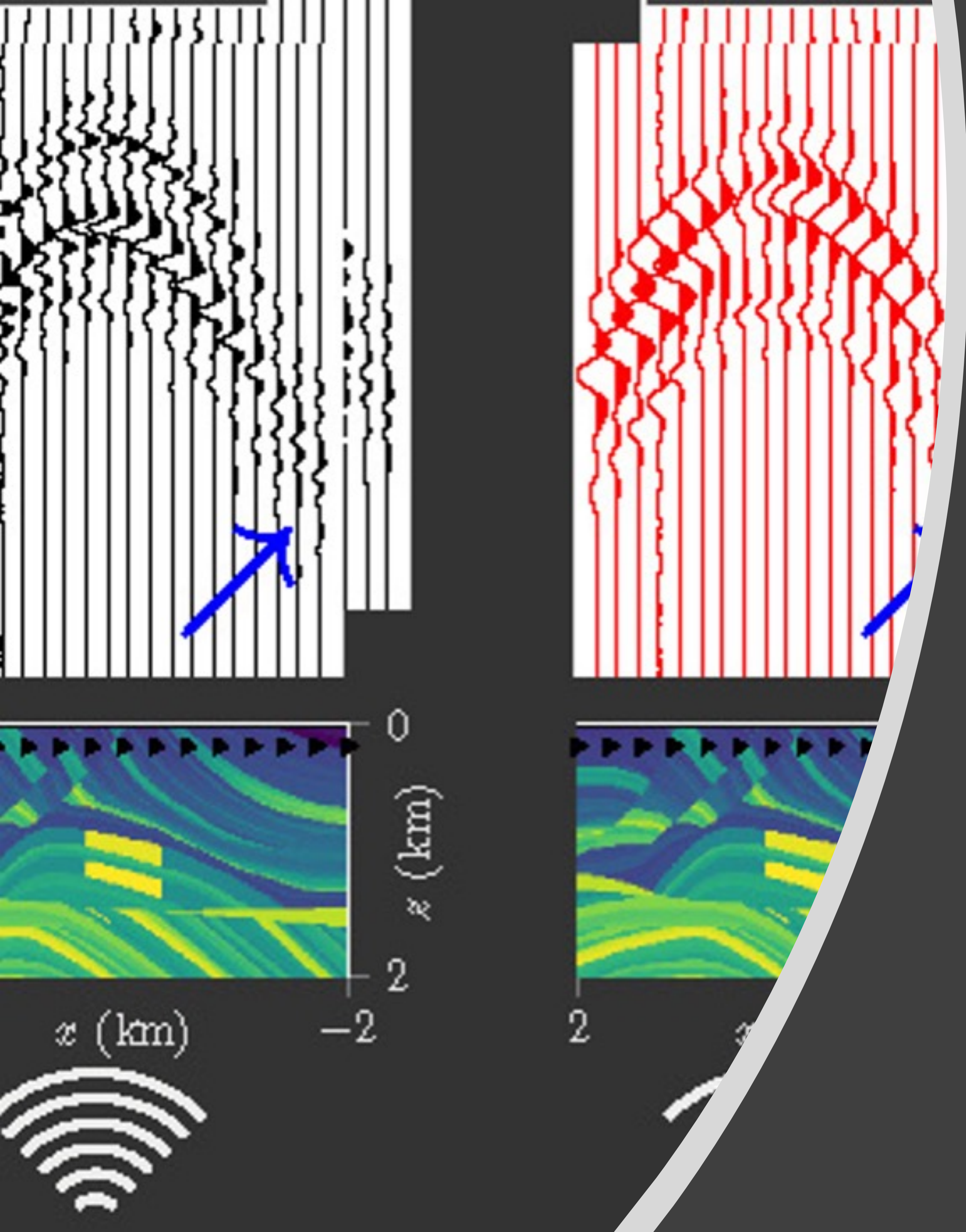


Recovery phenomena with symmetric autoencoders

Laurent Demanet
Mathematics, MIT
With Pawan Bharadwaj (IIS), Matt Li (MIT), Borjan Geshkovski (MIT)

ACM reunion conference
Caltech
November 11, 2023

Data

Black box predictor

Simulations

Black box model replicator

How deep nets help CSE

- Model reduction
- Solve PDE with NN
- Regress the PDE
- Physics-informed NN
- Subgrid closure models, etc.

Data

Black box predictor

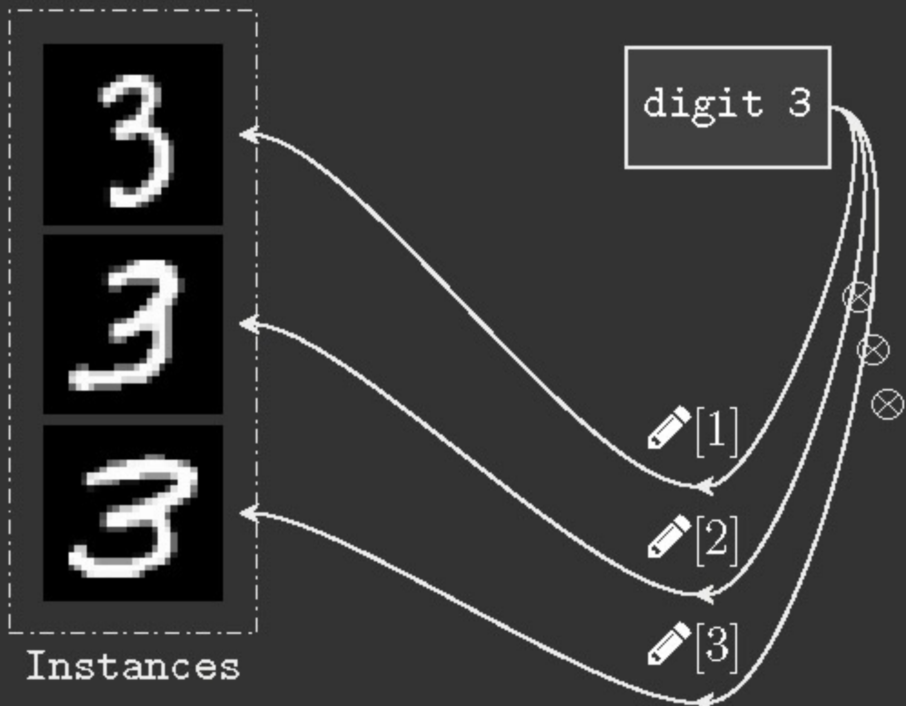
Simulations

Black box model replicator

What can a deep net really know?

- Disentangle explanatory components
- Generate unseen data (redatum)

>>> WHAT: Estimate *Similarity* or *Coherency* Among Instances



→ Coherent (G)

3; digit information is similar in all images

→ Unknown Operation $\otimes$

→ Dissimilar (Nuisance; W)

✎; writing style and orientation are dissimilar among instances

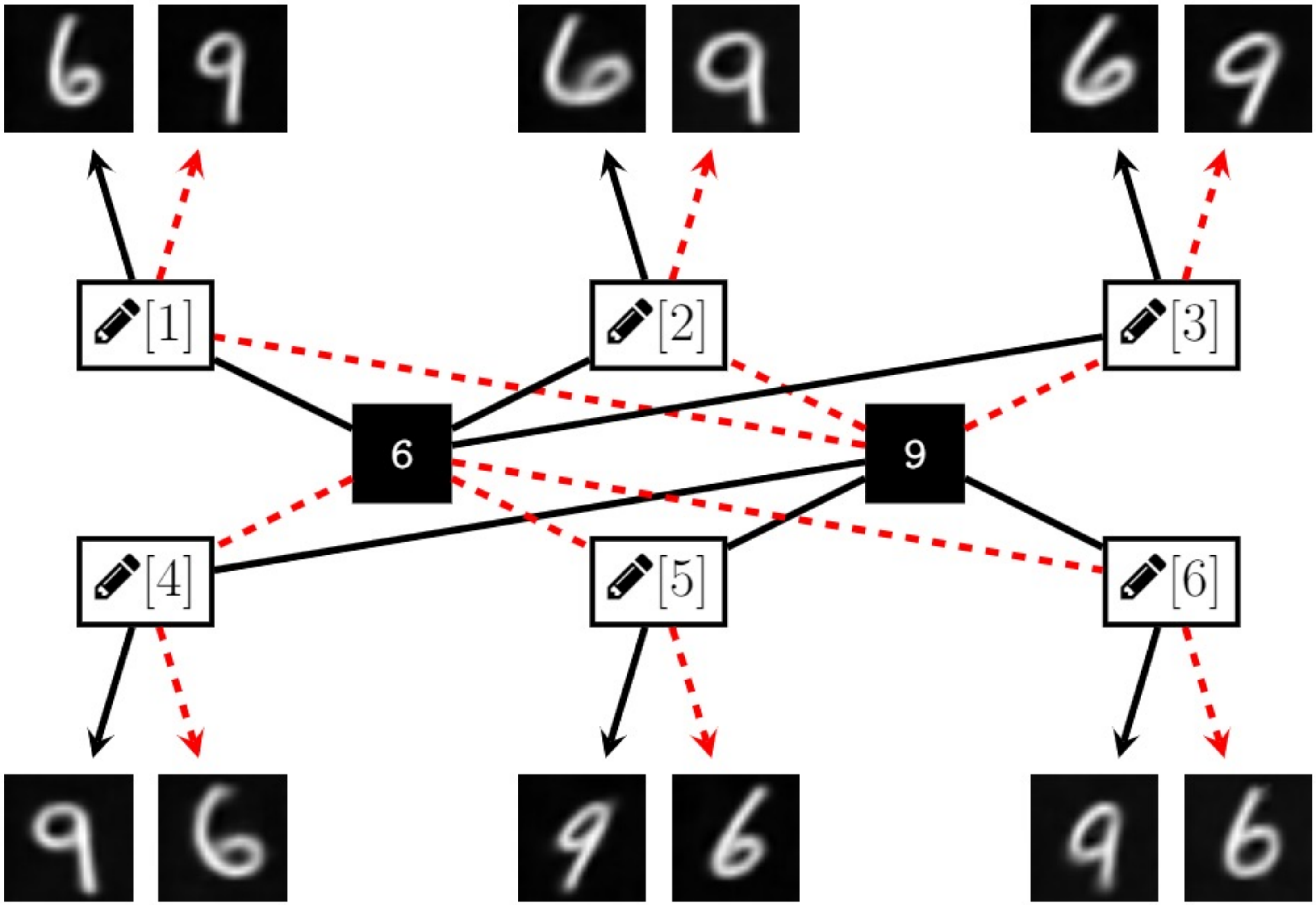
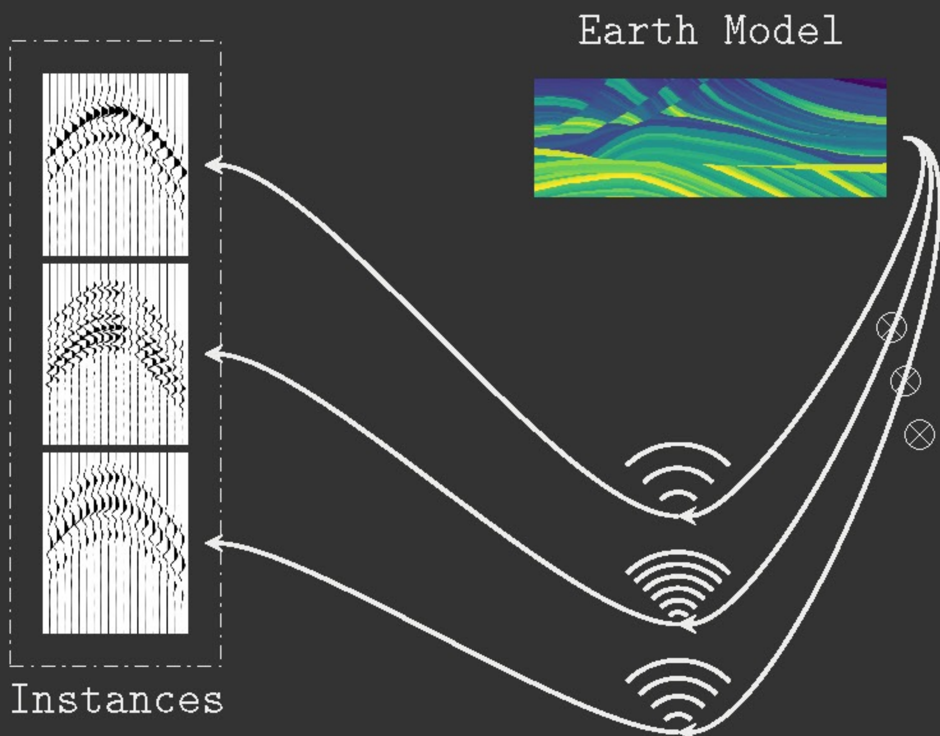


Fig. 1: In the MNIST experiment, the SIMO system responds to the input digit information (here, 6 and 9) by producing multiple dissimilar hand-written images as channel outputs. Here, six channel outputs are plotted. The first three channels didn't respond to 9, however SymAE produces their virtual outputs (dashed lines) that have identical writing style as in the true outputs. Similarly, virtual channel outputs of the last three channels are also plotted.

>>> WHAT: Estimate *Similarity* or *Coherency* Among Instances



→ Coherent (G)

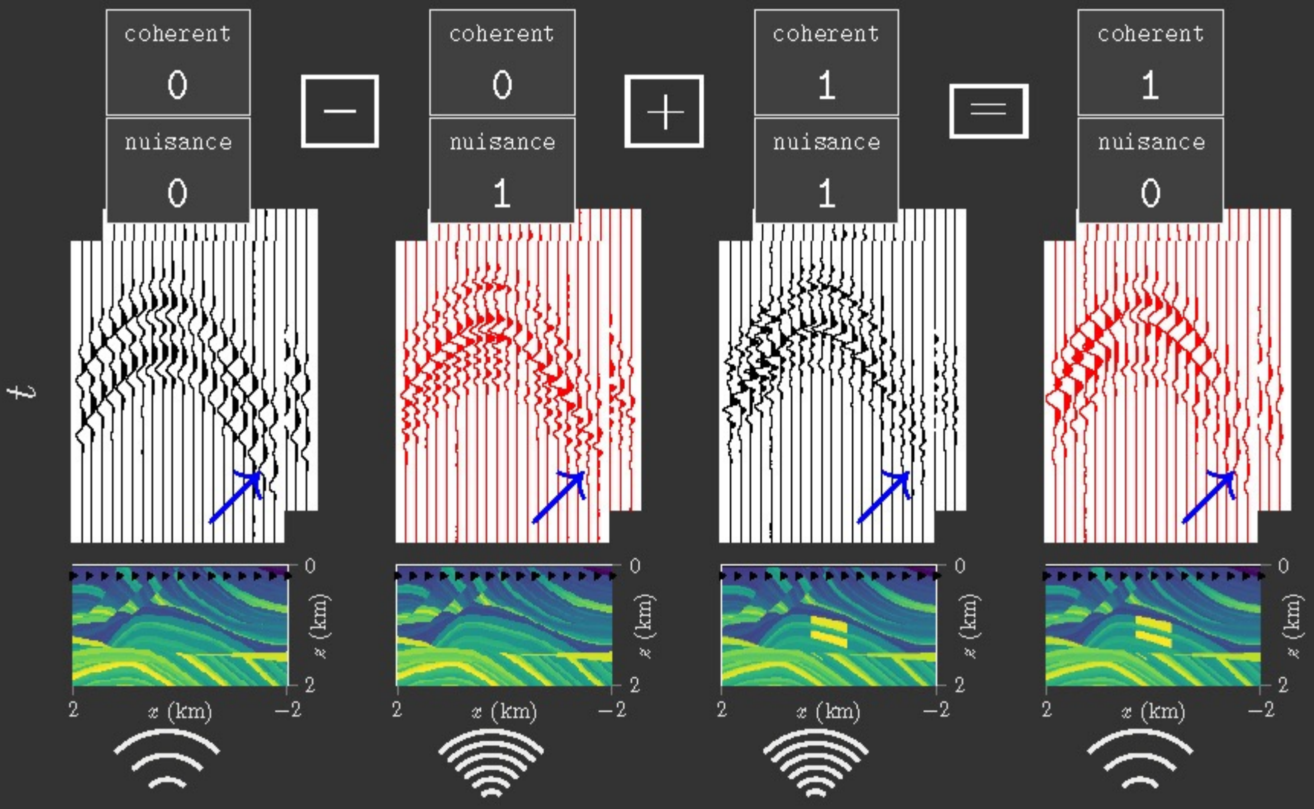
medium effects are similar in all passive shot gathers

→ Unknown Operation $\otimes$

→ Dissimilar (Nuisance; W)

source mechanism, signature and position are dissimilar among instances

>>> WHY: Redatum To Produce Virtual Gathers!



Solution: SymAE produces virtual baseline sources during monitoring

Why does this work?

There is an underlying model of the form

$$X_{ij} = \mathcal{F}(a_i, b_j)$$

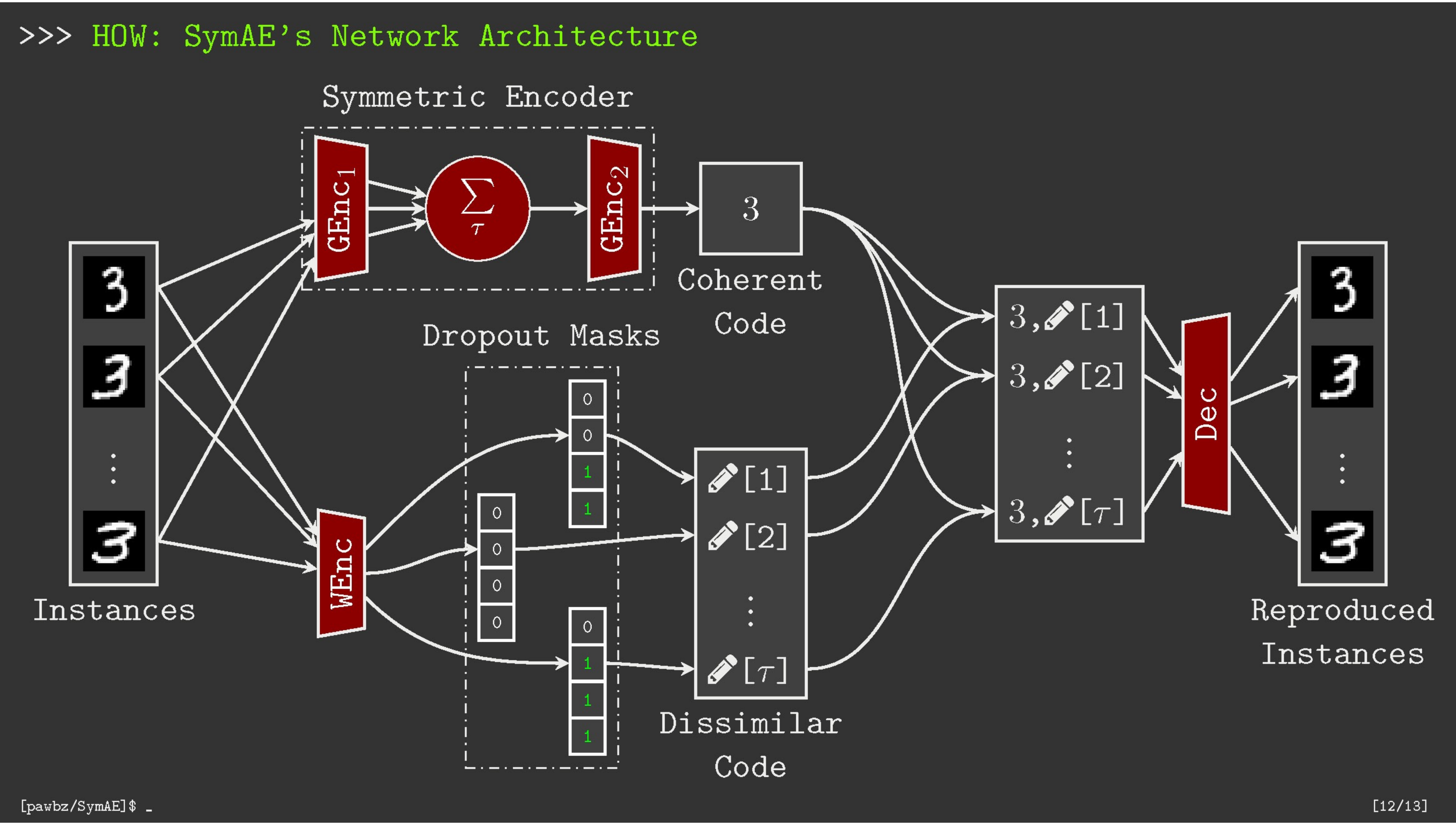
Symmetry under permutations
“Latent rank-1”

The network’s explanation is

$$X_{ij} = \mathcal{G}(z_i^{(a)}, z_j^{(b)})$$

-> Matrix recovery question

How does this work?



Regularization with $\mathcal{G}$

Inverse problem

$$y = \mathcal{A}(X)$$

Require $x \in \text{Ran}(\mathcal{G})$ for some deep generator $\mathcal{G}$. Then

$$\widehat{X} = \widehat{\mathcal{G}}(\widehat{z})$$

where

$$(\widehat{z}, \widehat{\mathcal{G}}) = \text{argmin}_{z,\mathcal{G}} \|y - \mathcal{A}(\mathcal{G}(z))\|_2$$

Measure of complexity:

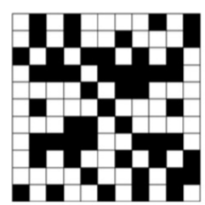
1. $z = (z^{(a)}, z^{(b)}) \in \mathbb{R}^{O(n_1+n_2)}$; and
2. $\mathcal{G}$ dense ReLu width w and depth d.

Regularization and recovery in Ran($\mathcal{G}$)

Matrix to complete/recover:

$$X_0 := \begin{bmatrix} I(1,1) & \dots & I(1,n_2) \\ \vdots & \ddots & \vdots \\ I(n_1,1) & \dots & I(n_1,n_2) \end{bmatrix} \in \mathbb{R}^{n_1 d_x \times n_2 d_y}$$

* Measurements: $I(i,j)$ for $(i,j) \in \Omega$, so $y = \mathcal{A}(X_0)$, where $\mathcal{A}$ is binary mask onto m blocks of size $d_x \times d_y$ indexed by Ω



* Idealized measurements: $y = \mathcal{A}(X_0)$, where $\mathcal{A}$ is Gaussian iid with m rows.

Regularization with $\mathcal{G}$

Recovery: Among all $X \in \mathrm{Ran}(\mathcal{G})$ for some $\mathcal{G}$, with zero training error ($\mathcal{A}(X) = \mathcal{A}(X_0)$), do we have $X = X_0$?

Complexity of $\mathcal{G}$: let $\Sigma =$ product of all squared spectral norms of weights of decoder

Theorem [D., Geshkovski '23]

If

$$m \gtrsim (n_1 + n_2)\Sigma^2 \quad \text{and logarithmic terms}$$

and $\mathcal{A}(X_0) = \mathcal{A}(X)$, then we have perfect recovery w.h.p.: $X = X_0$.

See also Bora et al, 2017 (fixed $\mathcal{G}$)

Proof ingredients

- Niceness of neural nets: $X \in \mathcal{G}$ and $X_0 \in \mathcal{G} \implies X - X_0 \in \mathcal{G}'$ (larger)
- Lower bound of "smallest singular value on a subset":

$$\inf_{Z \in \mathsf{Ran}(\mathcal{G}) \cap \mathbb{S}^{n-1}} \|\mathcal{A}(Z)\| > 0$$

$\implies$ Gordon's escape through the mesh lemma

- True, as long as $m \geqslant$ "complexity" of $\mathcal{G} \cap \mathbb{S}^{n-1}$
- "Complexity" $=$ average diameter of $\mathcal{G} \cap \mathbb{S}^{n-1}$ (Gaussian width). Can be computed by finding the covering number of $\mathsf{Ran}(\mathcal{G}) \cap \mathbb{S}^{n-1}$

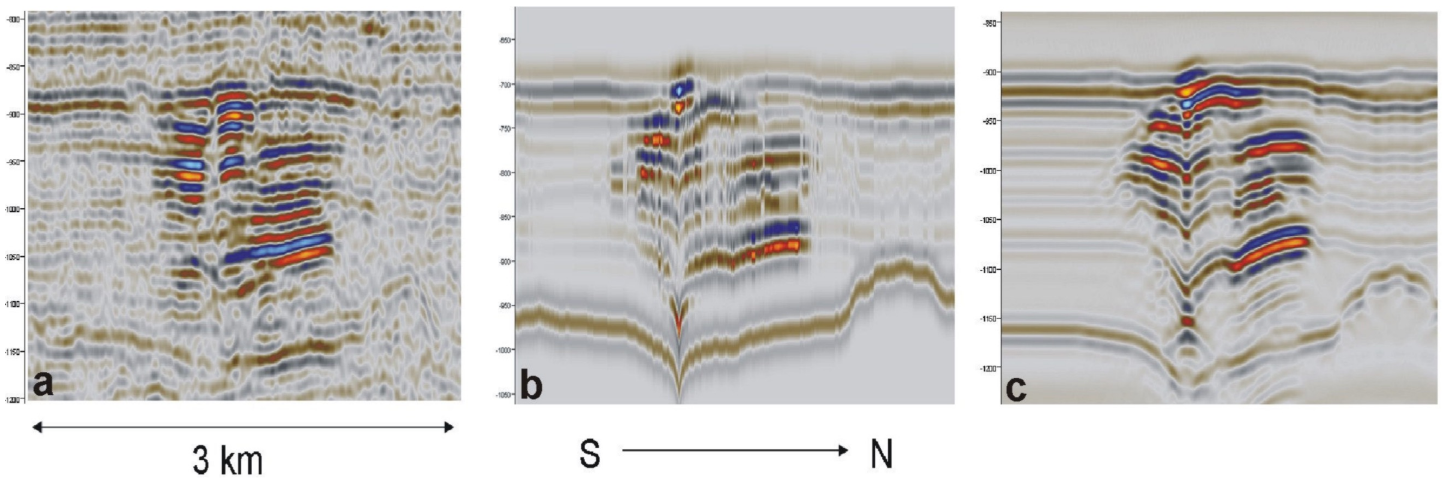
Why deep nets… and why not

Radically new inference tools

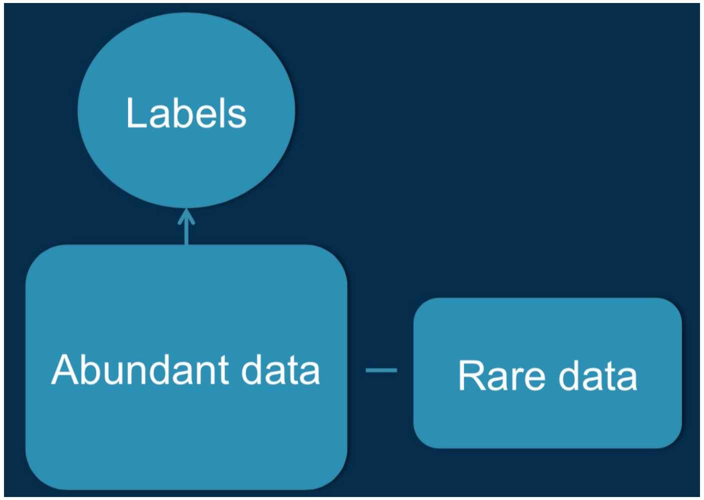
``Intelligence" = low-dimensional latency + low-complexity decoder

What are we giving up?

- interpretability
- guarantees (generalizablility)
- Science!
- … still not automated



Outlook



Example of data misfit: Arts et al, 2007 (Sleipner CO2 injection field)

Co-train from simulations and data

Explaining beyond modeling

Deep generative compressed sensing

Linear inverse problem

$$y = Ax_0 + \eta$$

with $A \in \mathbb{R}^{m \times n}$, Gaussian iid with $m < n$.
Classical regularization:

$$\min_x \|y - Ax\|_2^2 + \lambda \|x\|_*$$

New: require $x \in \text{Ran}(\mathcal{G})$ for some deep generator $\mathcal{G}$. Then

$$\widehat{x} = \mathcal{G}(\widehat{z})$$

where

$$\widehat{z} = \text{argmin}\{\|y - A\mathcal{G}(z)\|_2 : z \in \mathbb{R}^k, \|z\|_2 \leq r\}$$

Complexity of $\mathcal{G}$: $z \in \mathbb{R}^k$ and $\mathcal{G}$ is L-Lipschitz.

Deep generative compressed sensing

Theorem [Bora, Jalal, Price, Dimakis, 2017]

For every $\delta \in (0, 1)$, if

$$m \gtrsim k \log \frac{Lr}{\delta}$$

then it holds with probability $1 - e^{-O(m)}$ that

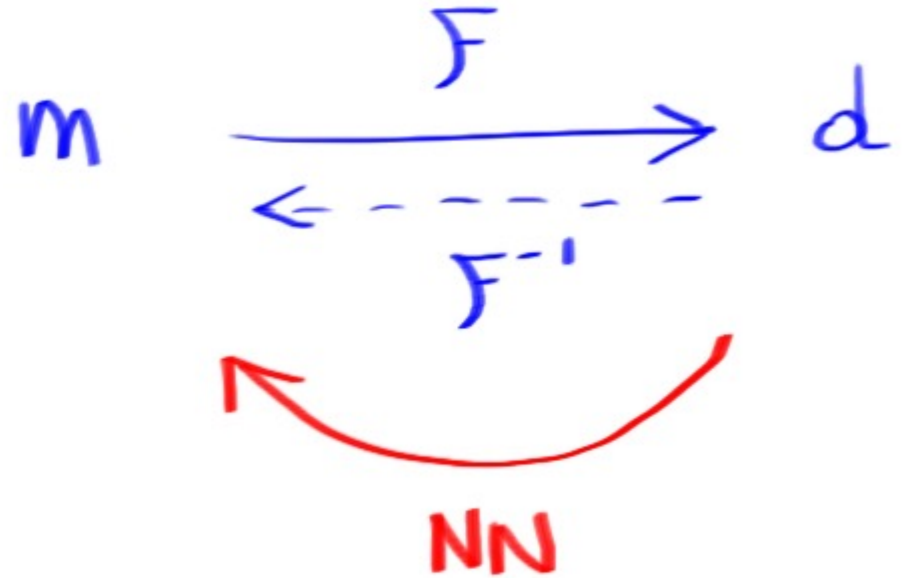
$$\|\widehat{x} - x_0\|_2 \leq 6 \min_{\|z\|_2 \leq r} \|G(z) - x_0\|_2 + 3\|\eta\|_2 + 2\delta$$

If $\mathcal{G}$ is a deep generator with width w and depth d (dense ReLU), then

$$L \lesssim w^{O(d)}$$

and r is a non-issue.
A version of the theorem deals with $\|y - AG(\widehat{z})\|$ within ε of optimum.

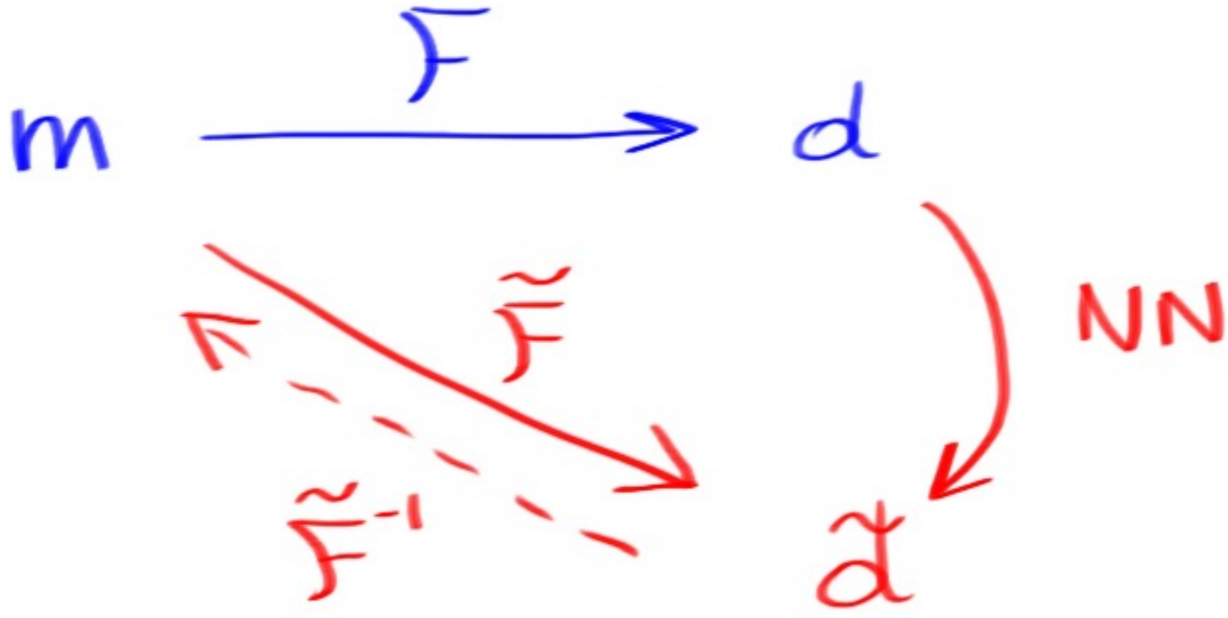
Neural networks for inverse problems



OK for small scale problems

Invert $\mathcal{F}$ (train on simulations) -- or bypass $\mathcal{F}$ (train on real data)

Neural networks for inverse problems

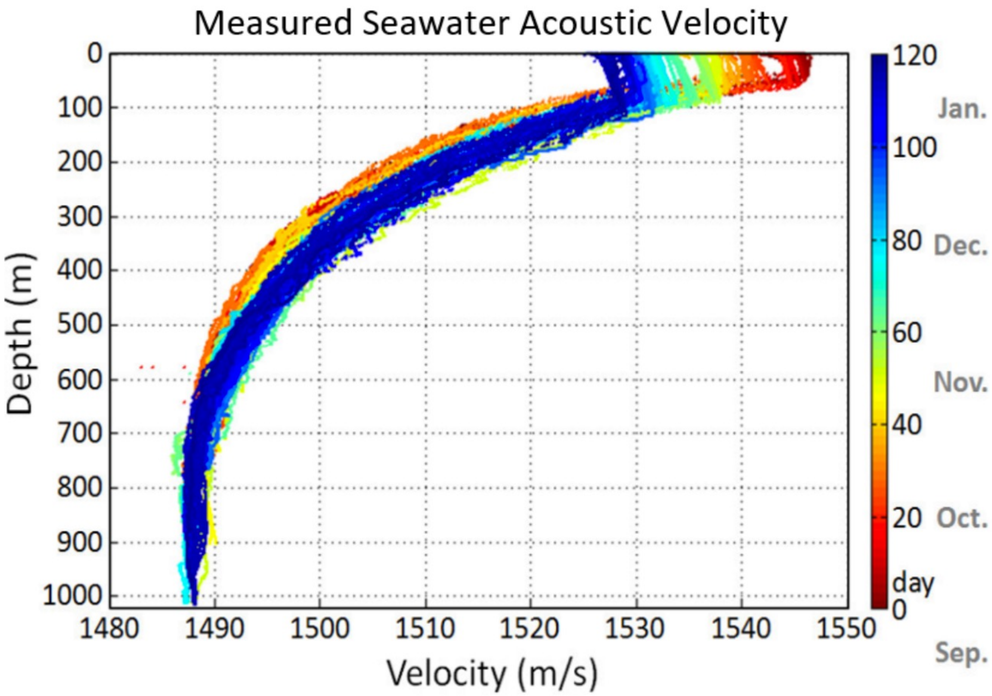


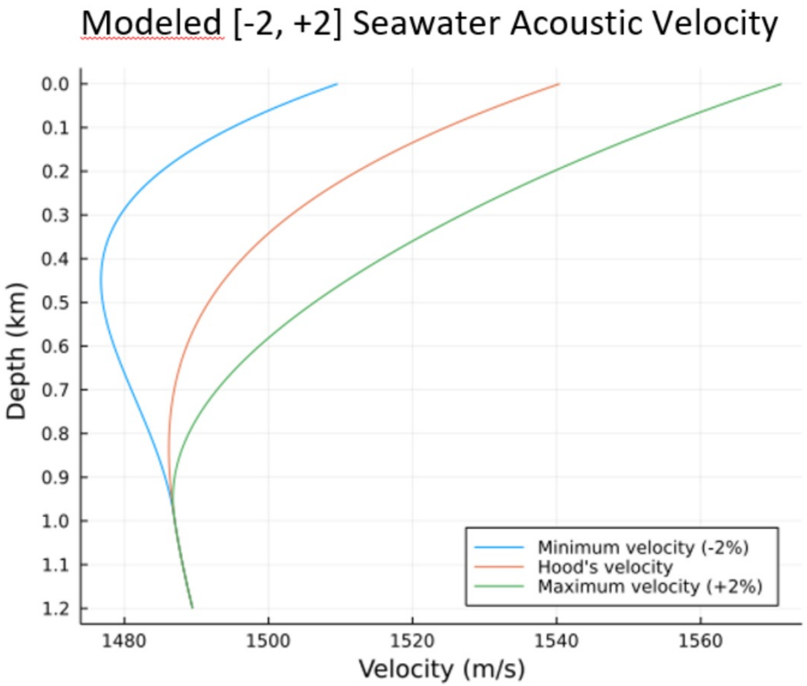
Auxiliary data extension task

More robust or favorable inversion

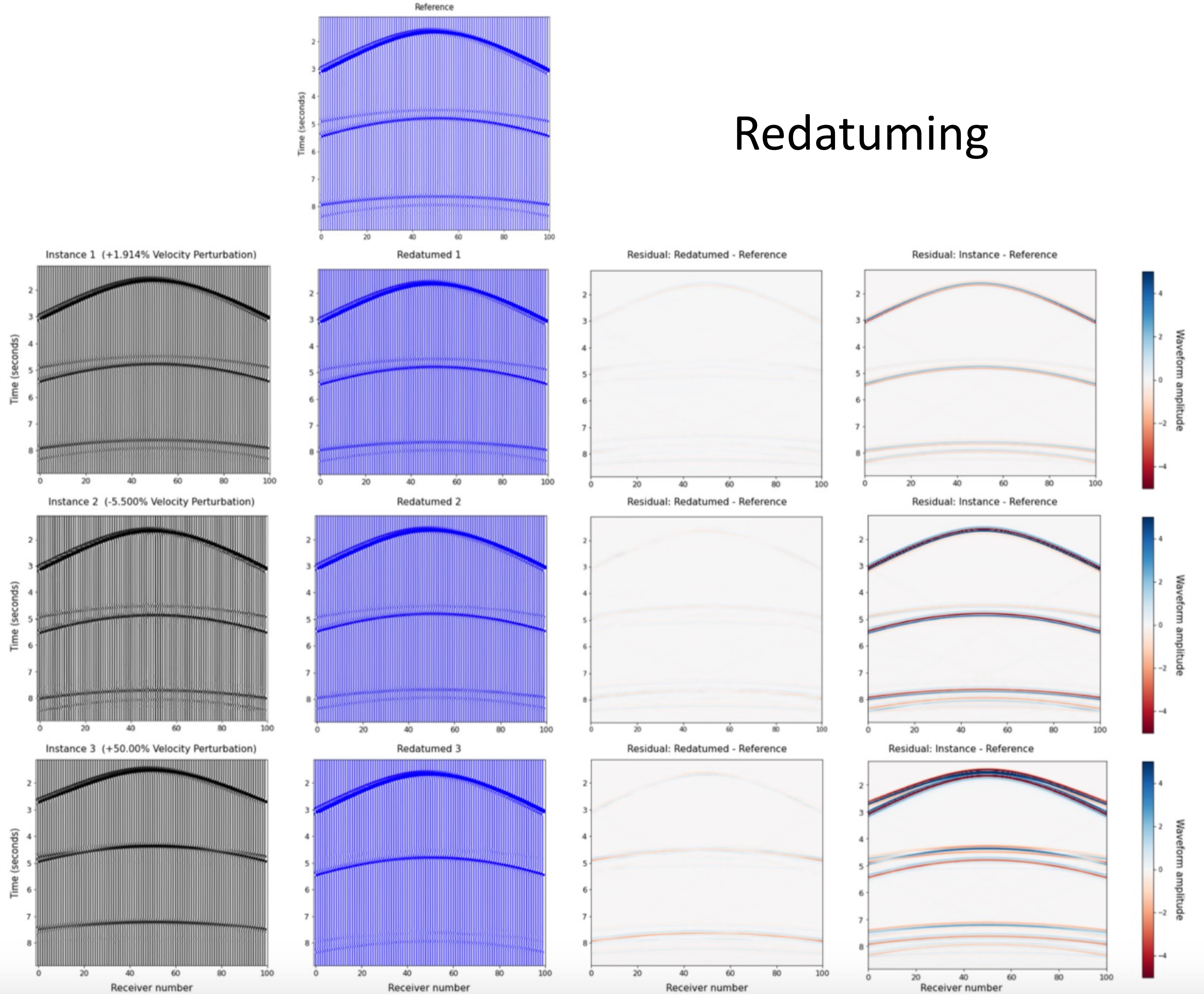
Bridge real vs synthetic divide

Example: Deepwater statics

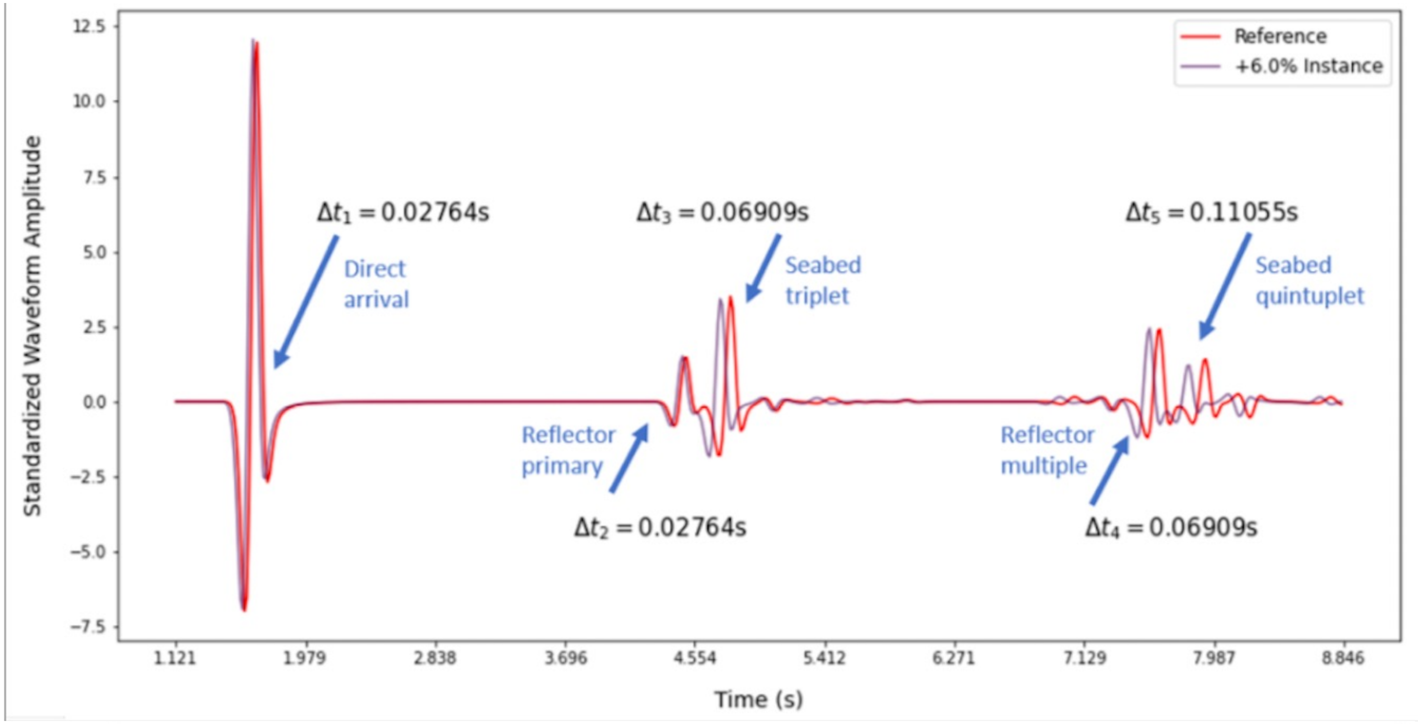




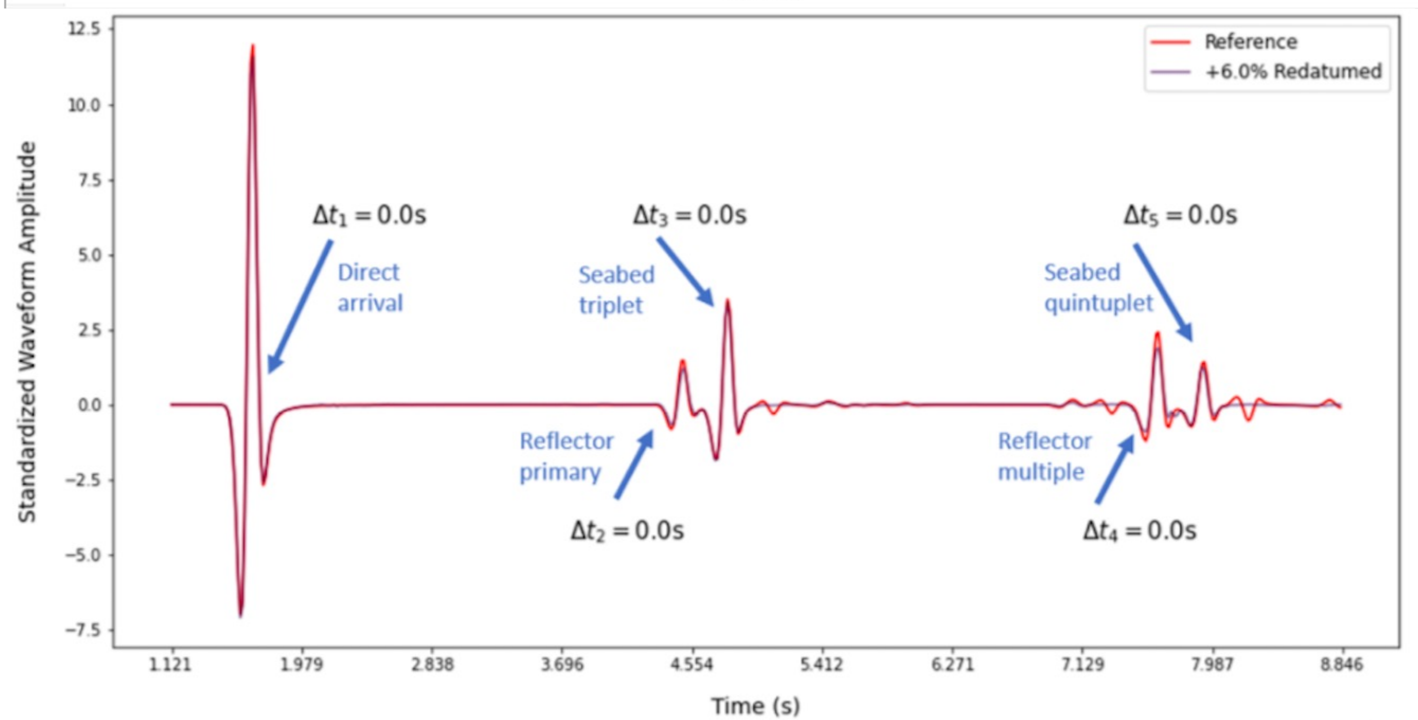
Redatuming



Reference
Instance 1 (+1.914% Velocity Perturbation)
Redatumed 1
Residual: Redatumed - Reference
Residual: Instance - Reference
Instance 2 (-5.500% Velocity Perturbation)
Redatumed 2
Residual: Redatumed - Reference
Residual: Instance - Reference
Instance 3 (+50.00% Velocity Perturbation)
Redatumed 3
Residual: Redatumed - Reference
Residual: Instance - Reference
Time (seconds)
Receiver number
Waveform amplitude



Reference
+6.0% Instance
Standardized Waveform Amplitude
Time (s)
Δt1 = 0.02764s
Direct arrival
Δt3 = 0.06909s
Seabed triplet
Δt5 = 0.11055s
Seabed quintuplet
Reflector primary
Δt2 = 0.02764s
Reflector multiple
Δt4 = 0.06909s



Reference
+6.0% Redatumed
Standardized Waveform Amplitude
Time (s)
Δt1 = 0.0s
Direct arrival
Δt3 = 0.0s
Seabed triplet
Δt5 = 0.0s
Seabed quintuplet
Reflector primary
Δt2 = 0.0s
Reflector multiple
Δt4 = 0.0s