Normalizing Flows for Uncertainty Quantification in Seismic Imaging

Alison Malcolm – Memorial University of Newfoundland

Rajiv Kumar – Schlumberger

Changxiao Sun – Innovation Academy for Precision Measurement Science and Technology, CAS, Visiting Student, Memorial University of Newfoundland

Maria Kotsi – Pangeo, formerly Memorial University of Newfoundland

MEMORIAL UNIVERSITY

NSERC CRSNG

Schlumberger

APM

China Scholarship Council

Acknowledgements

MEMORIAL UNIVERSITY
amalcolm@mun.ca

- Co-authors: Changxiao Sun, Rajiv Kumar, Maria Kotsi
- Past co-authors: Ben Clarke, Ali Siahkoohi
- SLIM group at Georgia Tech for making their codes available
- Developers of python and Julia
- Chrysoula and Noemi and AWM
- Our sponsors:

Scott MacLachlan

smaclachlan [at] mun.ca

Department of Mathematics and Statistics
Memorial University of Newfoundland
St. John's, NL
A1C 5S7
Canada
709-864-8095
Office: HH2019

CO_2 Sequestration

MEMORIAL UNIVERSITY
amalcolm@mun.ca

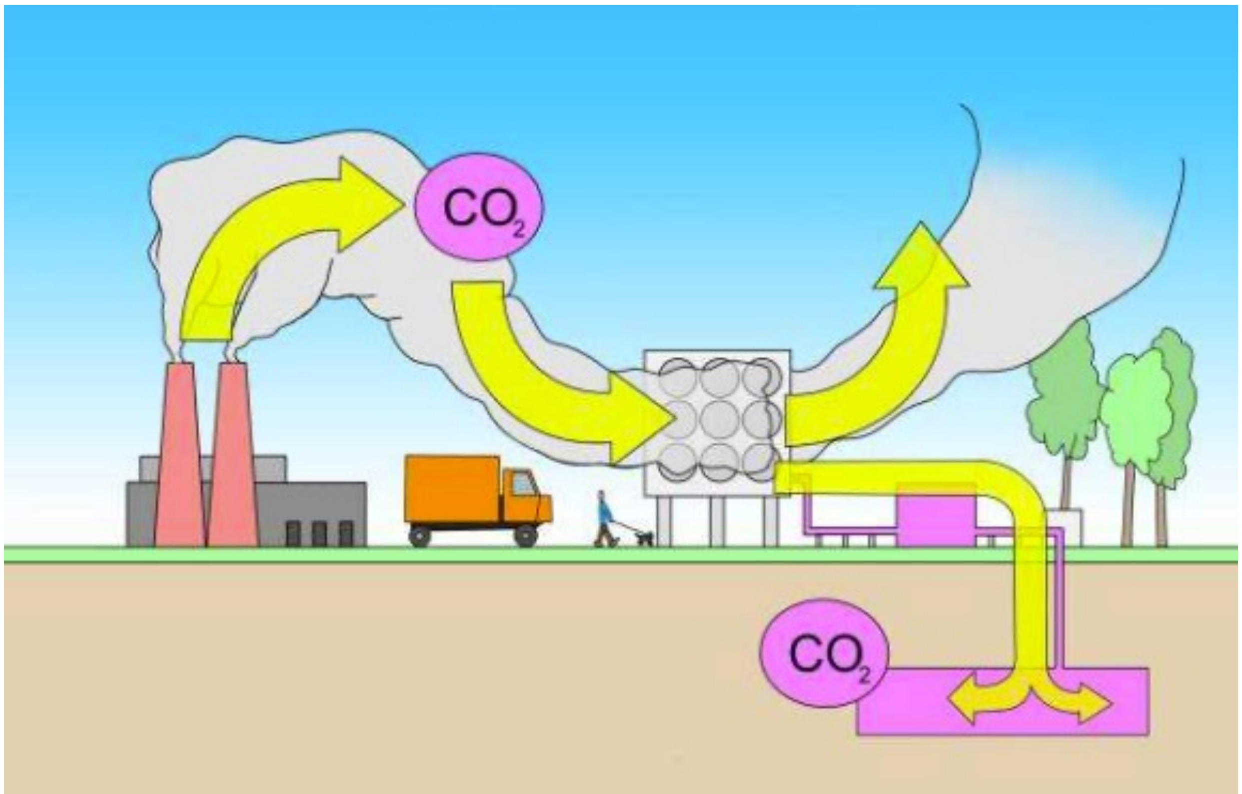
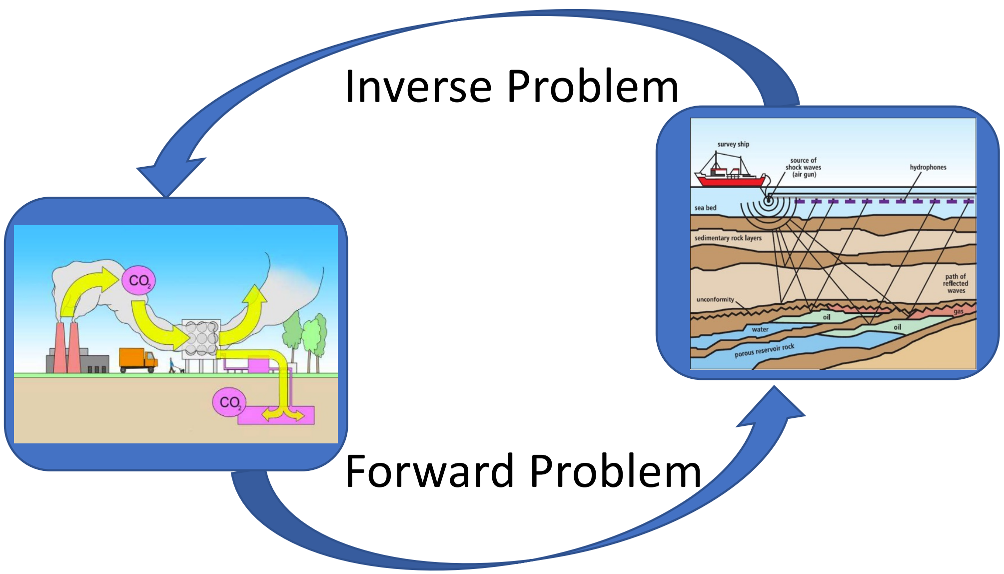


Figure From Don White, NRCAN.

We are talking about geologic storage, there are many other ways to ‘lock’ in Carbon.

Inverse Problems Basics

MEMORIAL UNIVERSITY
amalcolm@mun.ca



Some Unique Aspects:

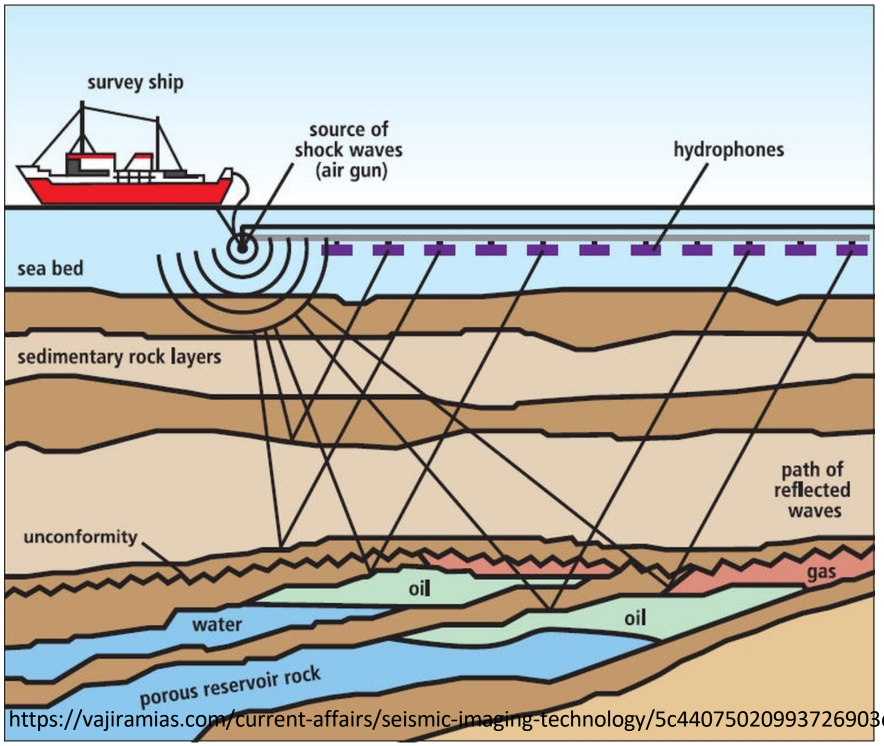
- We know how much CO2 was injected and to where
- Key goal is to understand if/where CO_2 is leaking
- We don't necessarily need a detailed point-by-point subsurface model

Goals of this project:

- Quickly generate an image and characterize its uncertainty

Seismic Imaging

MEMORIAL UNIVERSITY
amalcolm@mun.ca



https://vajiramias.com/current-affairs/seismic-imaging-technology/5c44075020993726903c4741/

Governing Equation:

$$\frac{\partial^2 u}{\partial t^2} + c^2 \nabla^2 u = f$$

Our data are:

$$d_{modelled} = Pu = F[m]$$

Our model is:

$$m = \frac{1}{c^2}$$

Our optimization problem is:

$$\underset{m}{\operatorname{argmin}}\{ || Fm - d_{observed} ||_2^2 \}$$

u – displacement; c – subsurface wavespeed; f – source of energy; P – projection operator; F – modelling operator

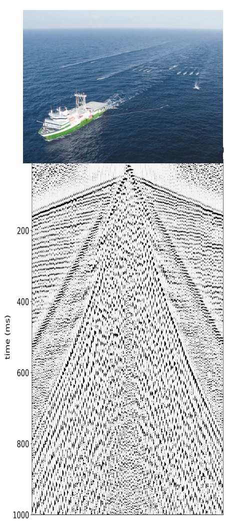
4D seismic imaging

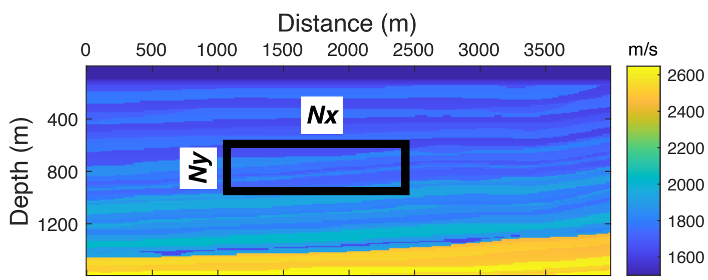
MEMORIAL
UNIVERSITY
amalcolm@mun.ca

4D Seismic:

1. Collect data, and solve your inverse problem
2. Change something (e.g. CO_2)
3. Re-collect the same data, matching everything you can
4. update your model

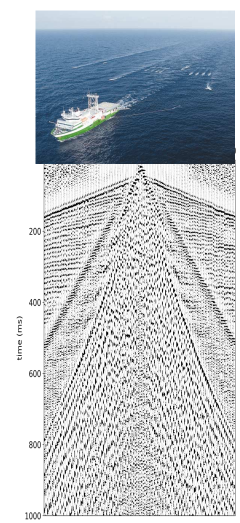
Before:

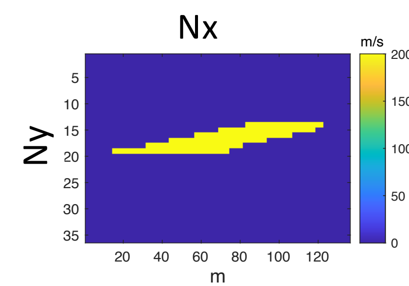




Injection:

After:

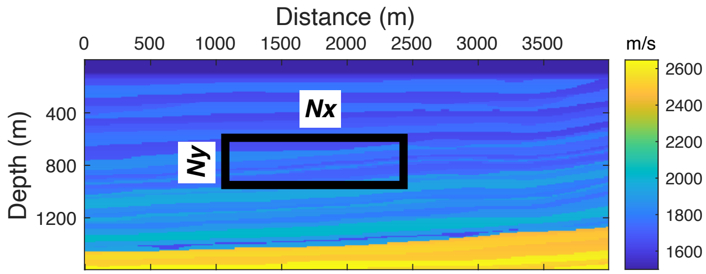


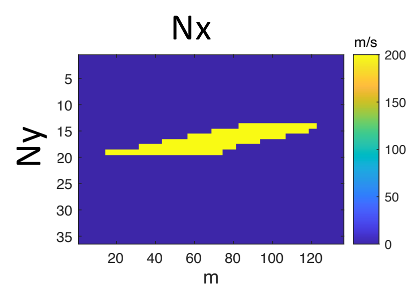


A Quick Full-Waveform Inversion introduction

MEMORIAL UNIVERSITY
amalcolm@mun.ca

- Our optimization problem is: $\underset{m}{\operatorname{argmin}}\{ || Fm - d_{observed} ||_2^2\}$
- Typically solve by:
  - Obtaining an initial model by solving a simplified problem
  - Using l-BFGS or similar starting from low-frequency and building to higher frequencies
  - We use a local solver (Willemsen & AM, 2016), to speed up the calculations and focus the model updates to a small region





Inverse Problems Basics -- Bayes

MEMORIAL UNIVERSITY
amalcolm@mun.ca

Our goal is to recover: $p(m|d)$ the probability of a particular model, given our observed data.

Likelihood: $\mathrm{L(m)} \propto \exp(-\frac{1}{2}(Fm-d)^T\,\Sigma^{-1}(Fm-d))$

Prior

$$p(m|d) = \frac{p(d|m)p(m)}{p(d)}$$

Posterior

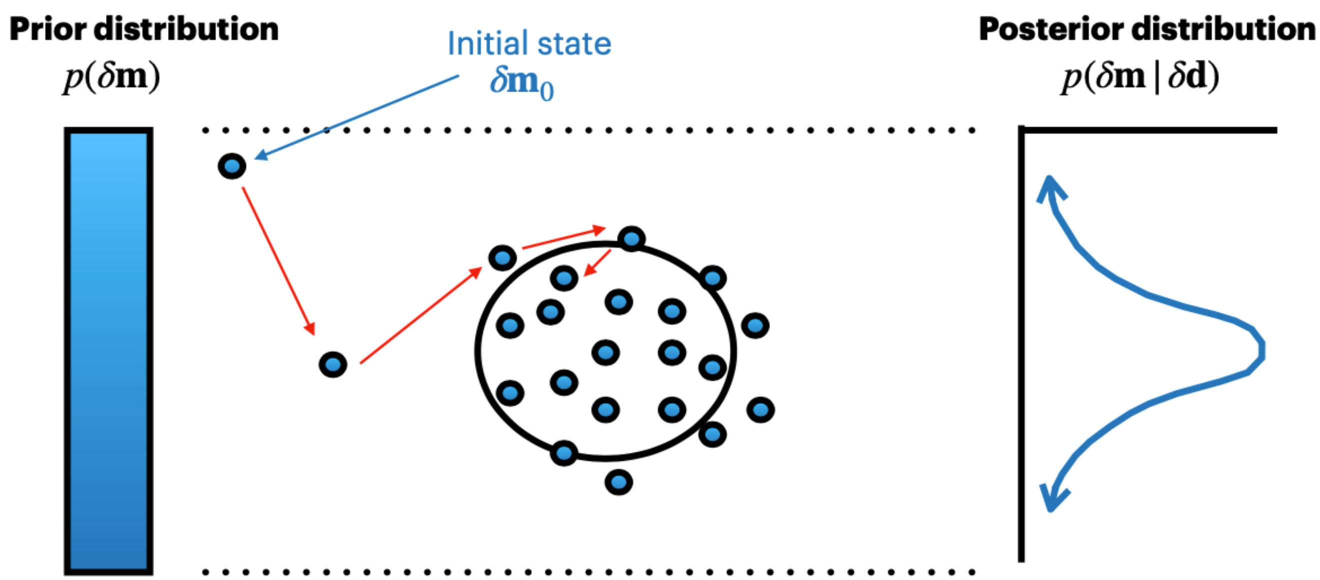
'data prior' normalization

Recall our original problem: $\underset{m}{\operatorname{argmin}}\{\,||\,Fm\ -d\,||_2^2\}$

How we can find $p(m|d)$

MEMORIAL UNIVERSITY
amalcolm@mun.ca

Markov-Chain Monte-Carlo (MCMC)



Each model we try requires an evaluation of our forward model. Not every model we try will be accepted.

Key Issues with MCMC

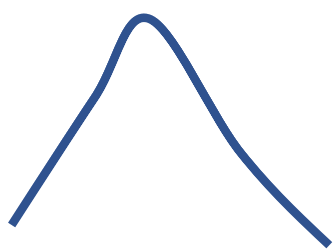
MEMORIAL UNIVERSITY
amalcolm@mun.ca

- Too slow!! We sampled 1-10 MILLION samples
- Not enough degrees of freedom
- Alternatives:
  - Hamiltonian Monte Carlo
    - Fichtner et al, 2018, 2019
    - Kotsi & AM, 2020
  - Stein Variational Gradient Descent (SVGD)
    - Nawaz & Curtis, 2018
    - Zheng & Curtis, 2021
  - **Normalizing Flows**
    - Siahkoohi & Hermann, 2021 – general image processing & codes!
    - Kumar, Kotsi, Siahkoohi & AM, 2021 – image interpolation
    - Zhao et al, 2022 – Full-waveform inversion, comparing to other methods
    - **We will show how to use NF to estimate uncertainties during a normal FWI**

Normalizing Flows?

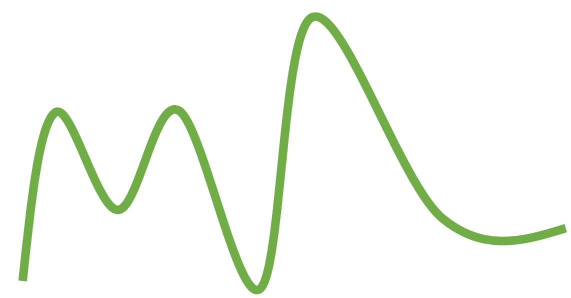
MEMORIAL UNIVERSITY
amalcolm@mun.ca

Simple distribution



Normalizing flows
Give an efficient
mapping in
both directions

Complicated distribution



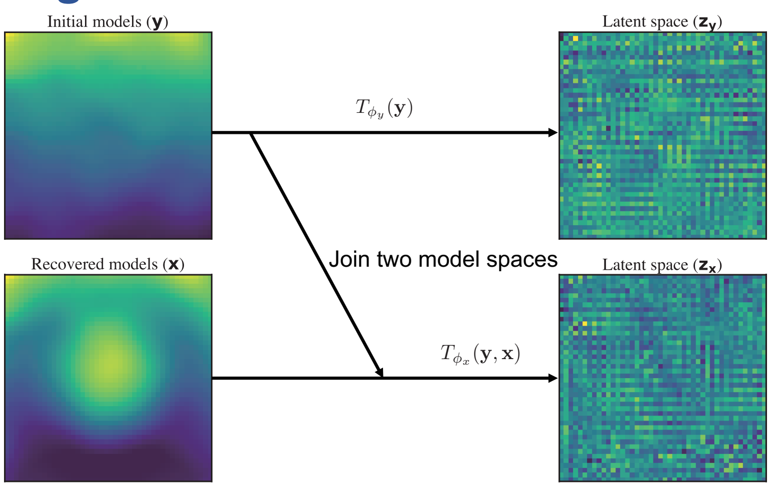
- Easy to sample
- Gives simple statistics
- Doesn't represent most statistics very well
- Allows for easier manipulations

- Difficult to sample
- Gives complicated statistics
- Represent most statistics very well
- Challenging to manipulate

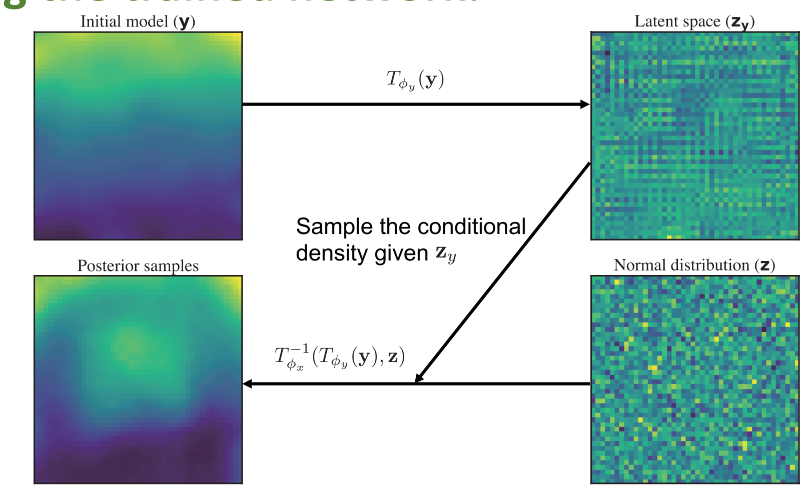
Basics of Normalizing Flows

MEMORIAL UNIVERSITY
amalcolm@mun.ca

Training:



Using the trained network:



Determine 3 mappings:

- Data -> normal
- Model -> normal
- Data -> model normal

Build a mapping:

- Data -> Distribution of model

E.g., we can estimate the mean, standard deviation etc of a new model given an initial model

How do we find this mapping?

MEMORIAL UNIVERSITY
amalcolm@mun.ca

$$\tau = \rho \circ T^{-1} |\det \nabla T^{-1}|$$

- Target distribution → τ
- Initial distribution → ρ
- Mapping → T^{-1}
- Jacobian (Chain rule) → $|\det \nabla T^{-1}|$

First split it into parts: $f_k \circ \cdots \circ f_2 \circ f_1 = T$

Then split the inputs into parts (this makes the Jacobian easy to compute): $f(x) = \begin{pmatrix} x_1 \\ C(x_2|x_1) \end{pmatrix}$

Define a simple function to couple variables: $C(x_2|x_1) = x_2 \odot \exp\big(s(x_1)\big) + t(x_1)$

How do we find this mapping?

MEMORIAL UNIVERSITY
amalcolm@mun.ca

This form: $f(x) = \begin{pmatrix} x_1 \\ C(x_2|x_1) \end{pmatrix}$ where $C(x_2|x_1) = x_2 \odot \exp\big(s(x_1)\big) + t(x_1)$

Neural Networks

lets only some variables connect, **makes our Jacobian easy to compute,**

$$\tau = \rho \circ T^{-1} |\det \nabla T^{-1}|$$

Target distribution

Initial distribution

Mapping

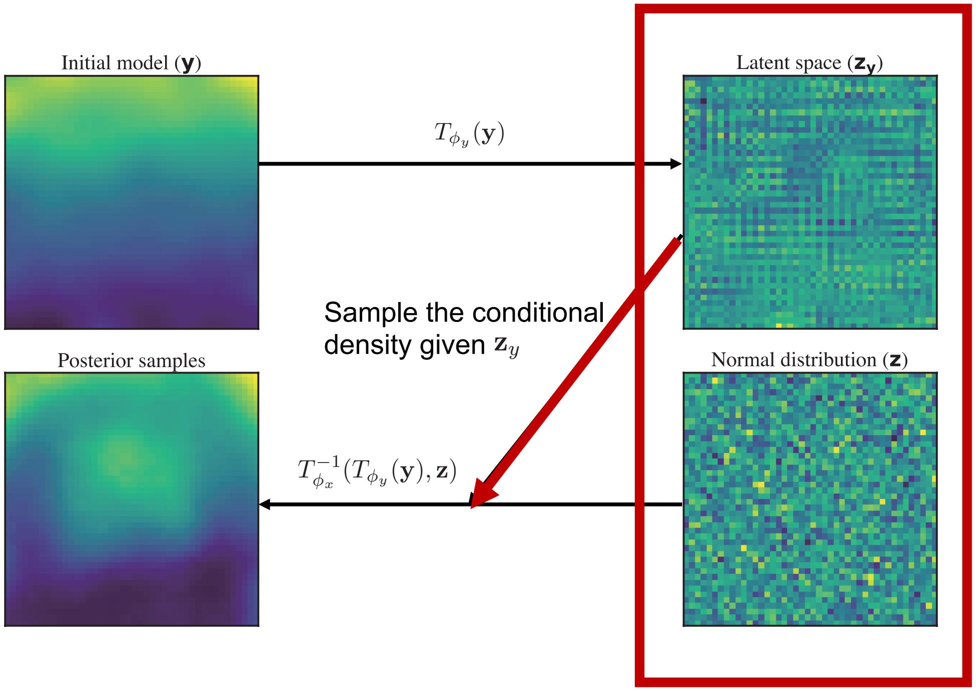
Jacobian (Chain rule)

Really only lets us compute p(m,d) not p(m|d)

From: Kruse et al, 2019

Calculating p(m|d)

MEMORIAL UNIVERSITY
amalcolm@mun.ca

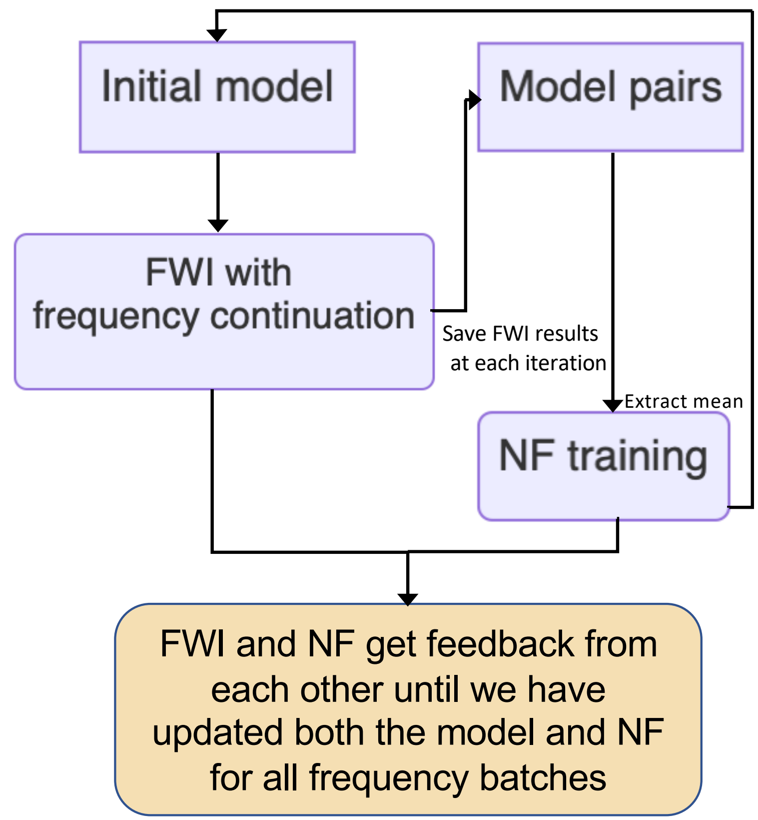


By connecting both initial and final models together, we can then fix our ‘data’ components and sample across the model

How it works for our problem

MEMORIAL UNIVERSITY
amalcolm@mun.ca

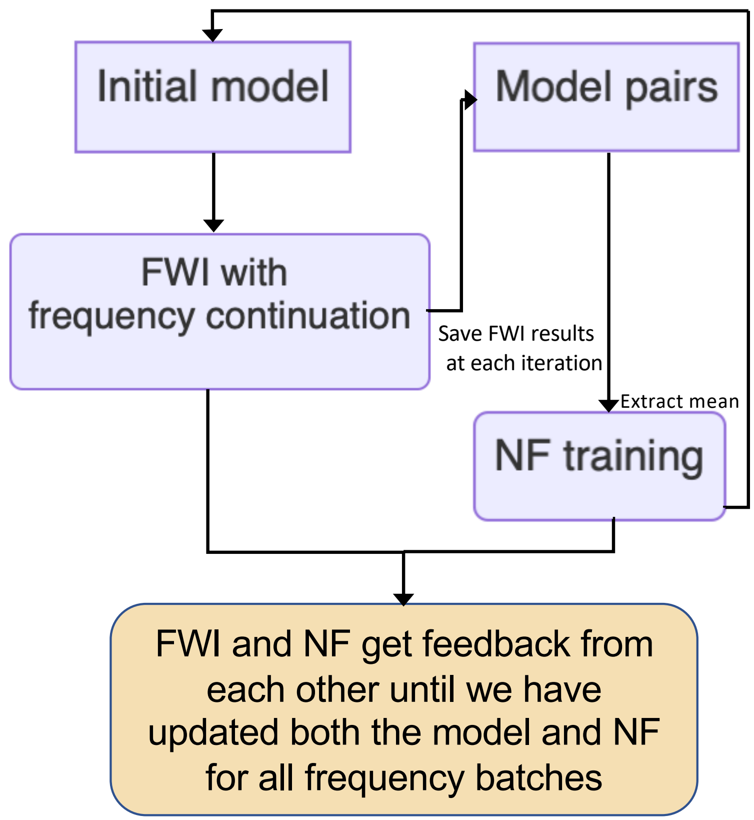
- For an initial model, find the associated point in the latent space (normal distribution)
- Fix the data components (updated model) and sample over the model
- Map our new samples back to the true model space, giving a range of updated models that fit our data



Connecting this to full-waveform inversion

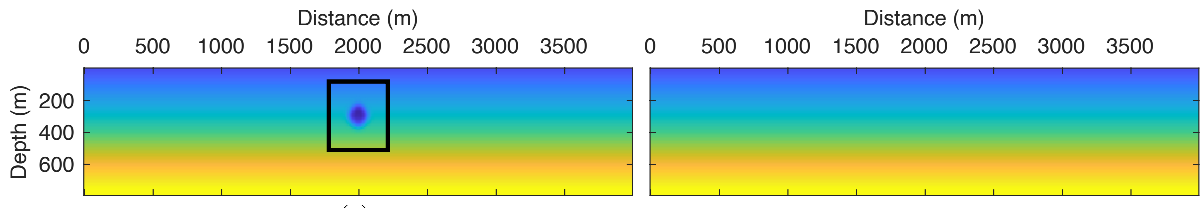
MEMORIAL UNIVERSITY
amalcolm@mun.ca

- We perform the UQ and FWI processes separately:
  - Perform FWI for one frequency batch, saving training pairs consisting of the initial and updated models.
  - Train a NF with these paired models and extract the mean model to be the new starting point in FWI for the next frequency batch.
  - For the next frequency batch, retrain the NF with the previously trained network.



How we setup the training

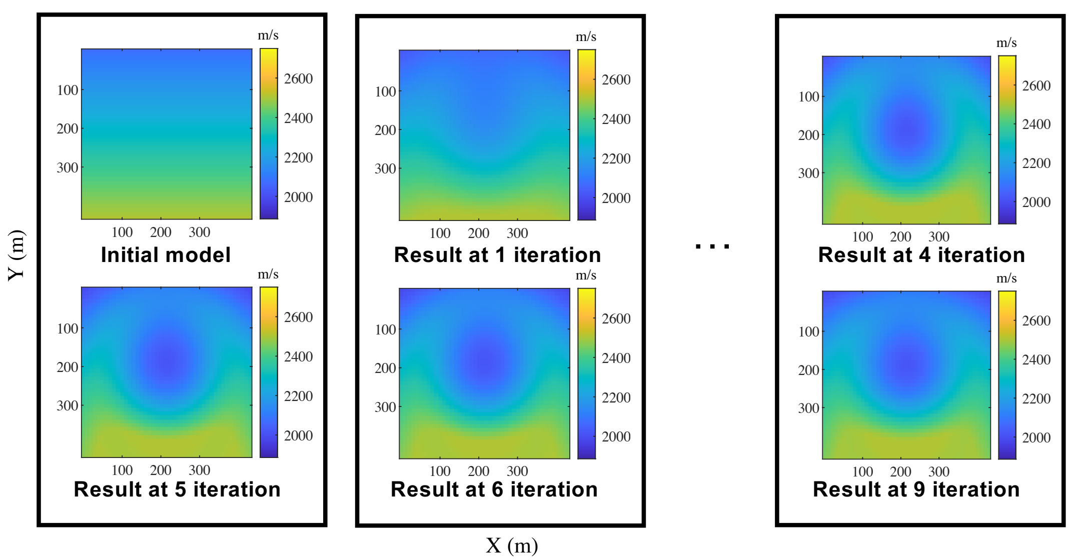
amalcolm@mun.ca



True velocity model

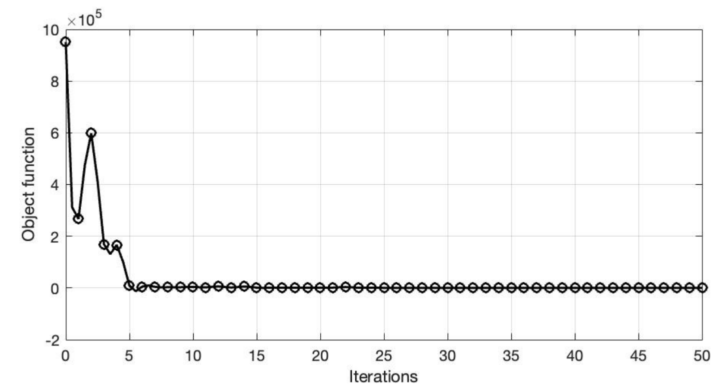
Background velocity model

Training dataset



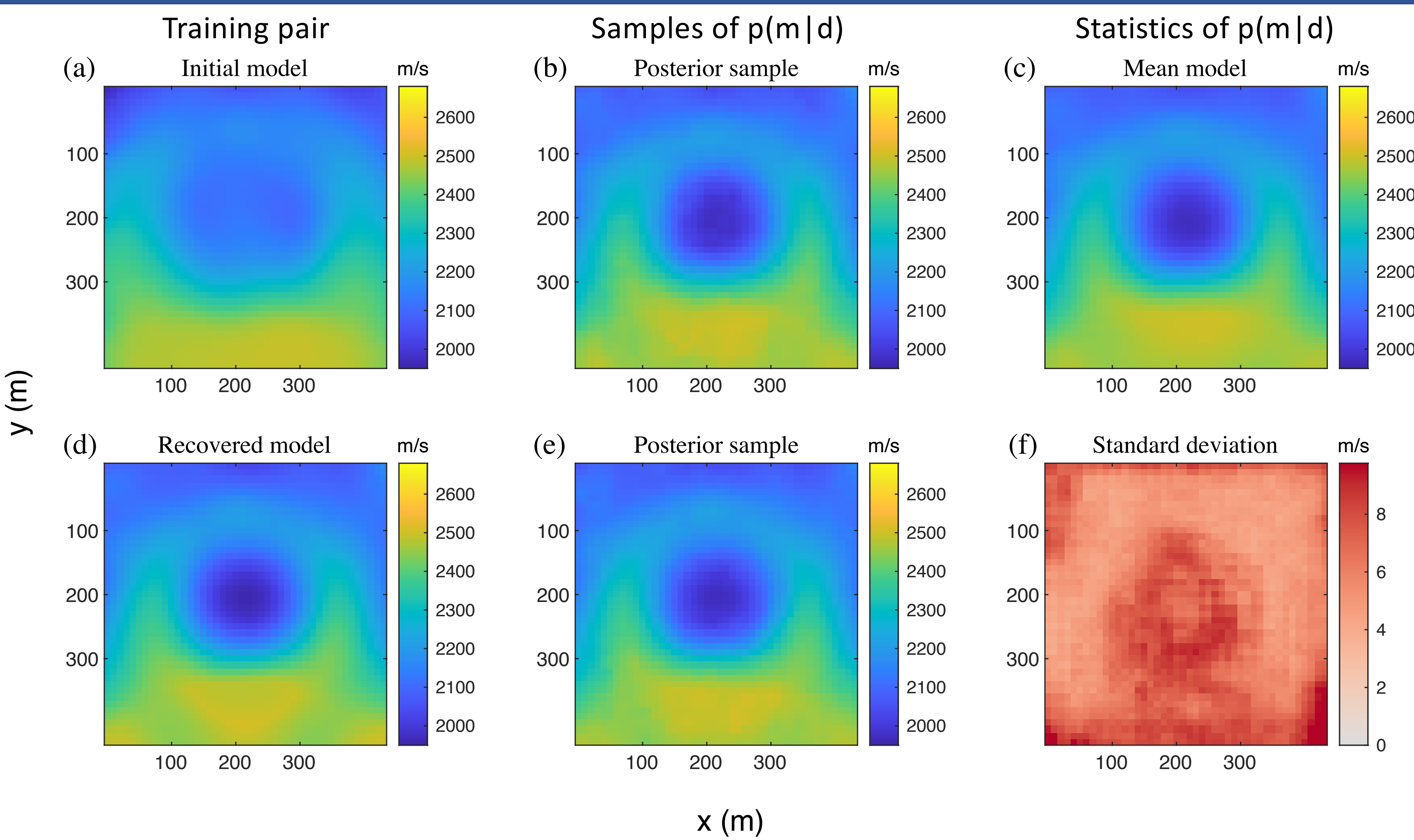
- FWI with local solver at 4Hz for 50 iterations.
- Save FWI results for 10 iterations as training data.

The convergence curve of the object function



Results – Simple Example

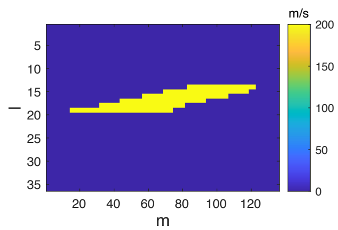
MEMORIAL UNIVERSITY
amalcolm@mun.ca

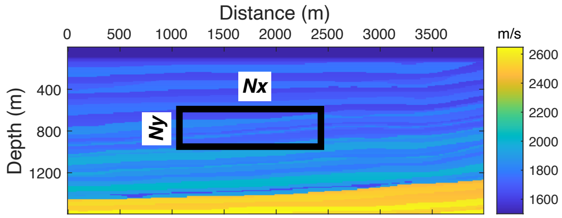


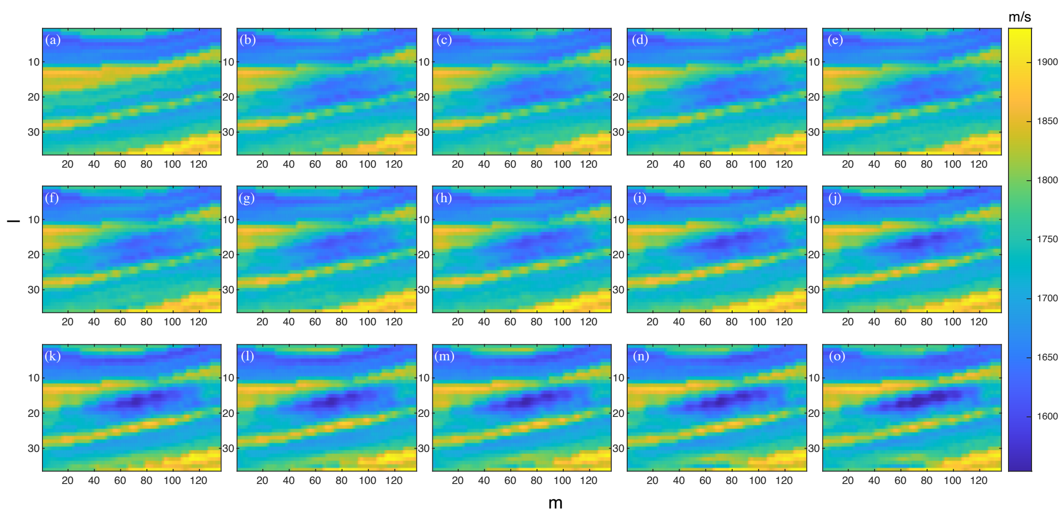
A Time-Lapse Example (for CO_2 Monitoring)

MEMORIAL UNIVERSITY
amalcolm@mun.ca

- Perform FWI with local solver at the first frequency batch for 10 iterations.
- Save FWI results to compose training data.



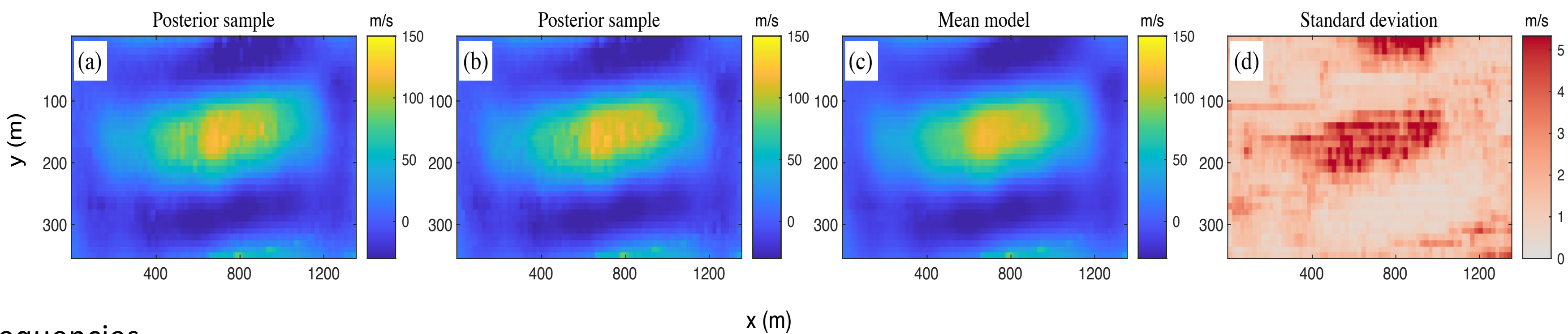




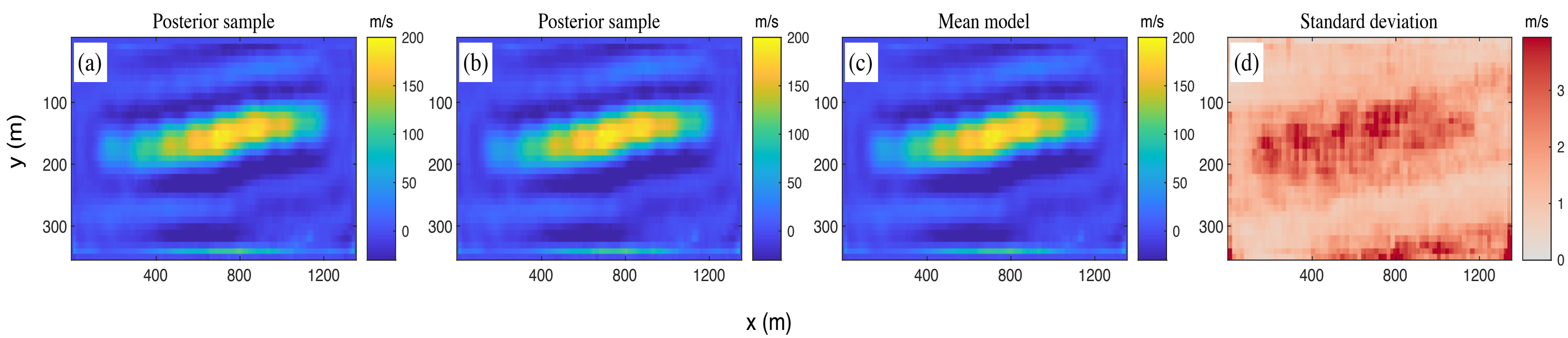
A Time-Lapse Example (for CO_2 Monitoring)

MEMORIAL UNIVERSITY
amalcolm@mun.ca

Low-f (3-4 Hz)



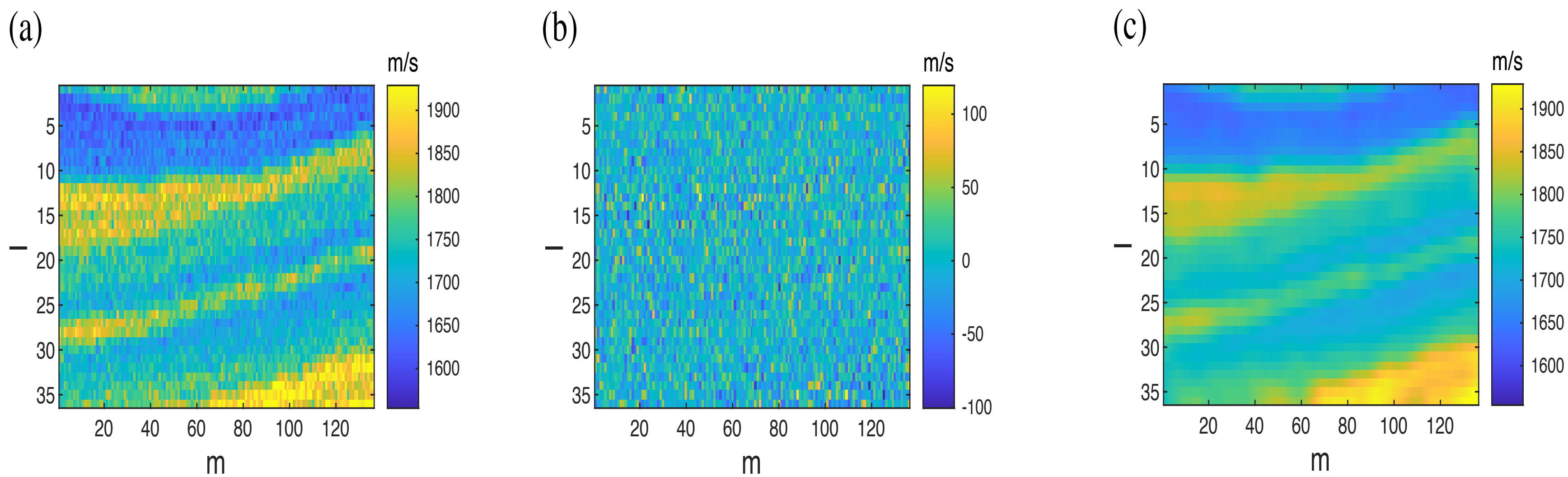
All Frequencies



A Time-Lapse Example (for CO_2 Monitoring)

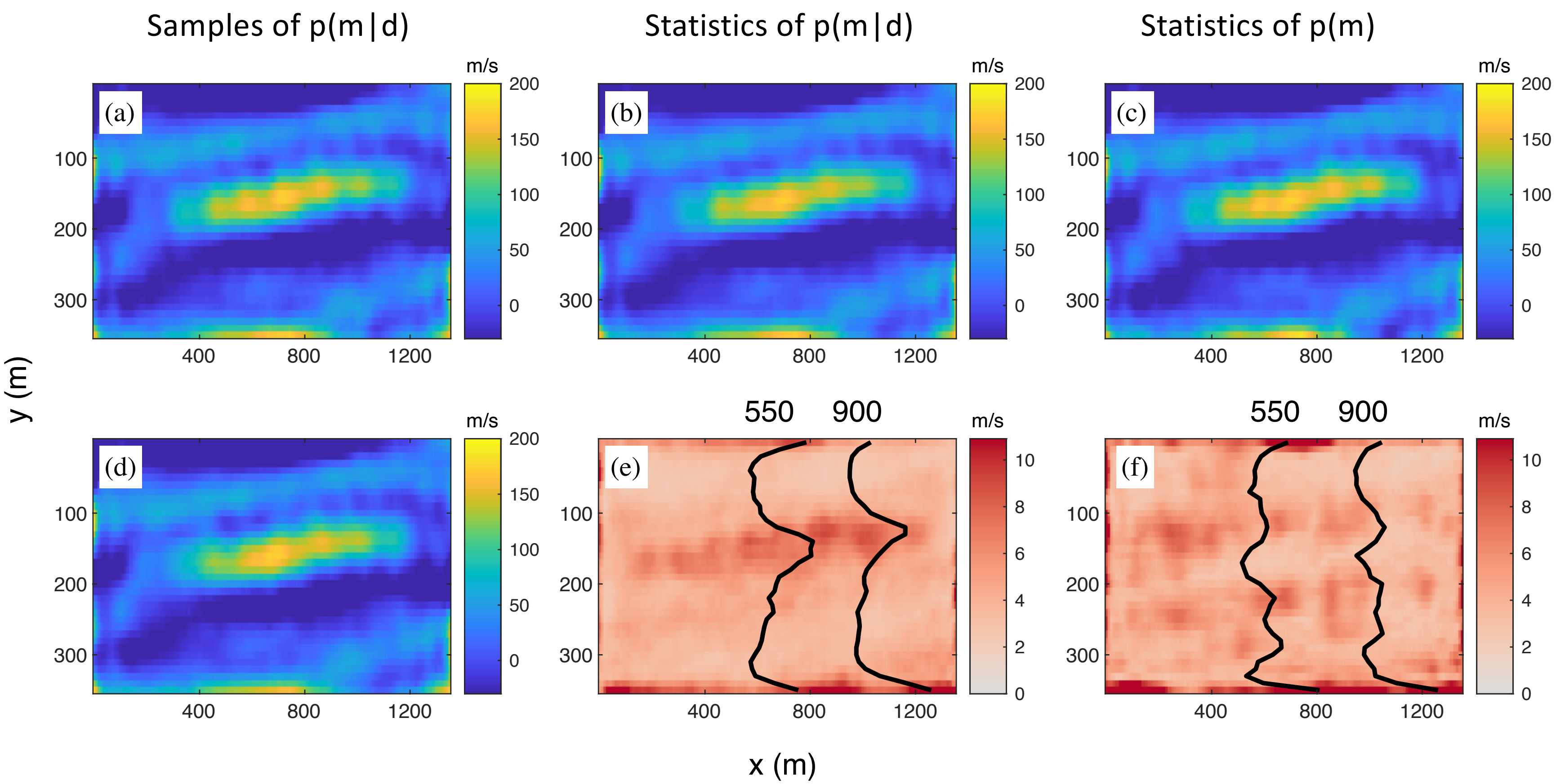
MEMORIAL UNIVERSITY
amalcolm@mun.ca

- Expand more training pairs by adding Gaussian noise into the initial model



A Time-Lapse Example (for CO_2 Monitoring)

MEMORIAL UNIVERSITY
amalcolm@mun.ca



Summary

MEMORIAL UNIVERSITY
amalcolm@mun.ca

We chose Normalizing Flows because:

- computationally feasible solution for large datasets, at least in the near future
- can handle high-dimensional space
- model agnostic – does not require any prior knowledge of distribution

Downsides:

- standard deviation from NF represents reliability measure, not an error bar like MCMC
- What distribution you are sampling depends on your training data and can be hard to tie down

Next Steps:

- Next steps are to move from just uncertainties to experimental design and answering specific questions (e.g. is it leaking?)

References

MEMORIAL UNIVERSITY
amalcolm@mun.ca

1. Kruse, J., G. Detommaso, R. Scheichl, and U. Köthe, 2019, Hint: Hierarchical invertible neural transport for density estimation and bayesian inference: arXiv preprint arXiv:1905.10687.
2. Siahkoohi, A., R. Kumar, and F. Herrmann, 2018, Seismic data reconstruction with generative adversarial networks: 80th EAGE Conference and Exhibition 2018, 1–5.
3. Siahkoohi, A., R. Kumar, and F. J. Herrmann, 2019, Deep learning based ocean bottom seismic wavefield recovery: SEG Technical Program Expanded Abstracts, 2232–2237. (SEG, San Antonio).
4. Siahkoohi, A., G. Rizzuti, M. Louboutin, P. A. Witte, and F. J. Herrmann, 2021, Preconditioned training of normalizing flows for variational inference in inverse problems: Presented at the 3rd Symposium on Advances in Approximate Bayesian Inference. AABI.
5. Siahkoohi, A., G. Rizzuti, P. A. Witte, and F. J. Herrmann, 2020, Faster uncertainty quantification for inverse problems with conditional normalizing flows: Technical Report TRCSE- 2020-2, Georgia Institute of Technology.
6. Witte, P., G. Rizzuti, M. Louboutin, A. Siahkoohi, and F. Herrmann, 2020, Invertiblenetworks.jl: A julia framework for invertible neural networks.
7. Zhao, X., A. Curtis, and X. Zhang, 2020, Bayesian seismic tomography using normalizing flows.
8. Zhao, X., and A. Curtis, 2021, Bayesian geophysical inversion using invertible neural networks: Journal of Geophysical Research: Solid Earth, 126, Issue 7.
9. Kumar, R., Kotsi, M., Siahkoohi, A. and Malcolm, A., 2021, September. Enabling uncertainty quantification for seismic data preprocessing using normalizing flows (NF)—An interpolation example. In *First International Meeting for Applied Geoscience & Energy* (pp. 1515-1519). Society of Exploration Geophysicists.
10. Willemsen, B., Malcolm, A. and Lewis, W., 2016. A numerically exact local solver applied to salt boundary inversion in seismic full-waveform inversion. *Geophysical Journal International*, *204*(3), pp.1703-1720.
11. Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. arXiv preprint arXiv:1505.05770, 2015.
12. Papamakarios, G., Nalisnick, E., Rezende, D.J., Mohamed, S. and Lakshminarayanan, B., 2021. Normalizing flows for probabilistic modeling and inference. *The Journal of Machine Learning Research*, *22*(1), pp.2617-2680.