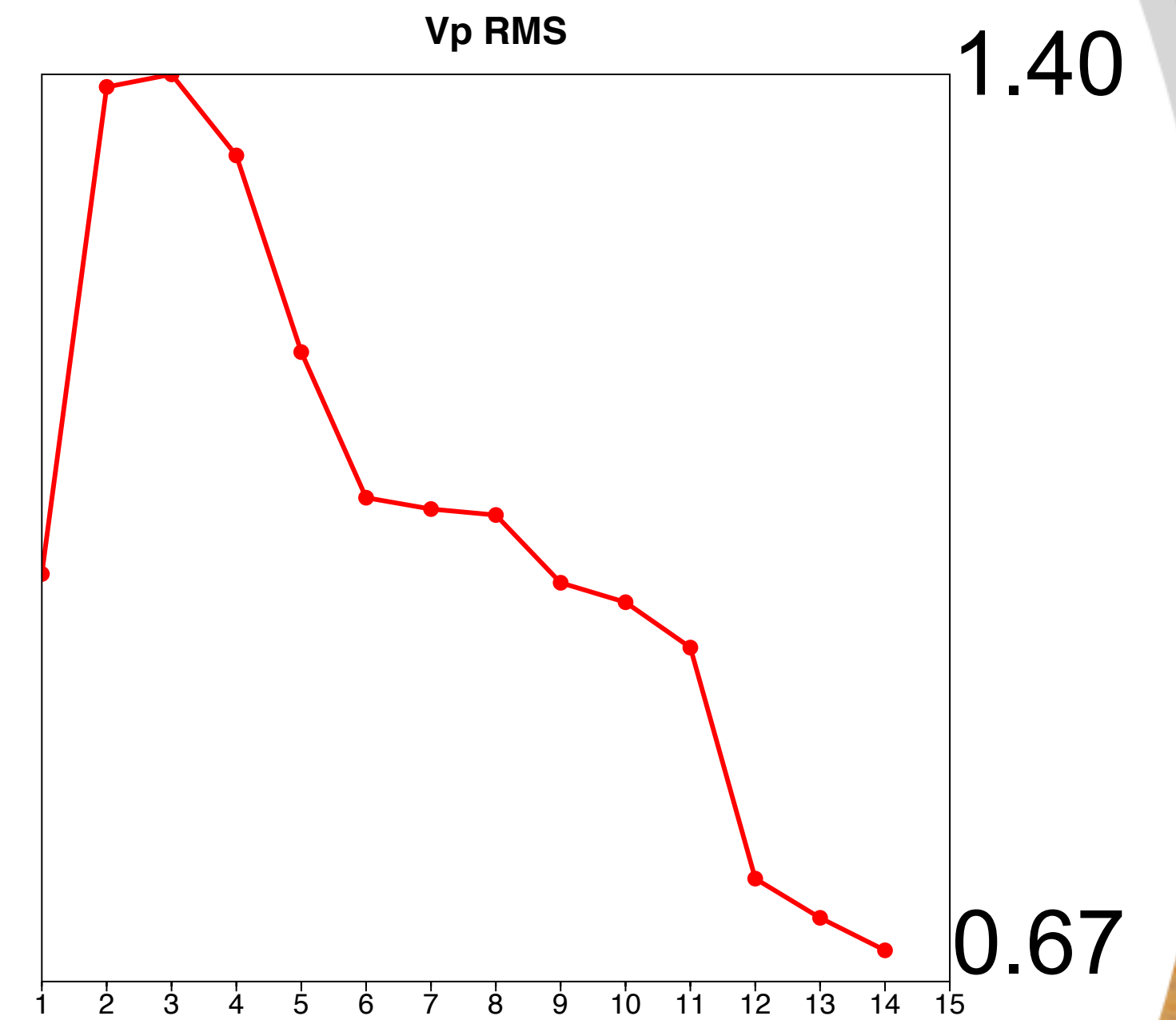
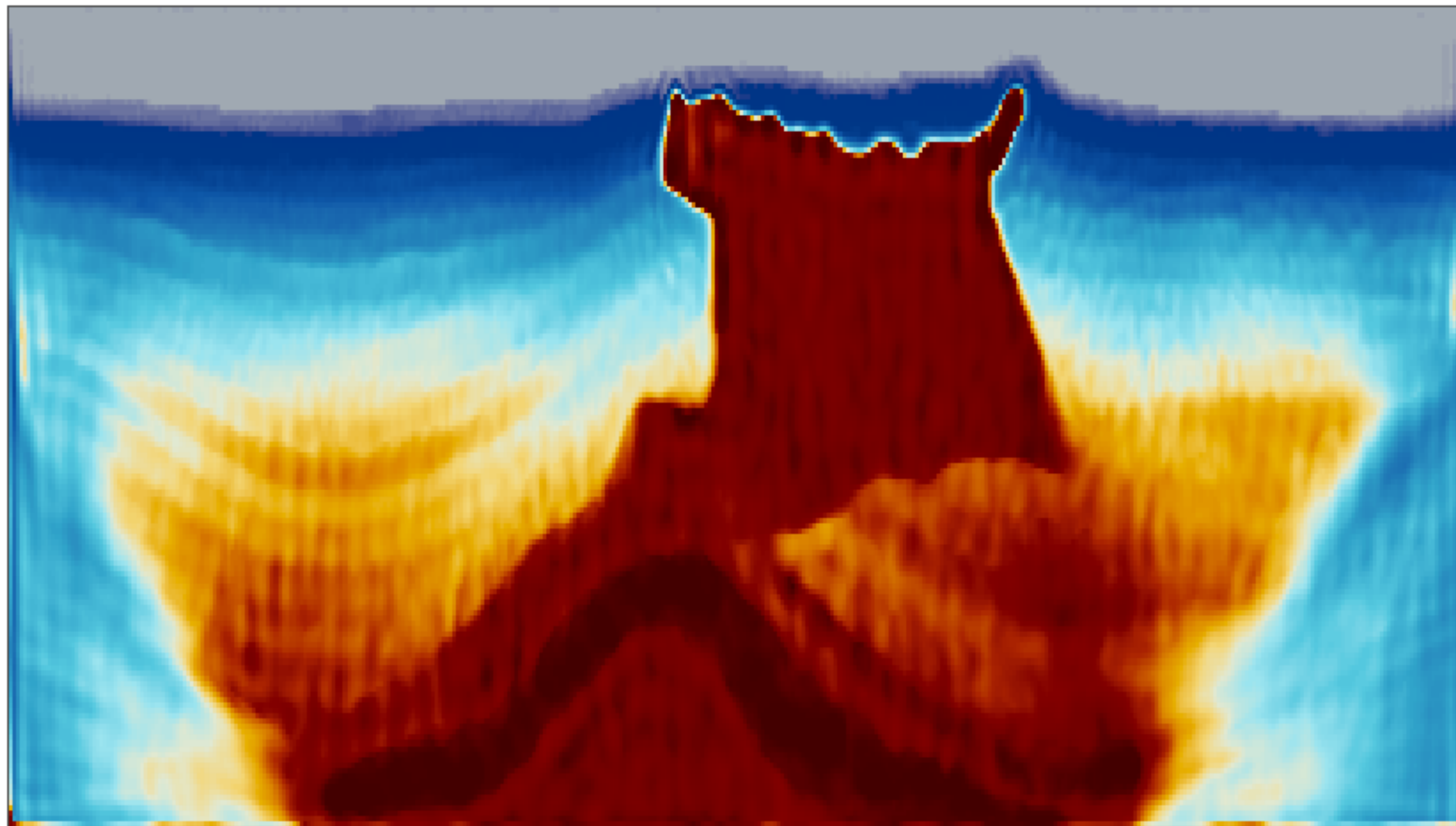
Adjoint-state – w/ TV-norm & hinge-loss projections



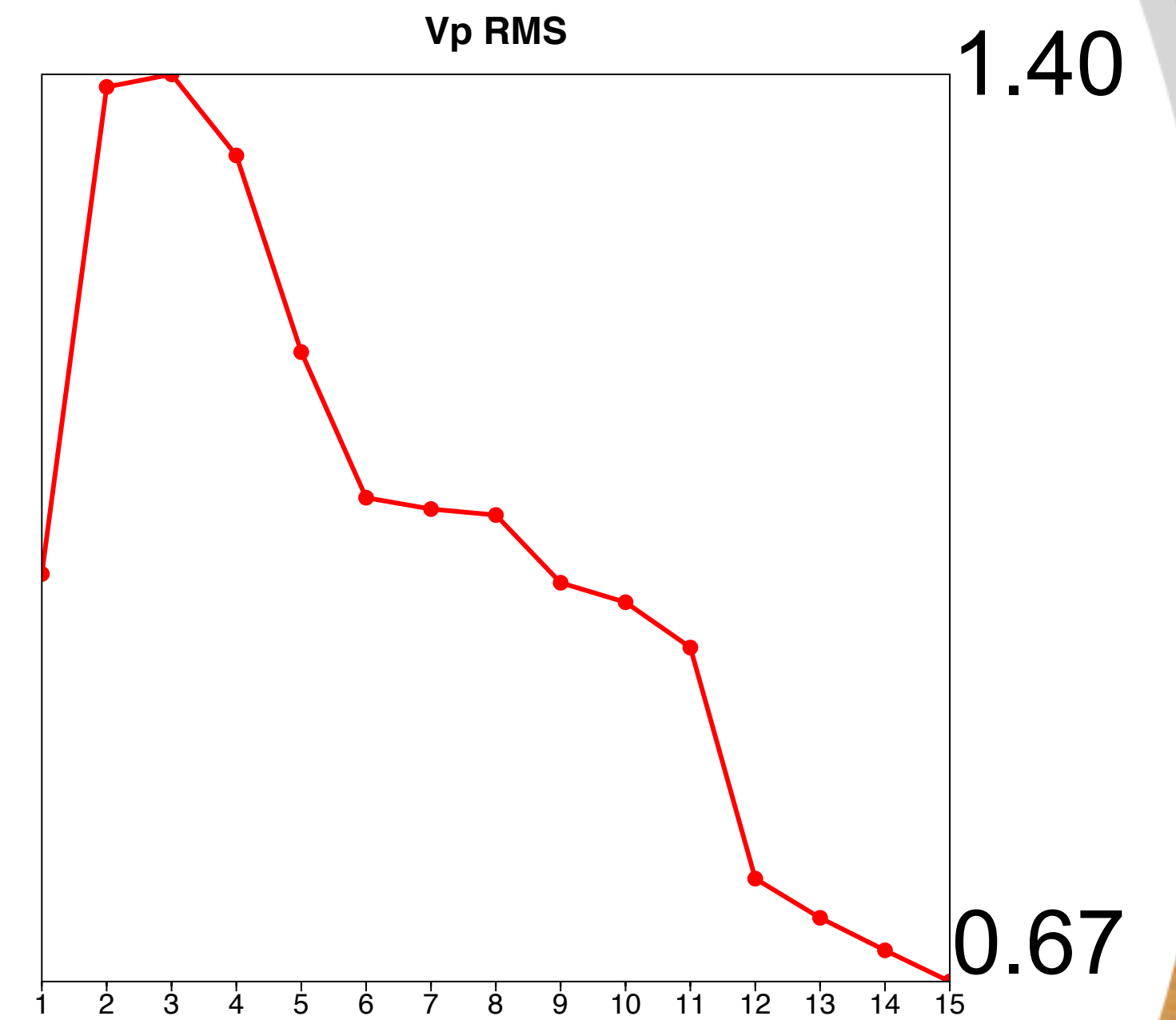
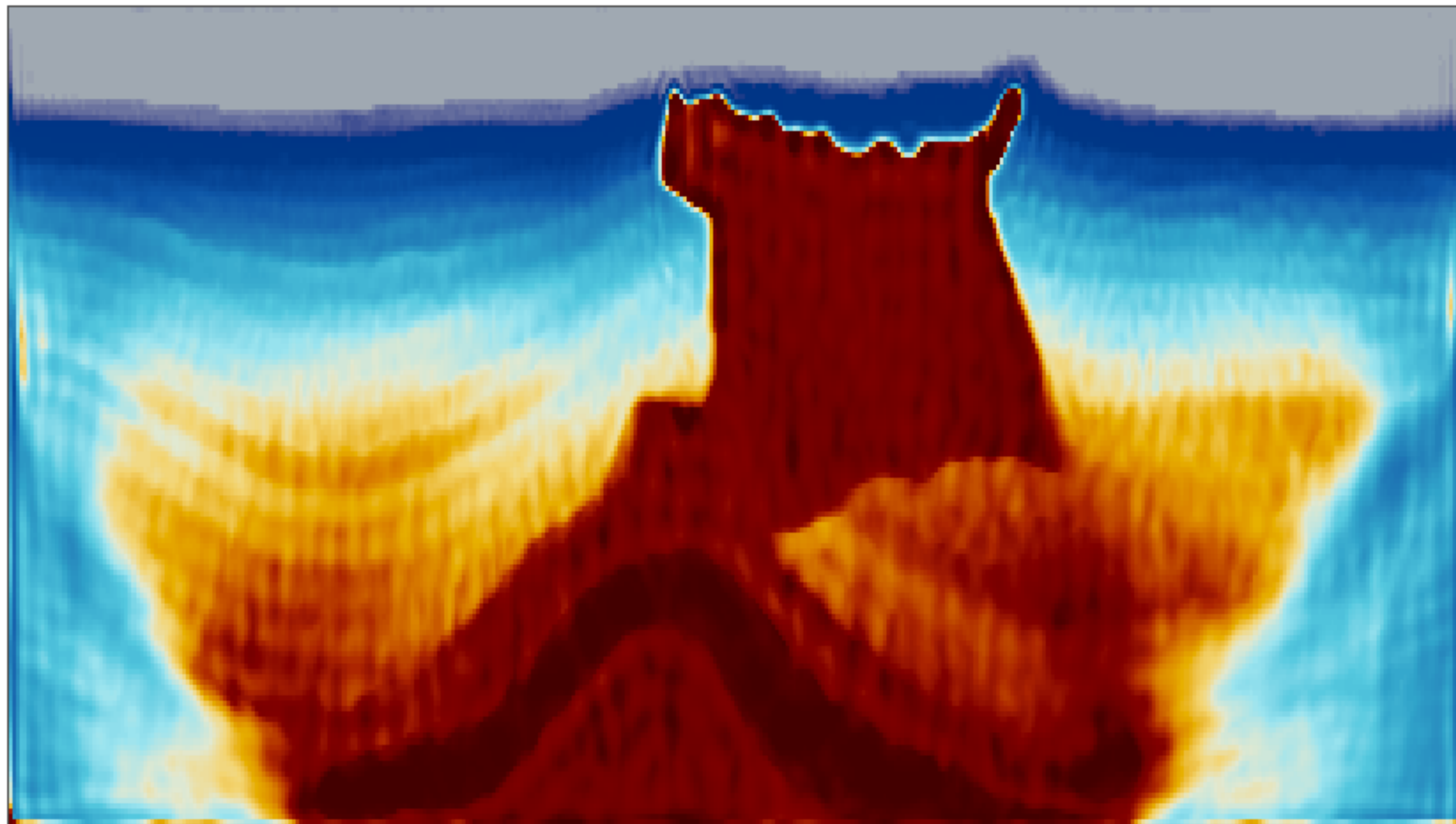
Vp (m/s)



Adjoint-state – w/ TV-norm & hinge-loss projections

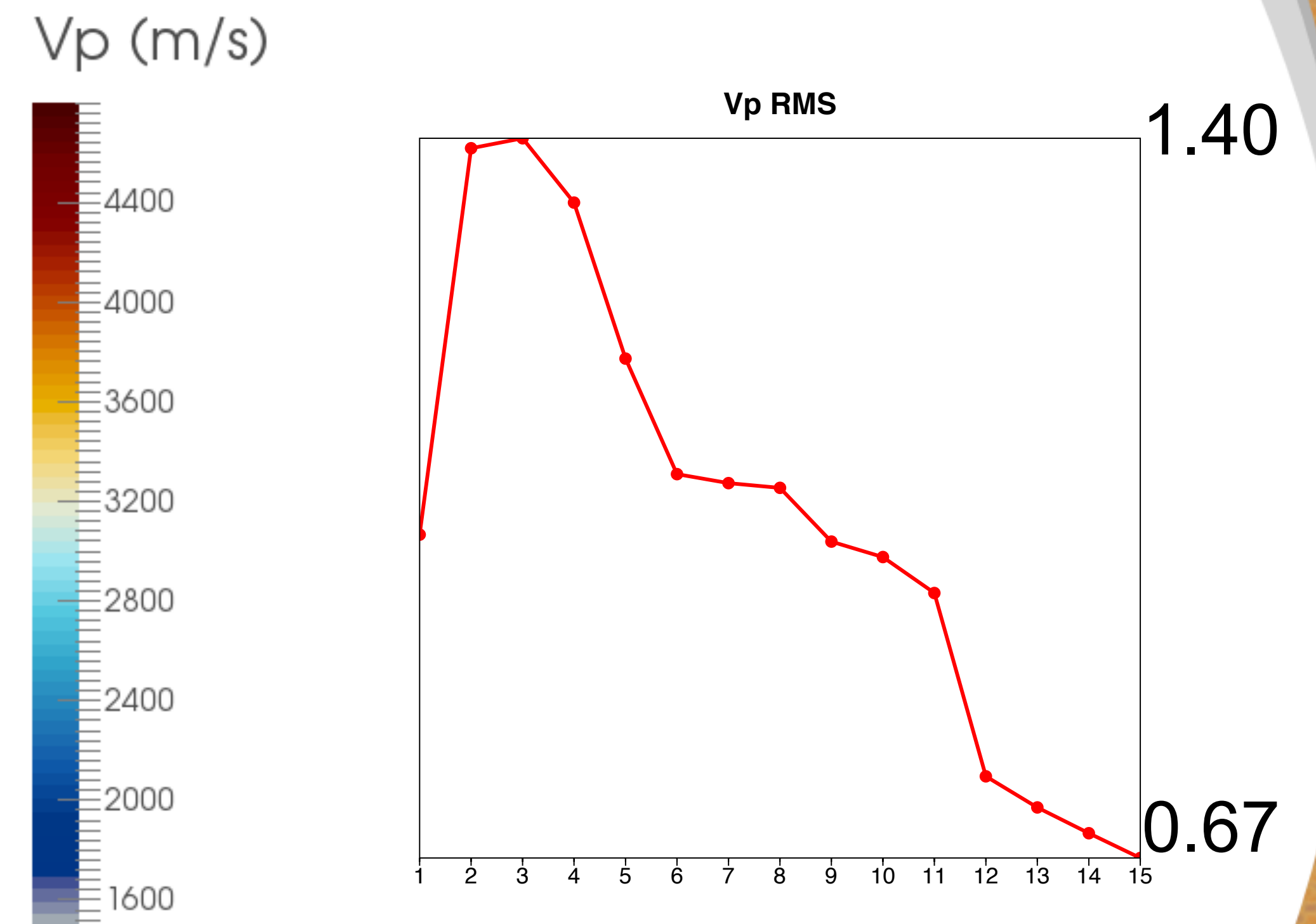
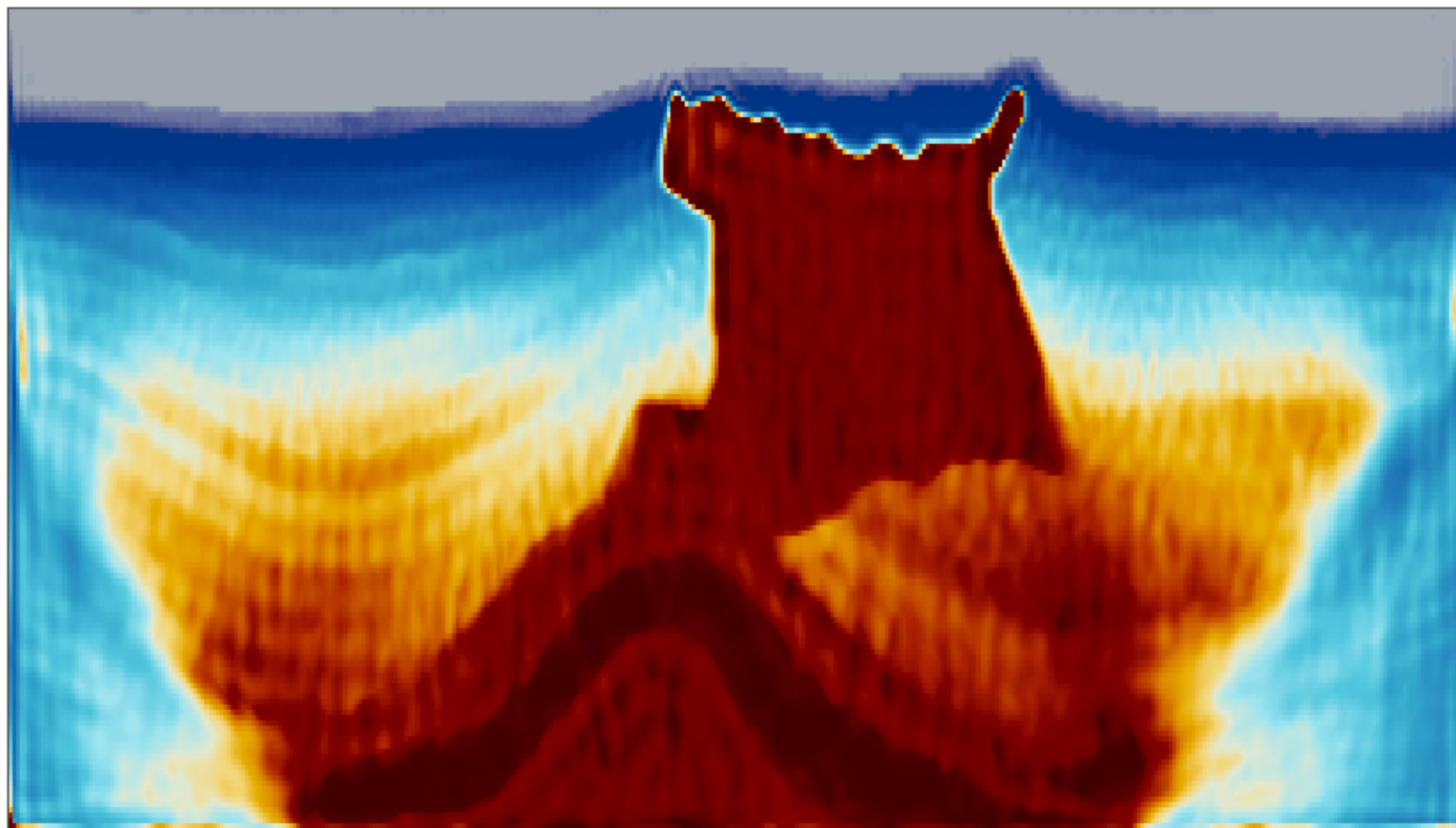


Adjoint-state – w/ TV-norm & hinge-loss projections



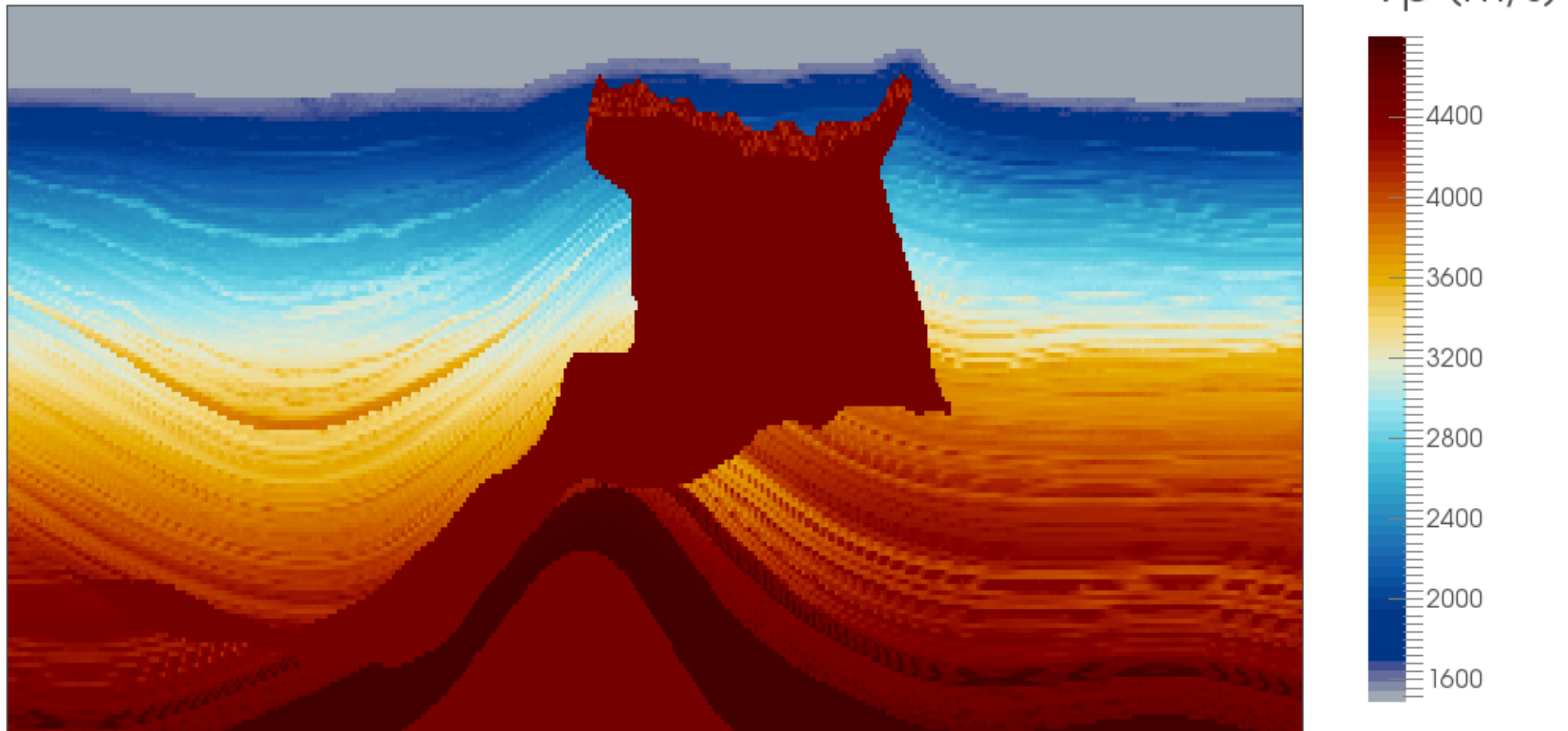
Adjoint-state – w/ TV-norm & hinge-loss projections

Final model



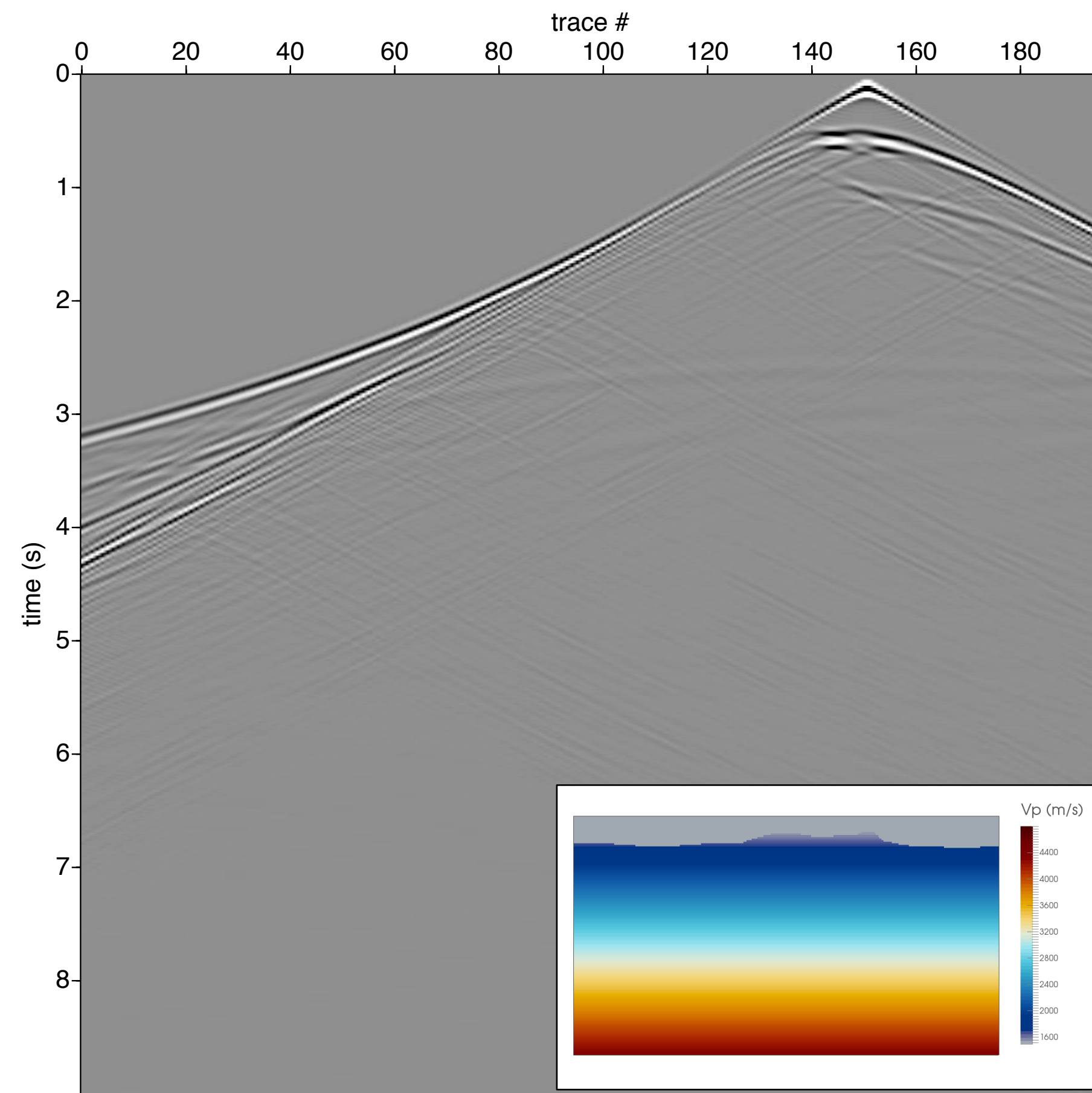
Adjoint-state – w/ TV-norm & hinge-loss projections

True model



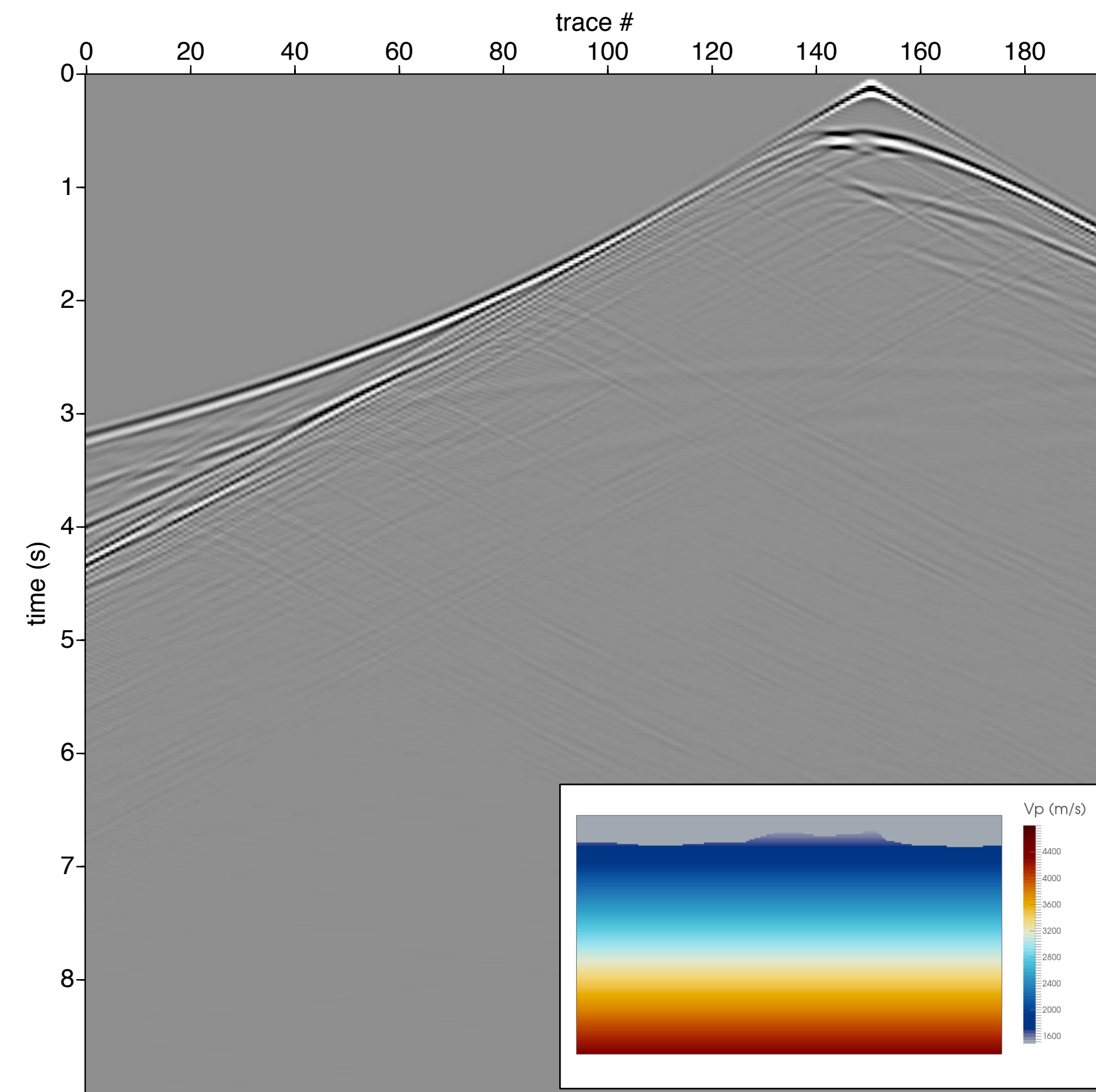
Adjoint-state – data comparisons

starting model



Starting model data

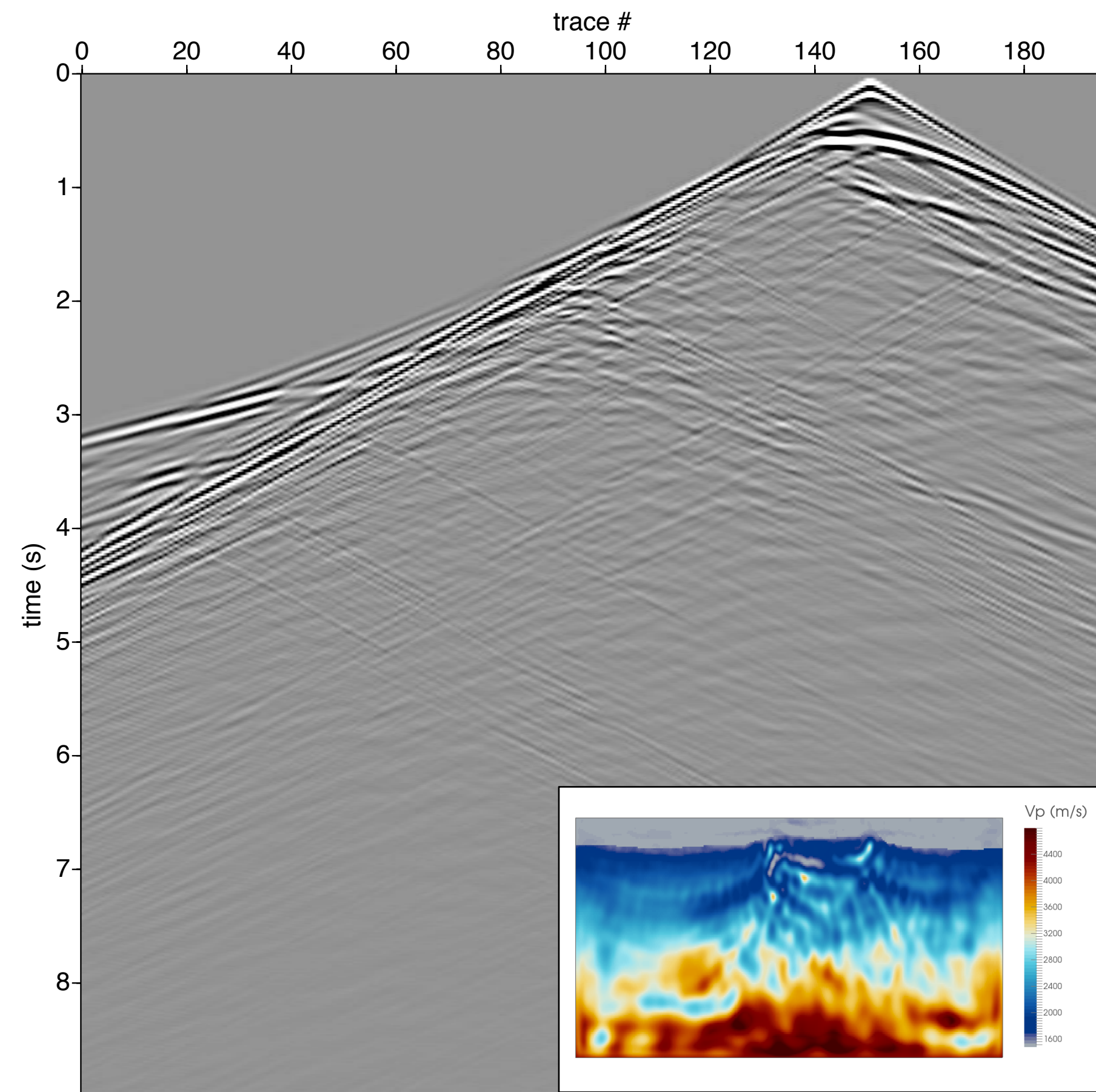
starting model



Starting model data

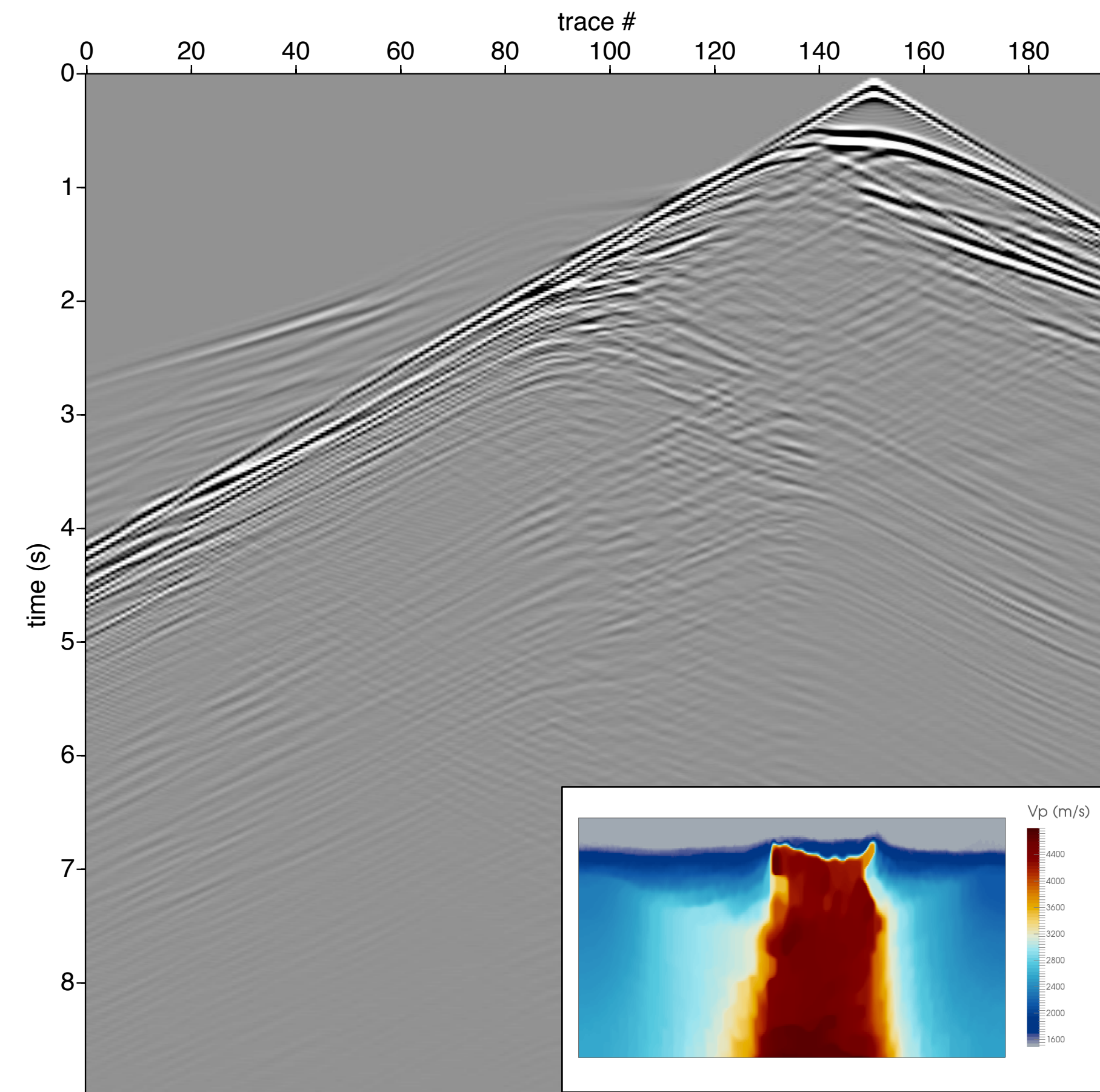
Adjoint-state – data comparisons

w/o TV & Hinge



FWI intermediate result data

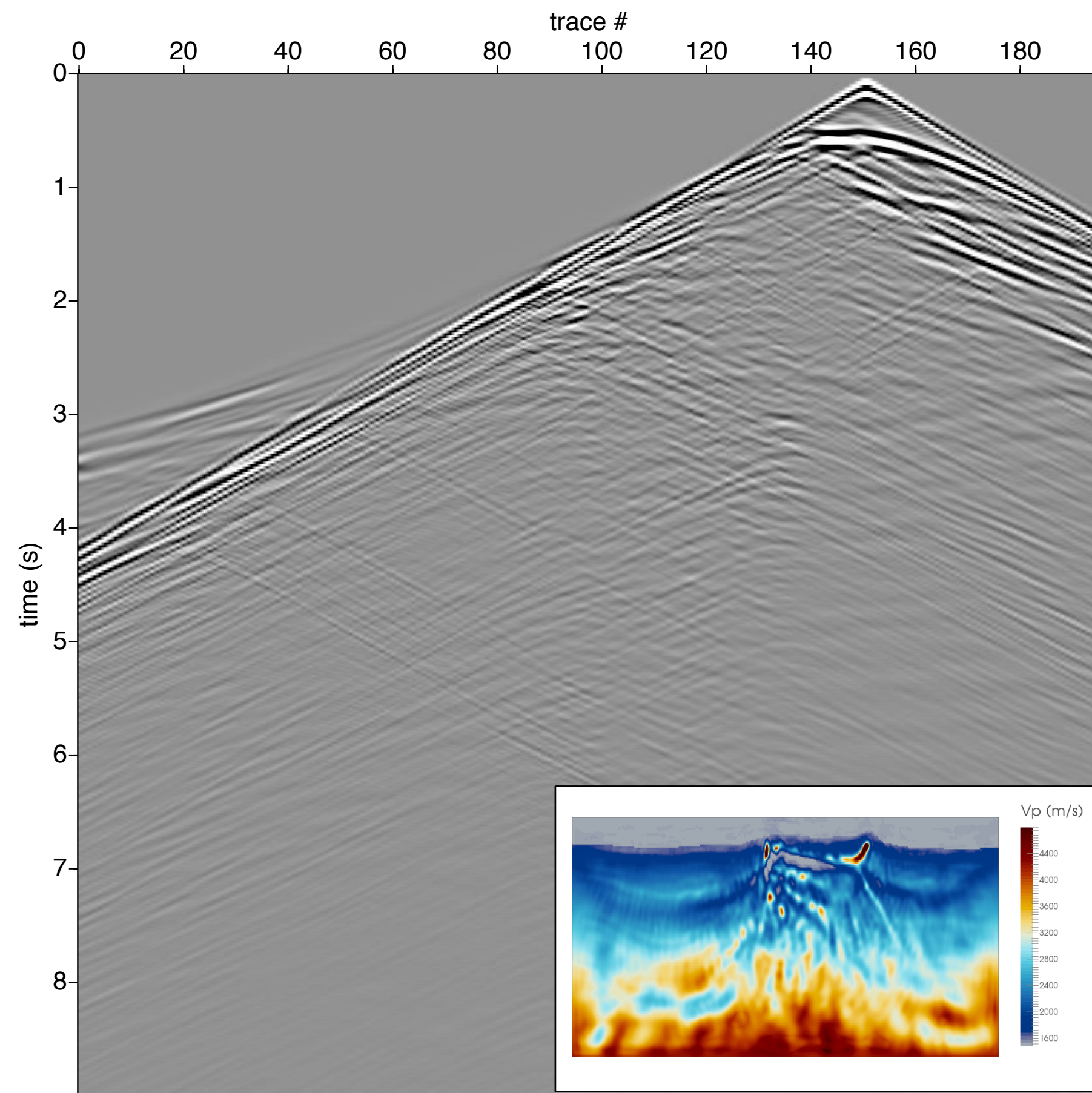
w/ TV & Hinge



FWI+TV middle result data

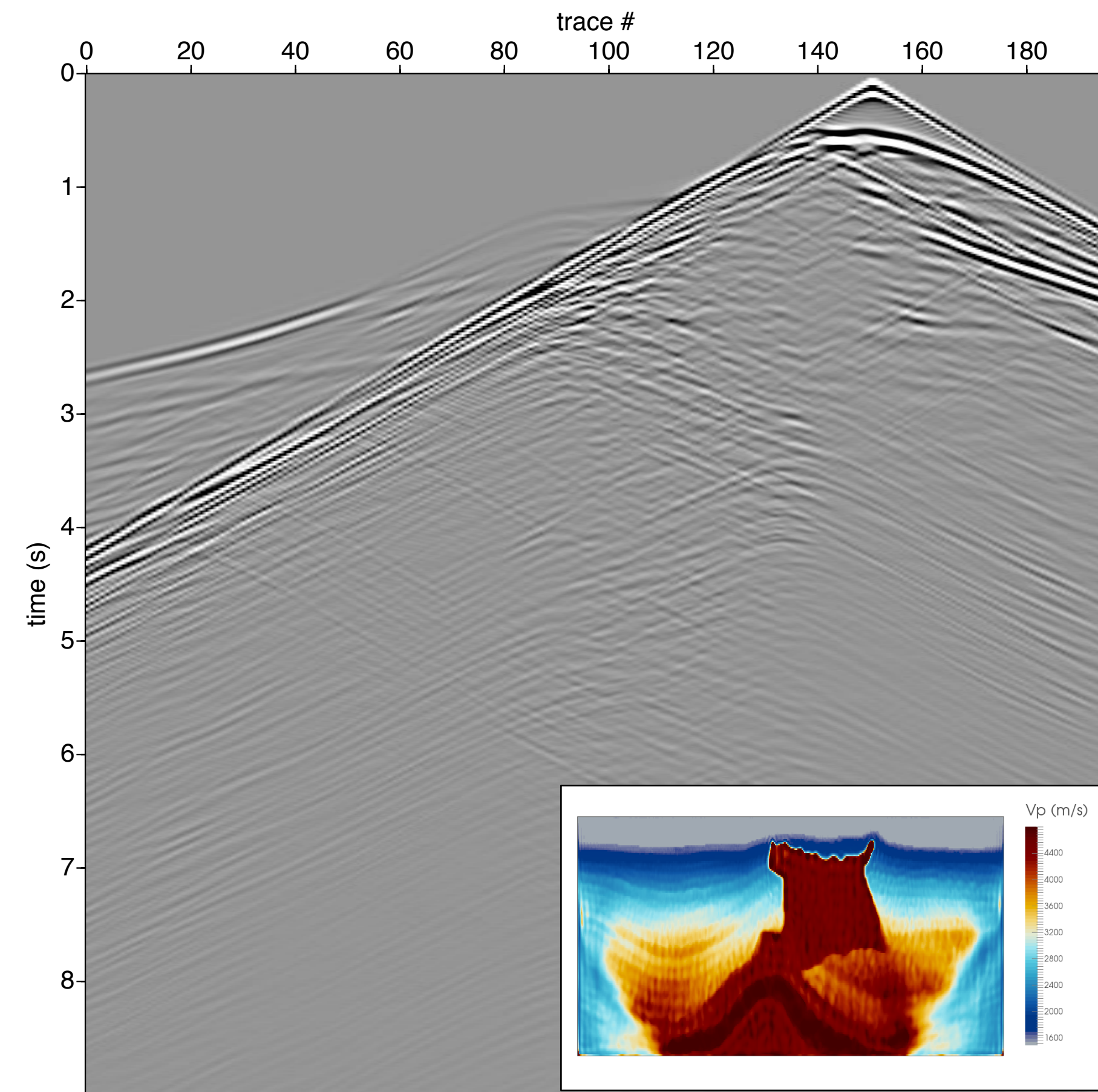
Adjoint-state – data comparisons

w/o TV & Hinge



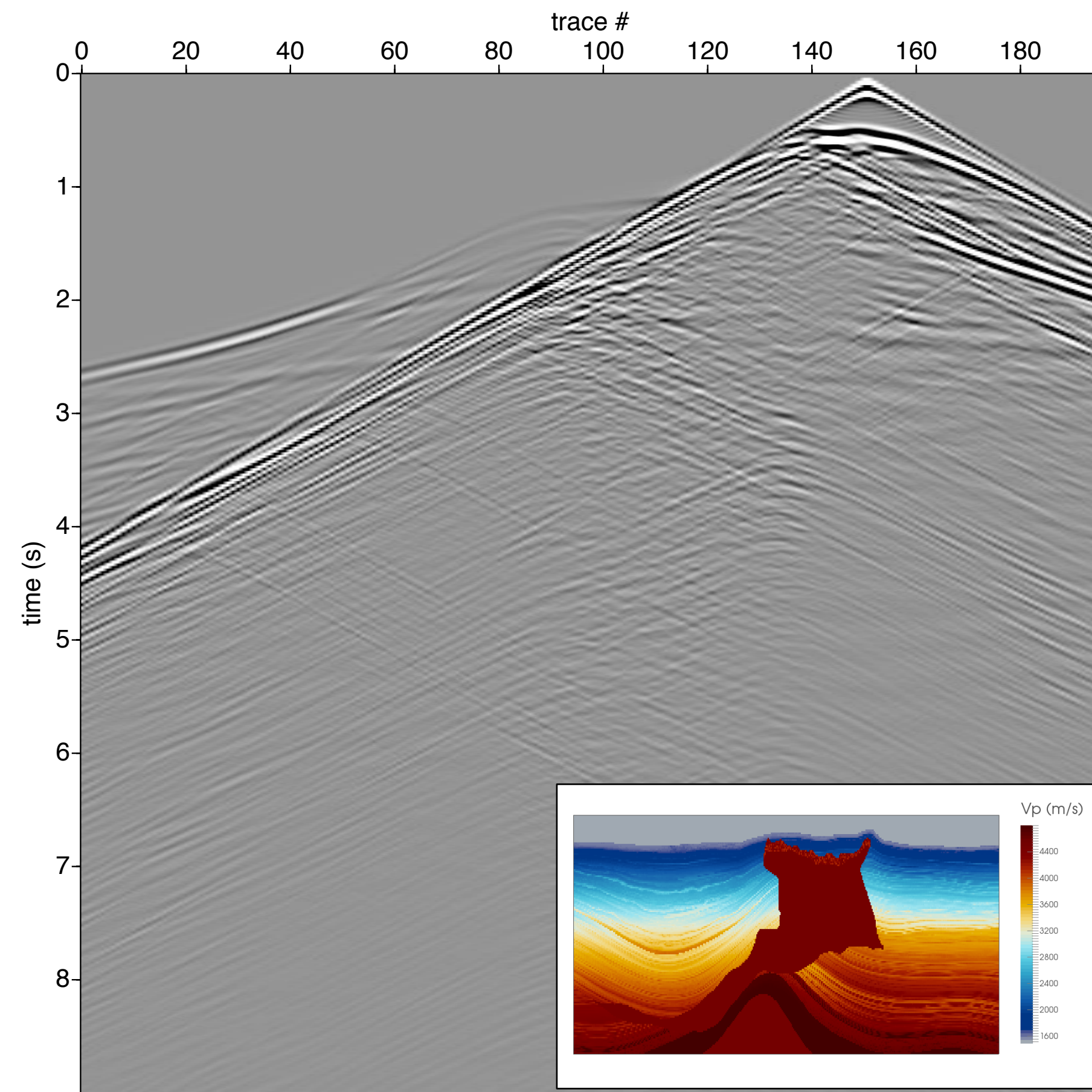
FWI final result data

w/ TV & Hinge

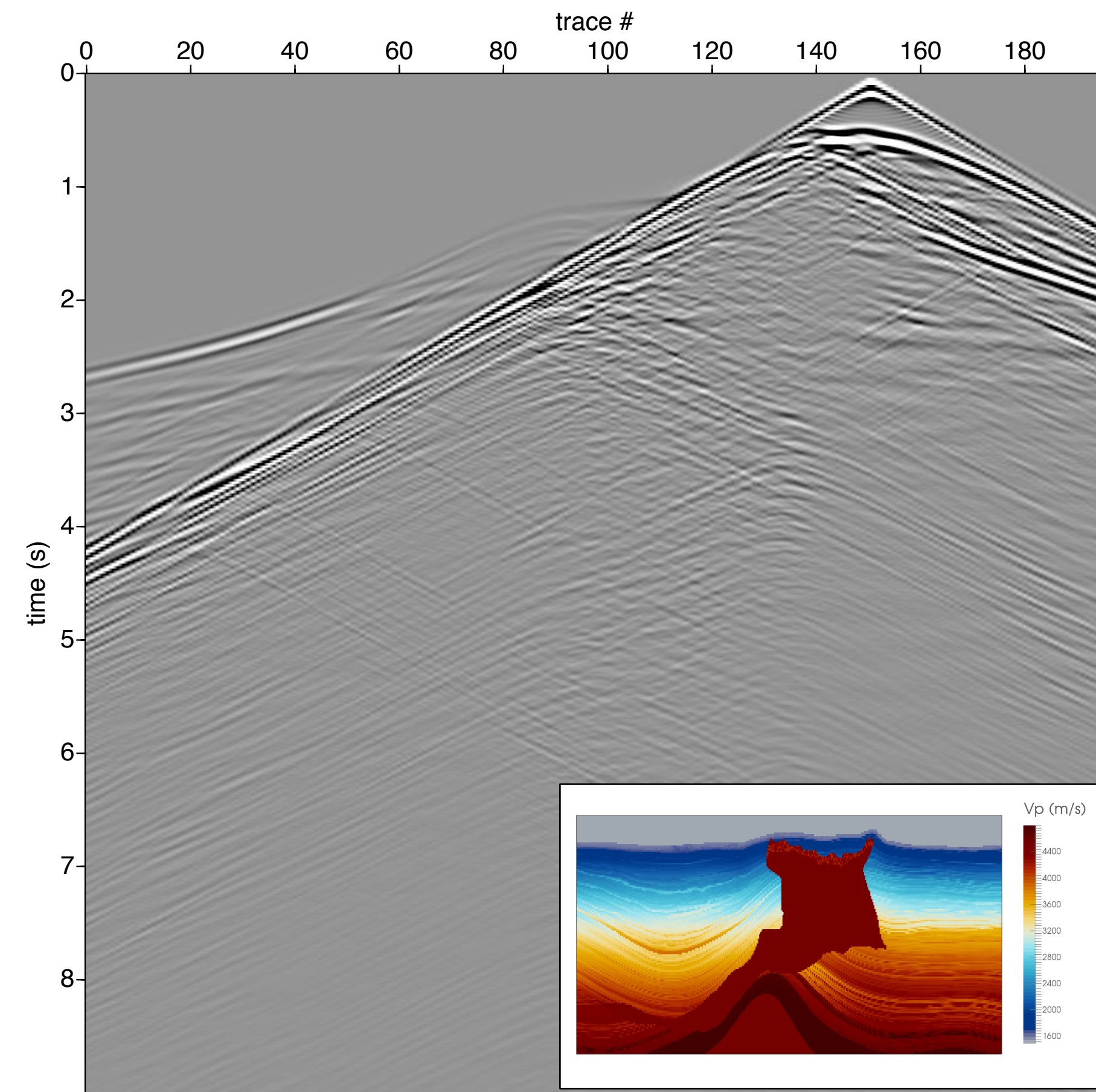


FWI+TV final result data

Adjoint-state – data comparisons



Field data



Field data

Conclusions

Constraints are enabled by WRI

- ▶ projected “second order” scheme w/ diagonal Gauss-Newton Hessian

Successful combination of

- ▶ multiple warm-started frequency cycles & relaxing constraints

Optimization framework w/ hinge-loss function on raw input data leads to

- ▶ recovery of salt, including top & bottom & sub-salt details (lows/highs)
- ▶ automatic salt flooding w/o picking & tomography
- ▶ cheap, reproducible & transparent workflows w/o handholding



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, DownUnder GeoSolutions, Hess, Petrobras, PGS, Subsalt Ltd, WesternGeco, and Woodside.