

OpenSeisML: Open Large-Scale Real Seismic and well-log Dataset for Generative AI

Ipsita Bhar¹, Huseyin Tuna Erdinc¹, Thales Souza¹, Charles Jones² and Felix J. Herrmann¹

¹Georgia Institute of Technology, Atlanta, USA ²Osokey Ltd, Henley-on-Thames, UK

SLIM



Session Title: Agentic Artificial Intelligence and Foundation Models

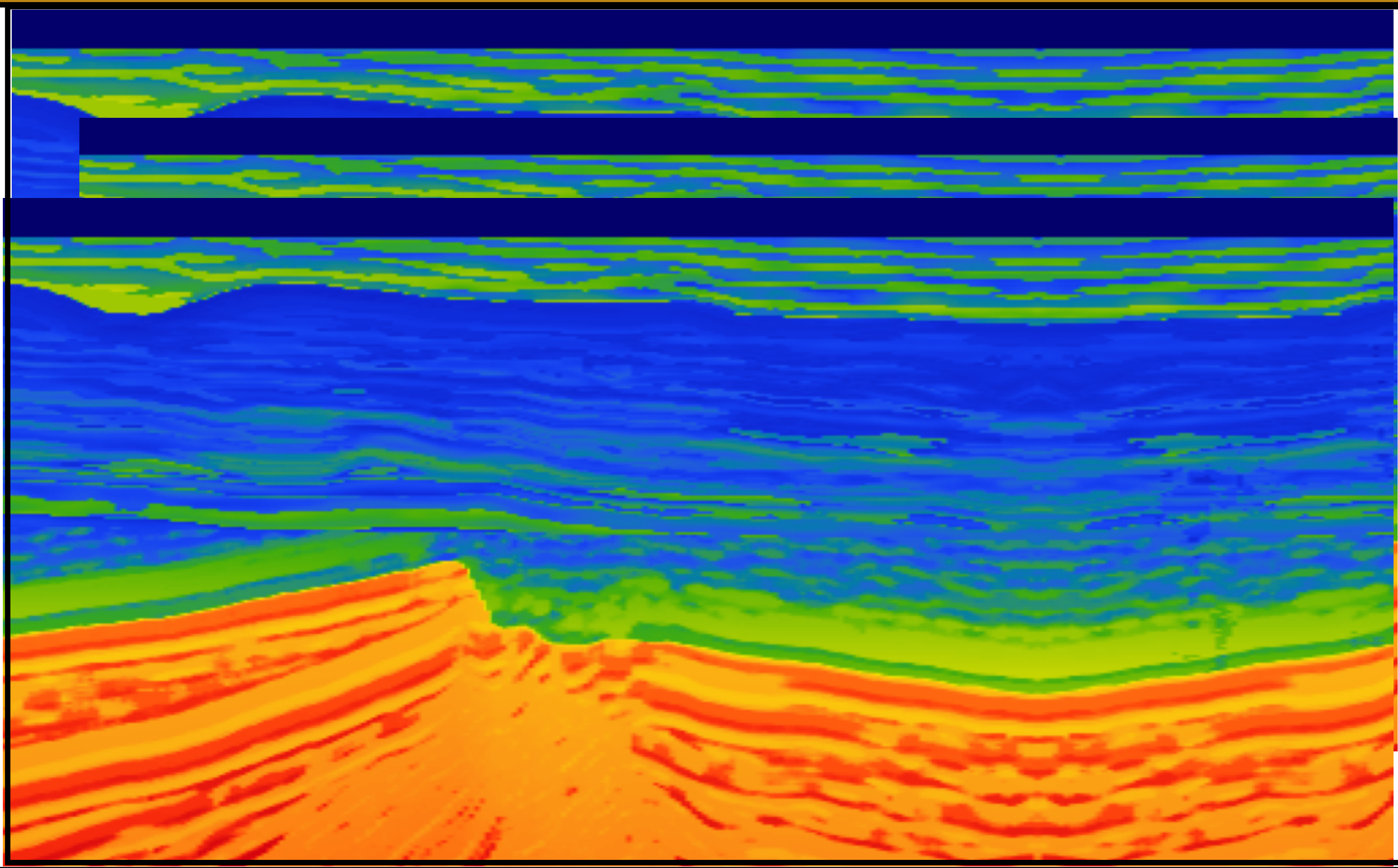
Session ID: DMLA P12

August 20, 2026

Time: 10:20 AM to 12:00 PM

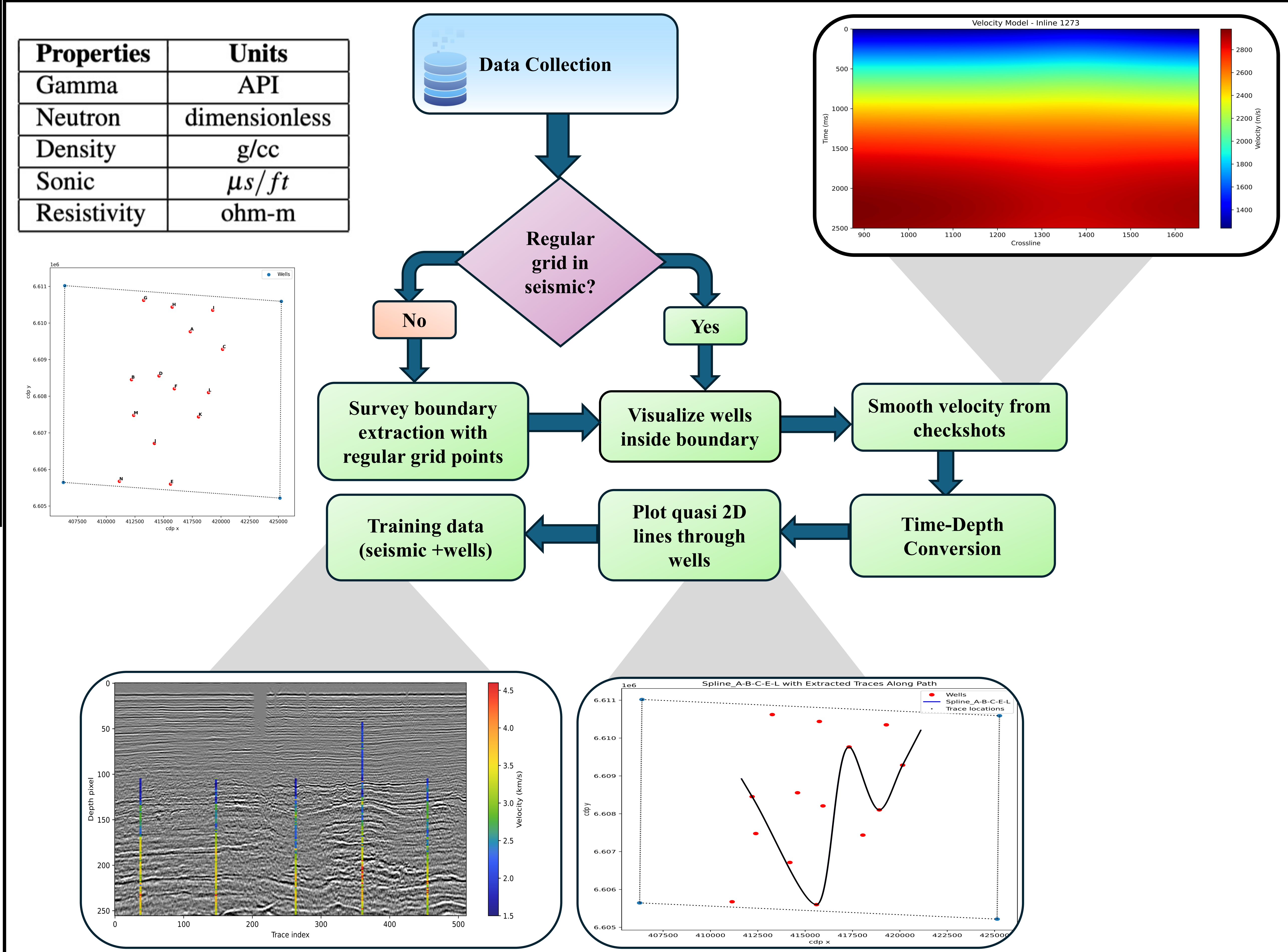


Introduction and Motivation



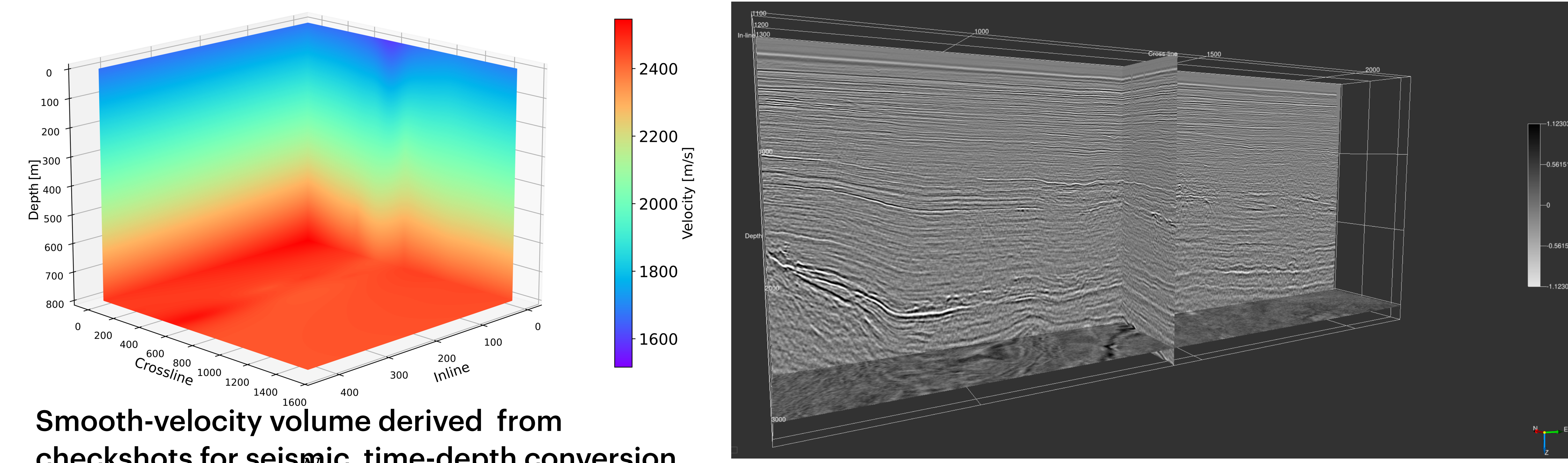
- Lack of publicly available realistic seismic property models for training generative ML-models
- Existing synthetic property models often lack realistic complexity like Compass model on the left
- Synthetic models often fail to capture subsurface heterogeneity
- Networks trained on synthetic models may fail when inferred on real data
- Real seismic data is not public not available
- Hence the need for real seismic data to train ML-models

Automated Data Pre-processing



Properties	Units
Gamma	API
Neutron	dimensionless
Density	g/cc
Sonic	$\mu s/ft$
Resistivity	ohm-m

Training Data: Seismic Time-Depth

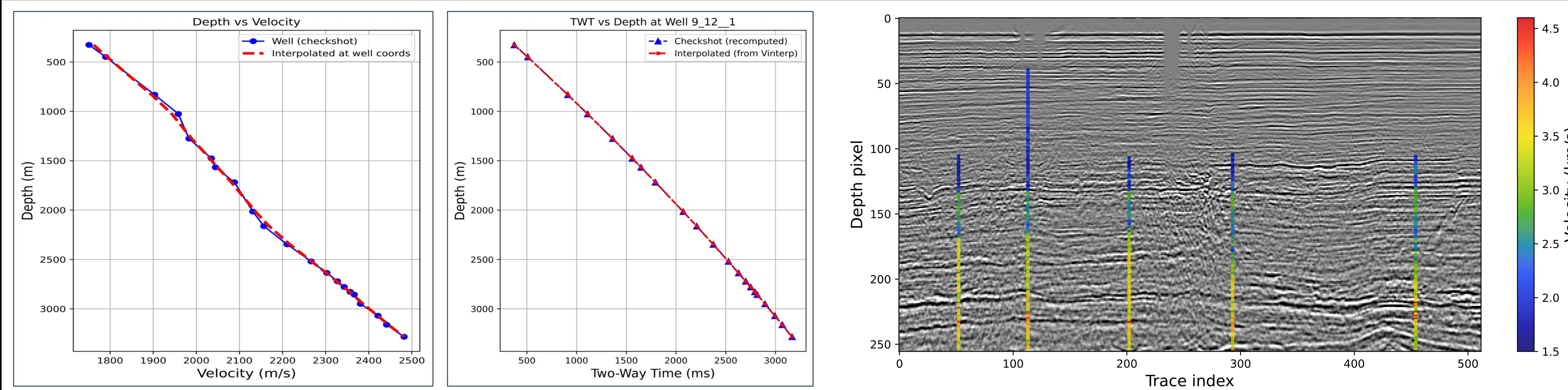


Smooth-velocity volume derived from checkshots for seismic time-depth conversion

$$f(x) = \sum_{i=1} \lambda_i \phi \left(\| x - x_i \| \right)$$

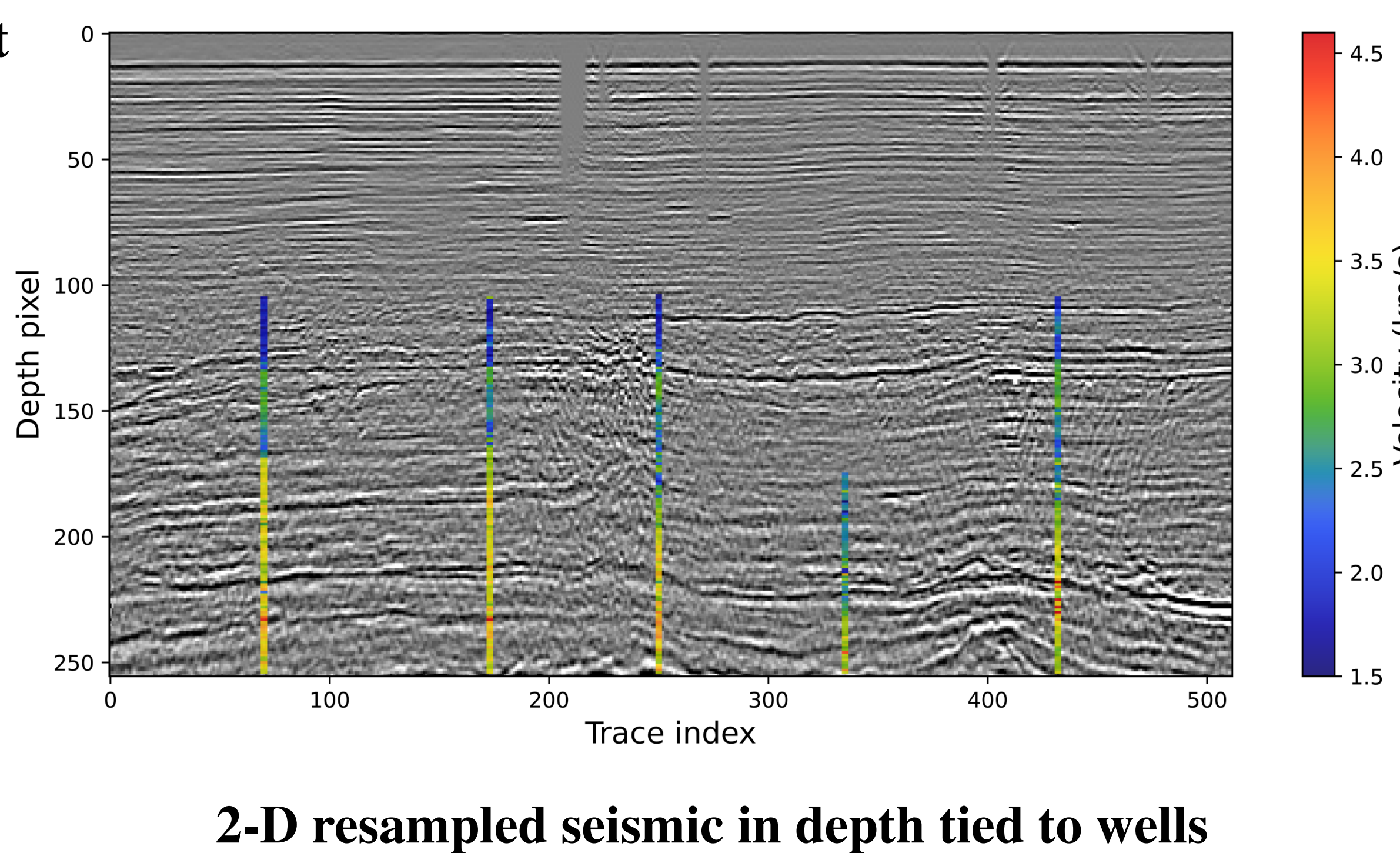
- use check shots to convert time-migration to depth
- smooth velocity ensures smooth velocity field and also interval velocities measured are often inaccurate

QC: Seismic Well Tie



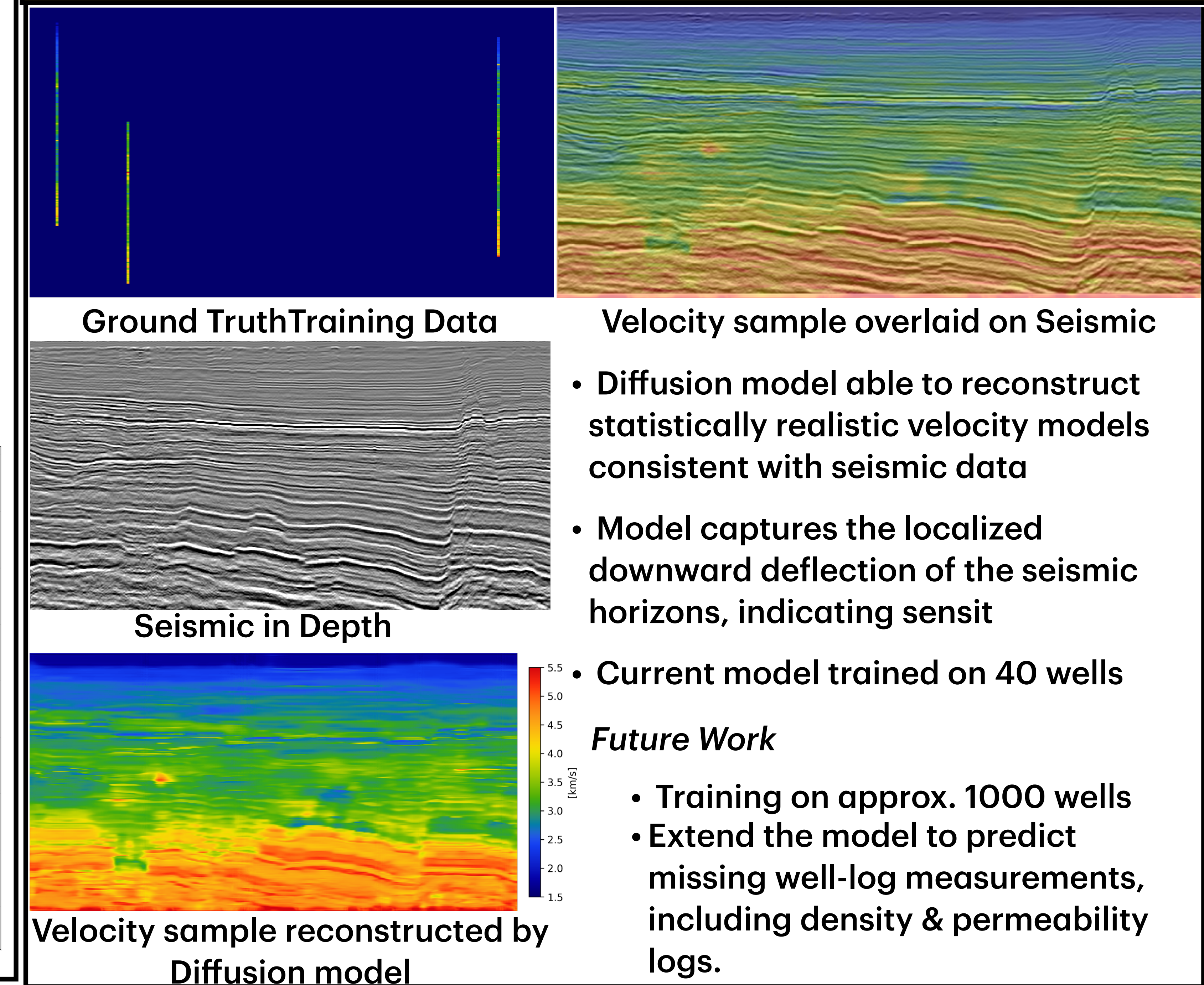
TWT vs depth and velocity vs depth for interpolated (red) vs checkshot (blue)

- **Seismic-well tie:** Velocity transitions along the wells align with major seismic reflectors and stratigraphic boundaries
- **Consistent horizon correspondence:** Similar velocity ranges occur at comparable seismic horizons across spatially separated wells
- **Lateral geological consistency:** The well-derived velocity variations follow lateral continuity and structural trends observed in the seismic image



2-D resampled seismic in depth tied to wells

Results & Future Work



Velocity sample overlaid on Seismic

Model captures the localized downward deflection of the seismic horizons, indicating sensit

Current model trained on 40 wells

Future Work

- Training on approx. 1000 wells
- Extend the model to predict missing well-log measurements, including density & permeability logs.

References

a) dGB Earth Sciences, 2026, **OpenText Pro & dGB Plugins Documentation, Version 7.0.** Available at: https://doc.opentext.org/7.0.0/doc/dgb_userdoc/Default.htm. Accessed February 12, 2026.

b) Contains information provided by the North Sea Transition Authority and/or other third parties.

c) Chen, H., J. Chen, M. D. Sacchi, J. Gao, and P. Yang, 2025, "Unsupervised Seismic Acoustic Impedance Inversion Based on Generative Diffusion Model," *Geophysics*, 90, M109–M121.

d) Donoho, D., 2024, "Data Science at the Singularity," *Harvard Data Science Review*, 6. Available at: <https://hdsr.mitpress.mit.edu/pub/g9mau4m0>.