Synthetic Geology: Structural Geology Meets Deep Learning

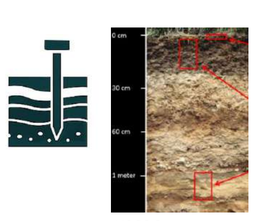
Simon Ghyselincks [1,2] **Valeriia Okhmak** [3] **Stefano Zampini** [3]
George Turkiyyah [3] **David Keyes** [3] **Eldad Haber** [1]

[1]University of British Columbia [2]Vector Institute [3]KAUST

Paper + Code

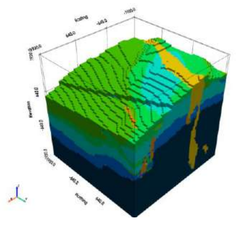
A Hidden Earth

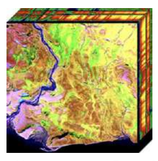
Geological observed data is highly heterogenous, multi-modal, and incomplete. There is no known **ground truth** for **structural geology** models.



USGS Survey

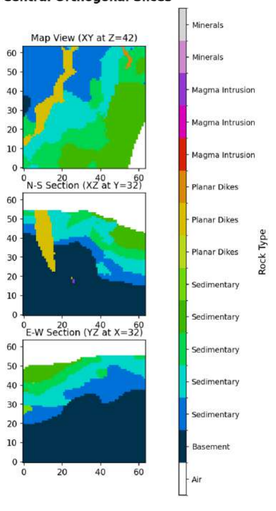
Ill-Posed Inverison





Hyperspectral Data, NASA

GTK Finland from Nironen et al. (2016)



This framework deals with **discrete categorical rock types** to make probabilistic reconstructions of the earth from observed data.

Inverse Problems

Classical Regularized Inversion:

Seek a single maximum likelihood solution using Maximum a Posteriori

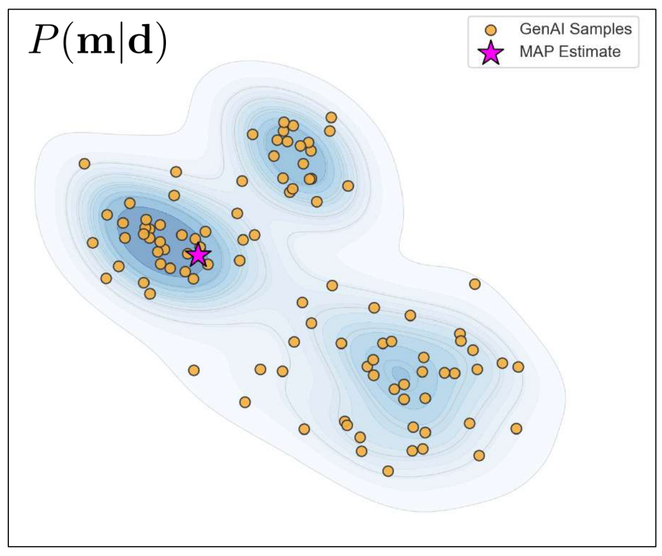
$$\mathbf{A}\mathbf{m} + \epsilon = \mathbf{d}$$

$$P(\mathbf{m}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}$$

$$\hat{\mathbf{m}}_{\text{MAP}} = \arg\max_{\mathbf{m}} \left[P(\mathbf{d}|\mathbf{m})P(\mathbf{m})\right]$$

$$\phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{A}\mathbf{m} - \mathbf{d}\|^2 + \alpha R(\mathbf{m})$$

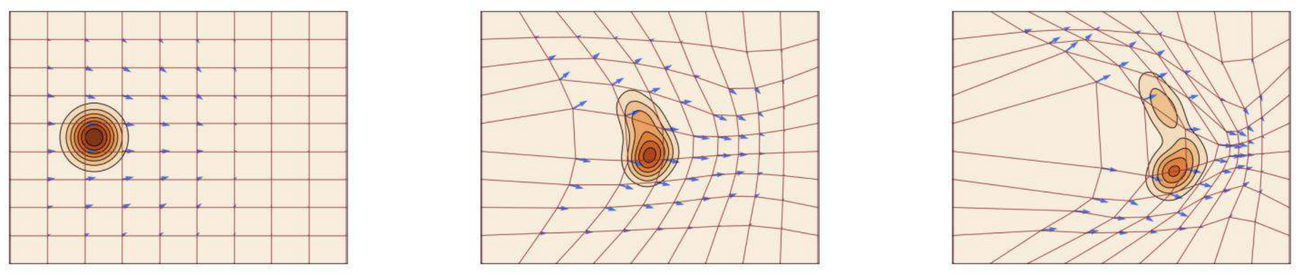
Misfit Prior



Posterior Sampling with Gen AI

Direct learning is intractable due to the partition function

$$P(\mathbf{m}|\mathbf{d}) = \frac{1}{Z(\mathbf{d})} e^{-E(\mathbf{m}|\mathbf{d})}, \quad \text{where } Z(\mathbf{d}) = \int_\Omega e^{-E(\mathbf{m}|\mathbf{d})} d\mathbf{m}$$



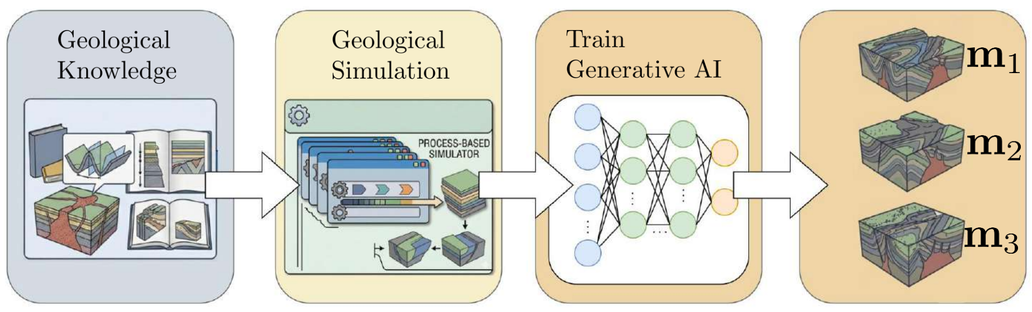
(Lipman et. al, 2024)

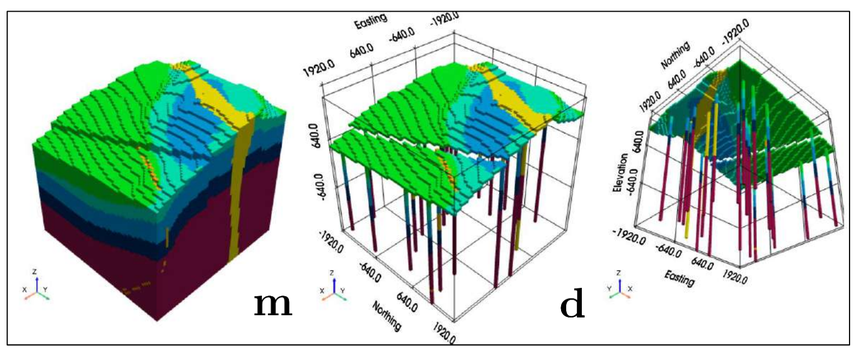
In **conditional flow matching,** the conditional probability distribution over all models is indirectly learned via a time dependent velocity field $\hat{\mathbf{v}}(\mathbf{m}, t; \boldsymbol{\xi})$ with model params $\boldsymbol{\xi}$. Discrete data must first be embedded into a continuous space.

$$\frac{d\mathbf{m}}{dt} = \hat{\mathbf{v}}(\mathbf{m}_t^{(e)}, t; \boldsymbol{\xi}) \quad \mathbf{m}_0^{(e)} \sim N(0, \mathbf{I}) \quad t \in [0, 1]$$

Data Driven Geology

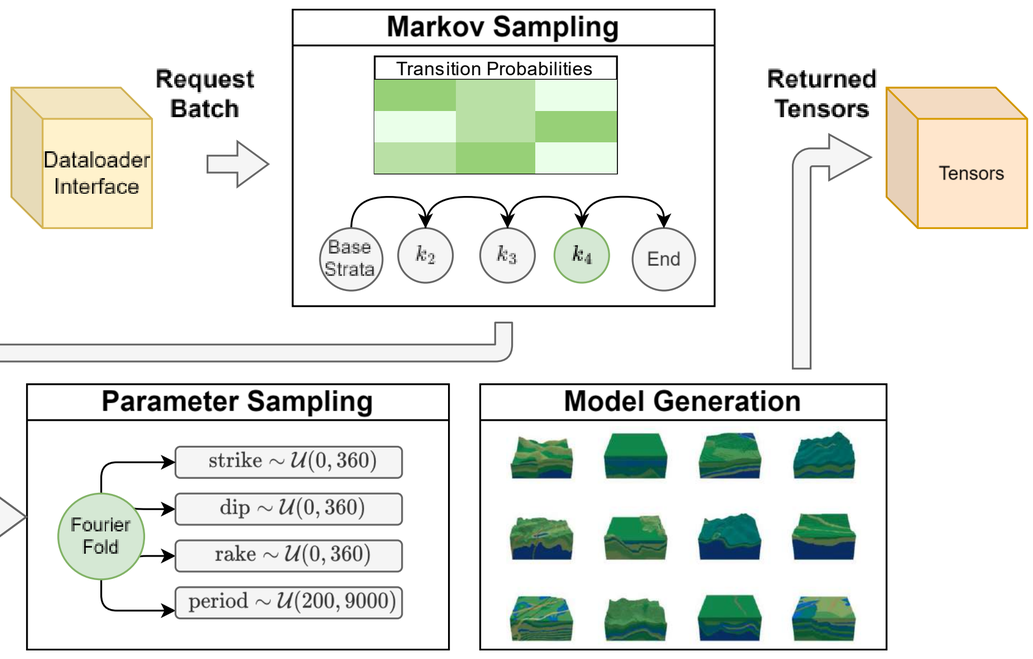
A **simulator** encodes **geological knowledge** into complete ground truth samples and observables to **train a generative model** that outputs plausible geology



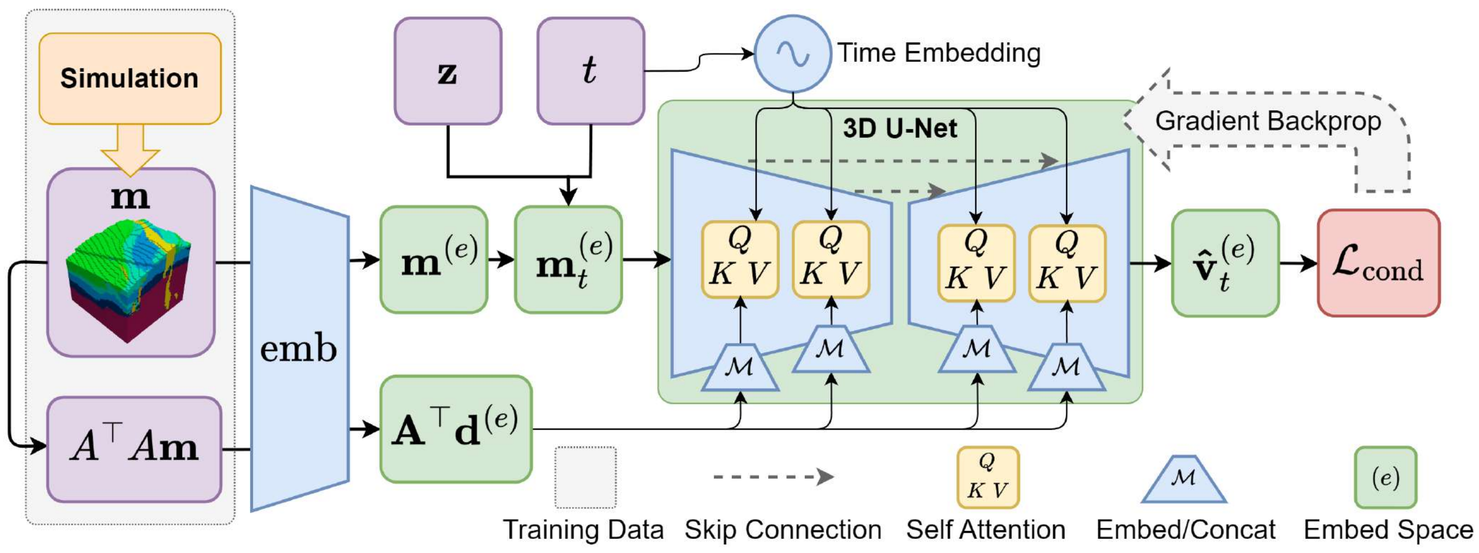


Synthetic Data Pipeline

- Unlimited sampling
- Zero disk storage
- Customizable / adaptable

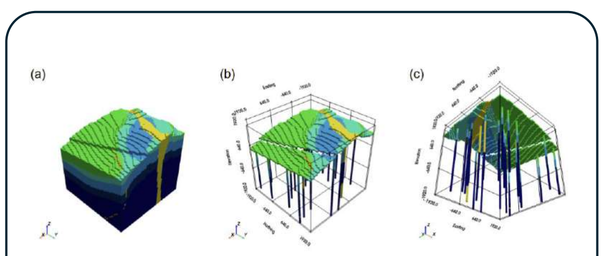


Training Architecture

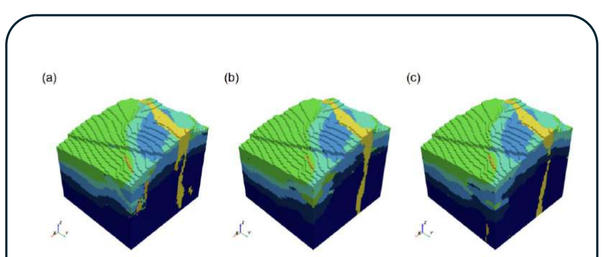


$$\mathcal{L}_{\text{cond}} = \|\hat{\mathbf{v}}(\mathbf{m}_t, \mathbf{d}, t; \boldsymbol{\xi}) - \mathbf{v}_t\|_2^2 + \lambda \cdot t \cdot \left\|\hat{\mathbf{d}} - \mathbf{d}\right\|_2^2.$$

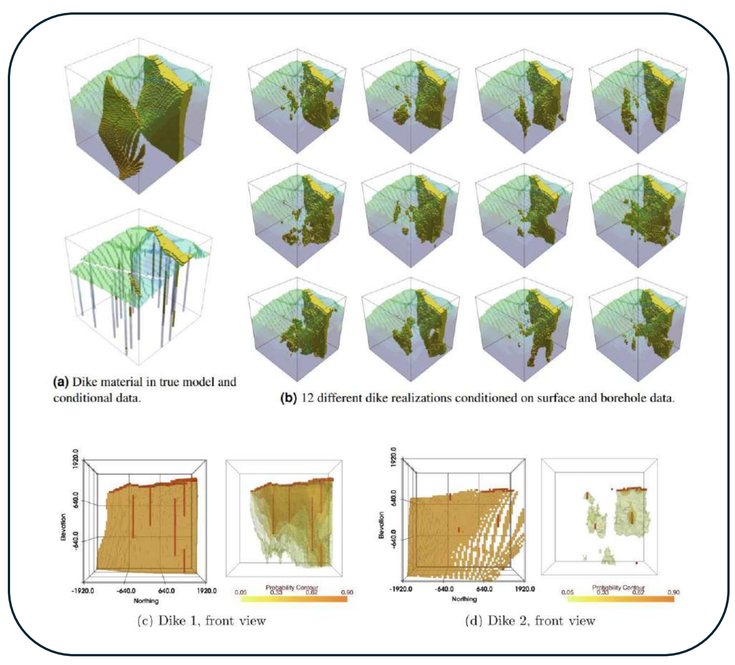
Results and Applications



(a) Ground truth; (b) and (c) borehole and surface data



(a)-(c) Three conditional realizations



(a) Dike material in true model and conditional data.

(b) 12 different dike realizations conditioned on surface and borehole data.

(c) Dike 1, front view

(d) Dike 2, front view

THE UNIVERSITY OF BRITISH COLUMBIA

VECTOR INSTITUTE | INSTITUT VECTEUR

NSERC CRSNG

KAUST

Hammer Lab