Bayesian Imaging with Uncertainty Analysis

- Invertible Neural Networks and Normalizing Flows -

Andrew Curtis

Professor of Mathematical Geoscience
University of Edinburgh
United Kingdom

Contents

- Bayesian Inference

- Invertible Neural Networks
  - *1D Surface wave depth-inversion*
  - *2D Travel time tomography*

- Normalizing Flows
  - *2D Travel time tomography across the UK*

Bayesian Inference

- Forward and Inverse Problems

$$\mathbf{d} = \boldsymbol{f}(\mathbf{m})$$

$$\hat{\mathbf{m}} = \arg_{\mathbf{m}} \min |\boldsymbol{f}(\mathbf{m}) - \mathbf{d}_{obs}|$$

$$\{\mathbf{m}:\ \boldsymbol{f}(\mathbf{m}) - \mathbf{d}_{obs} \text{ acceptable}\}$$

$\mathbf{m}$ – model

$\mathbf{d}$ – data

$\boldsymbol{f}$ – forward function

Bayesian Inference

- Bayes' theorem

likelihood

Prior pdf

Can we estimate this using Neural Networks?

$$\underbrace{\boxed{p(\mathbf{m} \mid \mathbf{d}_{obs})}}_{\text{Posterior pdf}} = \frac{\overbrace{\boxed{p(\mathbf{d}_{obs} \mid \mathbf{m})}}^{\text{likelihood}}\ \boxed{p(\mathbf{m})}}{\underbrace{\boxed{p(\mathbf{d}_{obs})}}_{\text{evidence}}}$$

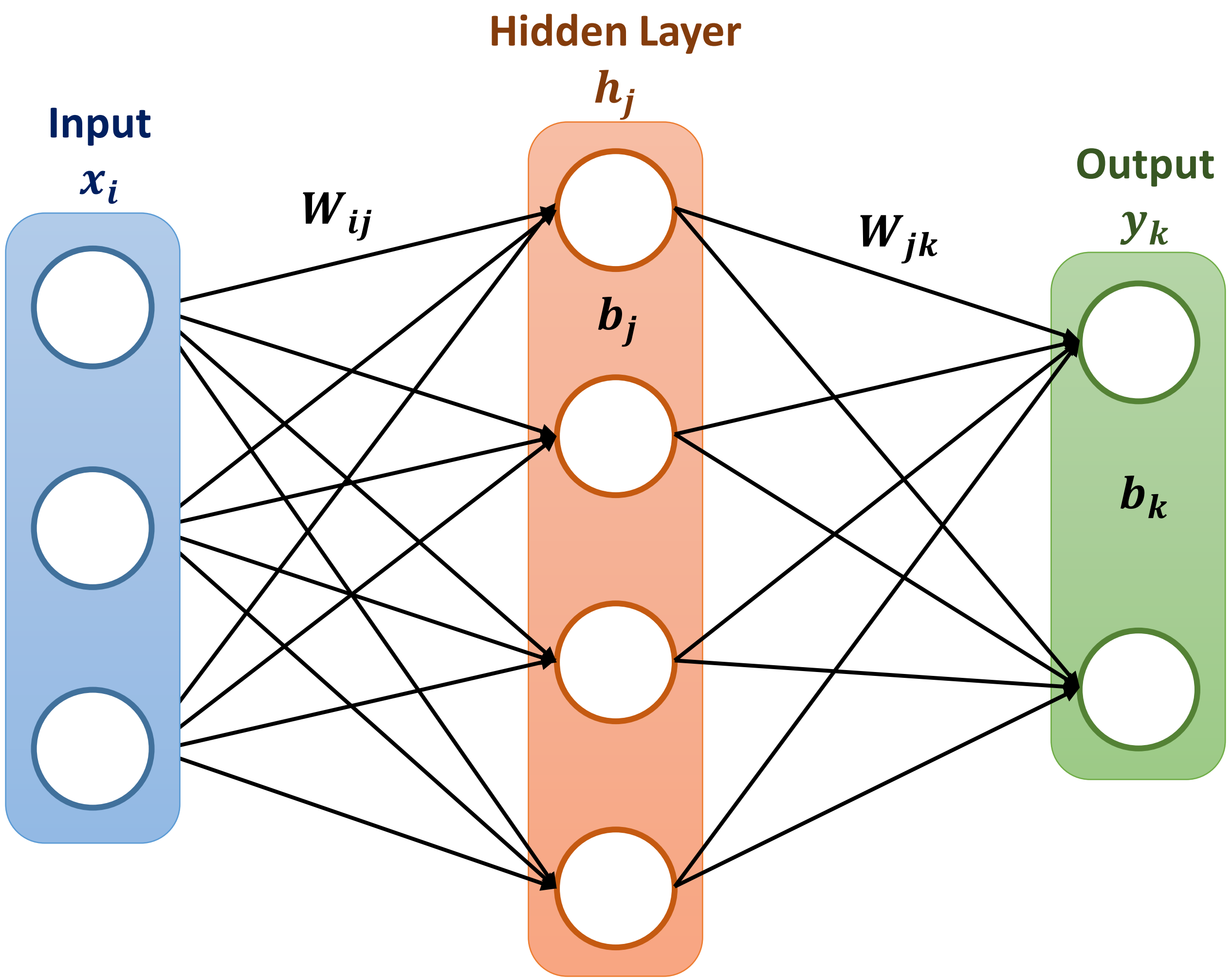
Posterior pdf

evidence

$\mathbf{m}$ = model , $\mathbf{d}_{obs}$ = observed data

pdf = probability distribution function

Neural Network

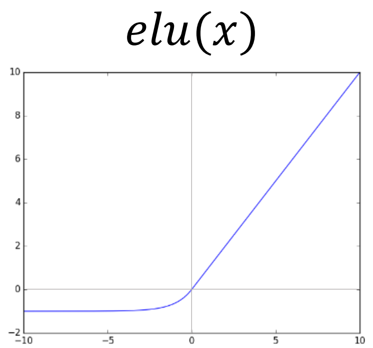


$$h_j = \sum_i W_{ij} x_i + b_j$$

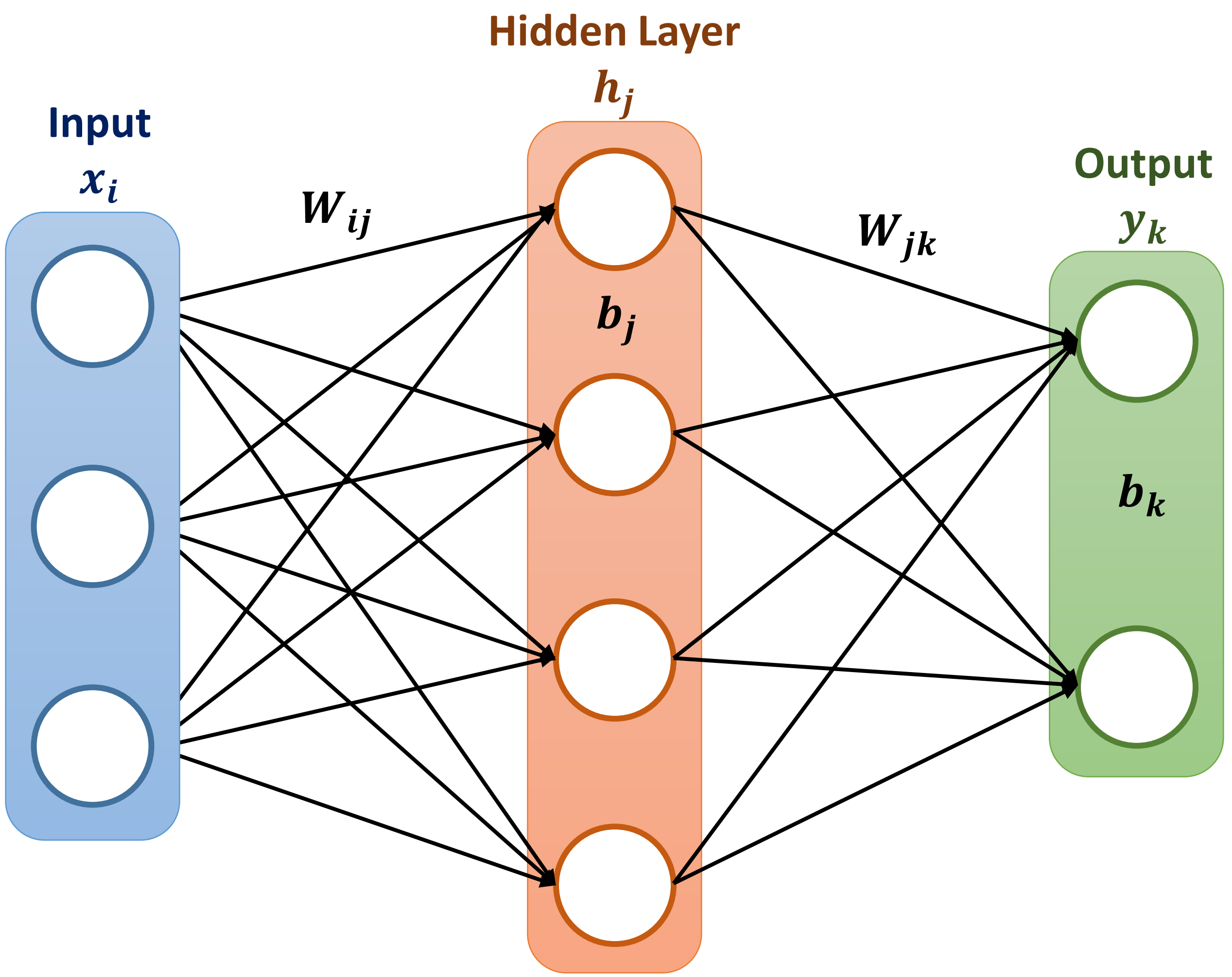
$$h'_j = elu(h_j)$$

Activation function

$$y_k = \sum_j W_{jk} h'_j + b_k$$



Neural Network



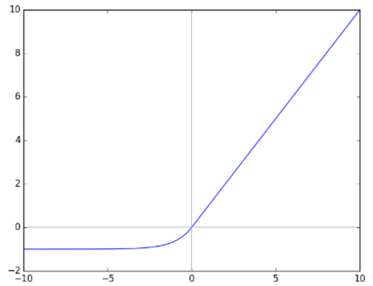
$$h_j = \sum_i W_{ij} x_i + b_j$$

$$h'_j = elu(h_j)$$

Activation function

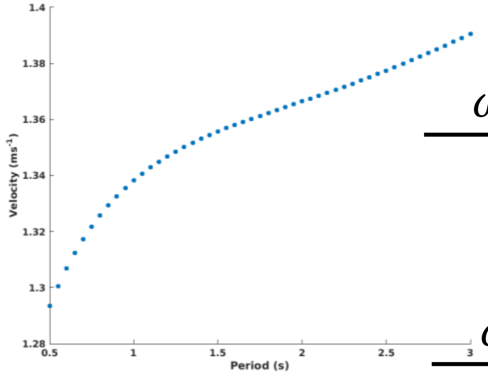
$$y_k = \sum_j W_{jk} h'_j + b_k$$

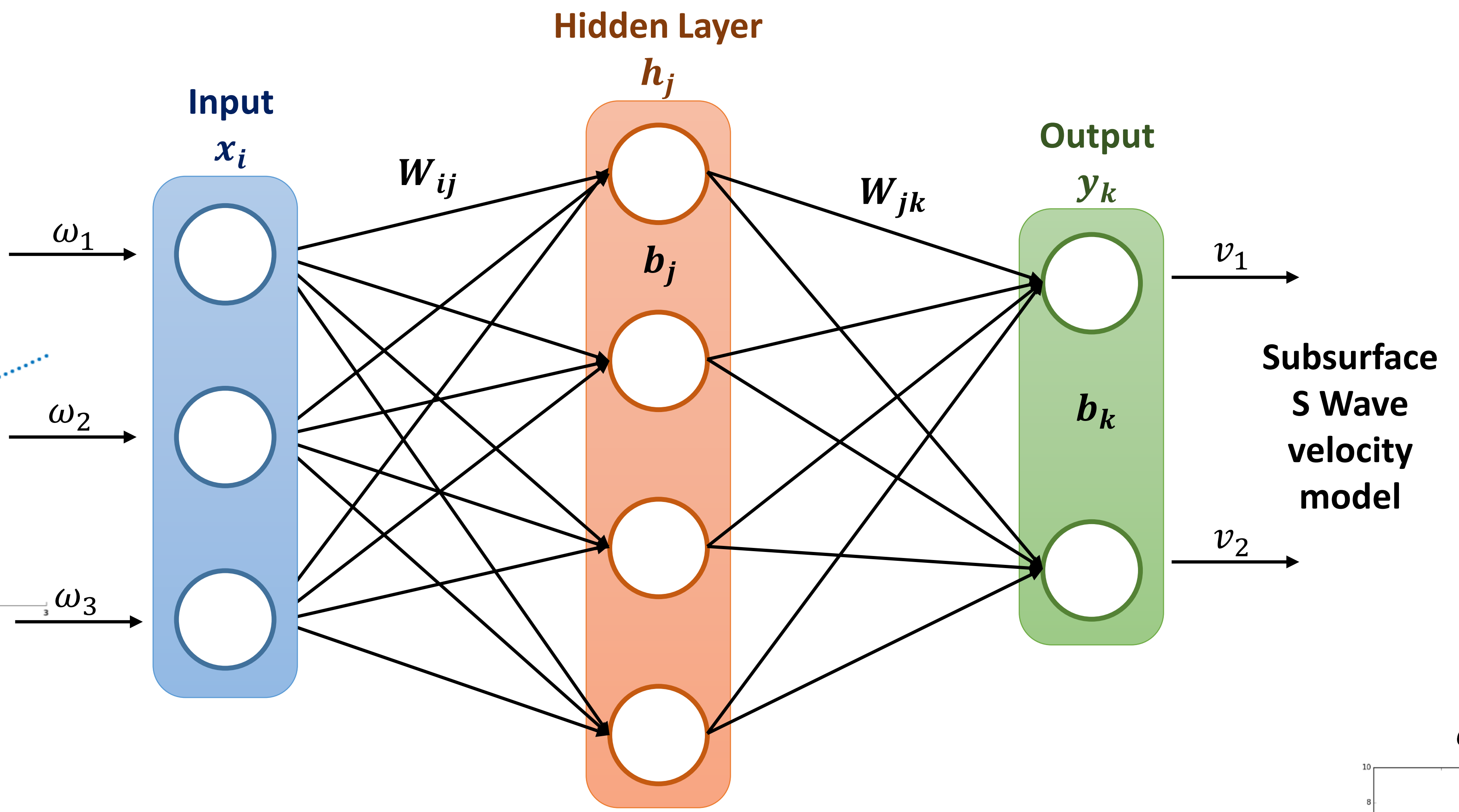
$elu(x)$



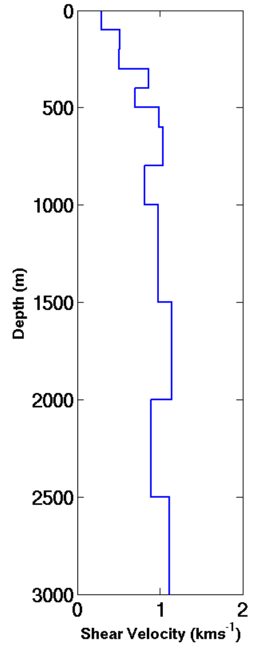
Neural Network

Dispersion Curves





Subsurface S Wave velocity model



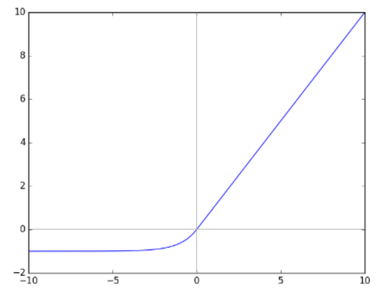
$$h_j = \sum_i W_{ij} x_i + b_j$$

$$h'_j = elu(h_j)$$

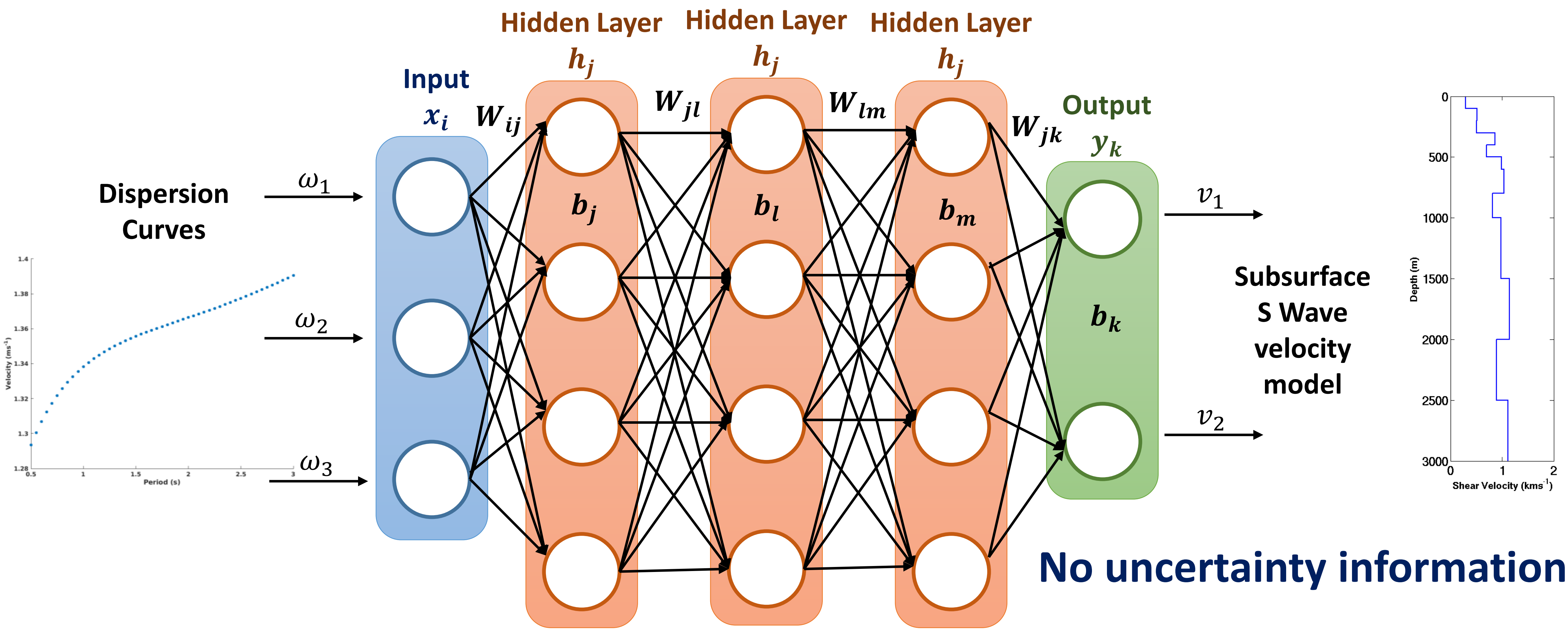
Activation function

$$y_k = \sum_j W_{jk} h'_j + b_k$$

$elu(x)$



Neural Network



No uncertainty information

Recursive, convolutional, adversarial, PINN, etc.

Mixture Density Networks

- Standard Neural Network (NN) gives no uncertainty information

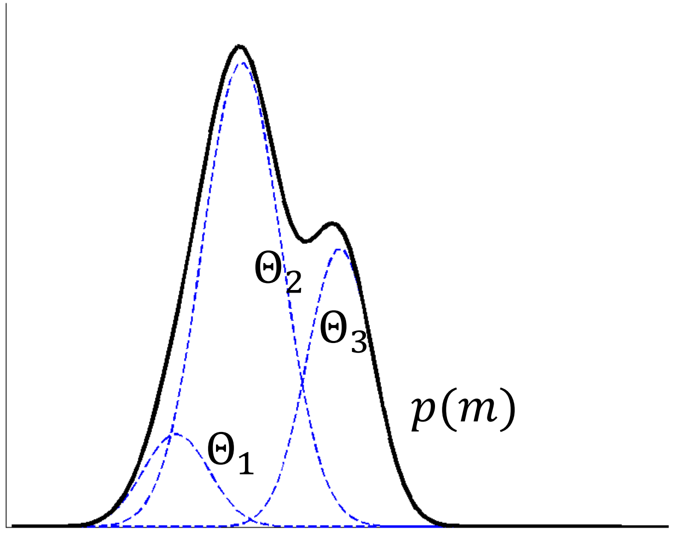
- Parameterise uncertainty using mixture of densities (→ MDN)

Weights in sum

$$p(\mathbf{m}) = \boxed{\sum_{k=1}^{M}} \boxed{\alpha_k(\boldsymbol{d})} \boxed{\Theta_k(\boldsymbol{m}|\boldsymbol{d})}$$

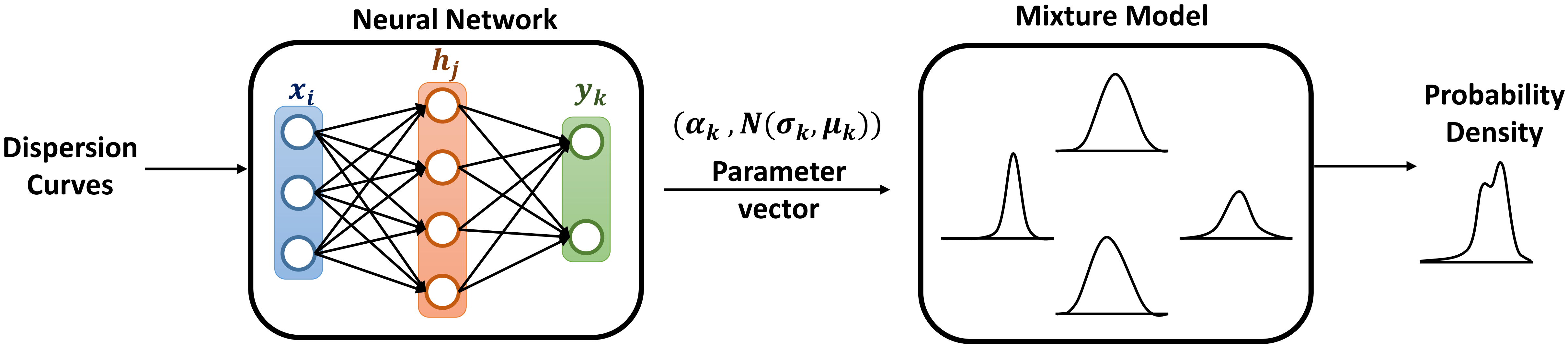
Mixture

Probability Density Kernel (Gaussian)



Mixture Density Networks

$$p(\mathbf{m}) = \sum_{k=1}^{M} \alpha_k(\boldsymbol{d})\Theta_k(\boldsymbol{m}|\boldsymbol{d})$$



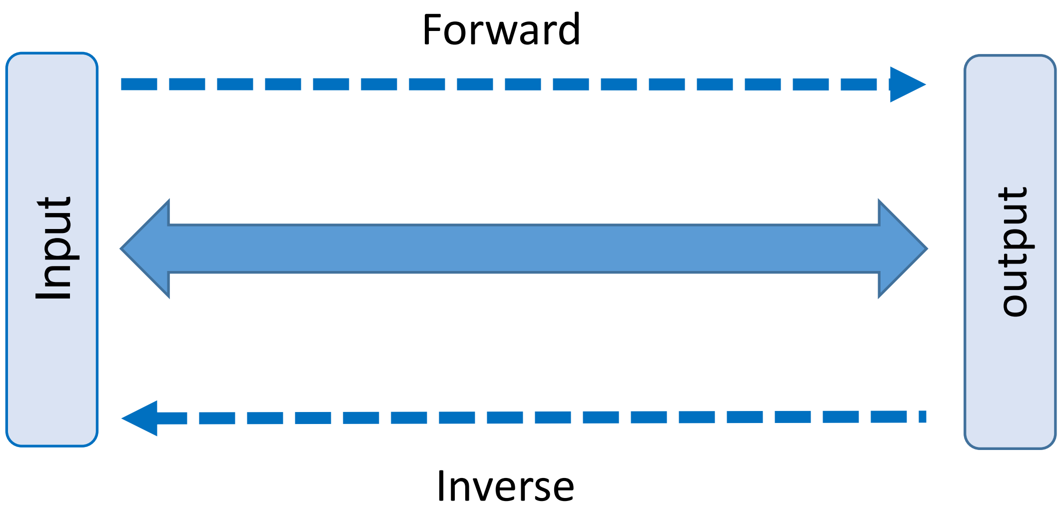
Travel-Time Tomography

Earp & Curtis 2020, *Neural Comp. Applic.*

Hard to obtain inter-parameter correlations

Invertible neural networks

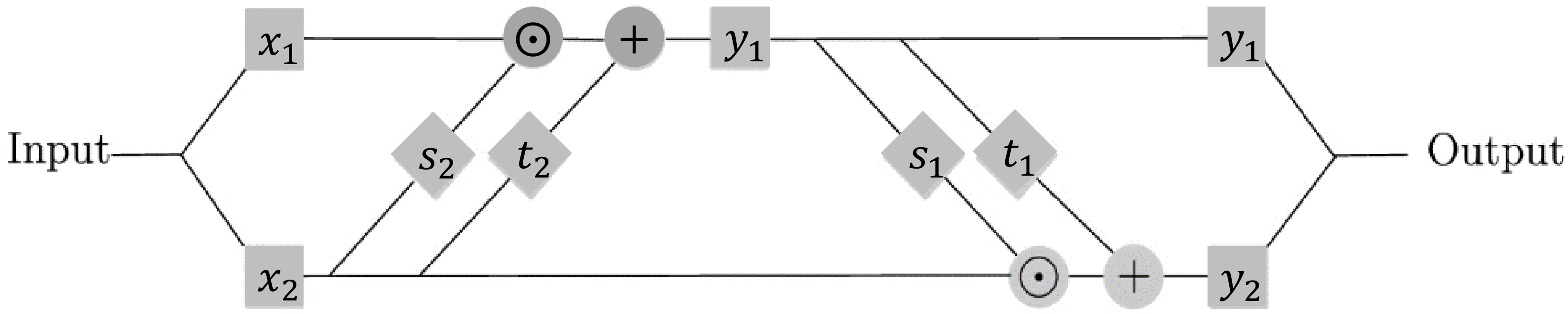
→ Approx. posterior pdf with correlations



Ardizzone et al. (2018 – arXiv)
Zhang & Curtis (2021 – J. Geophys. Res.)

Invertible neural networks

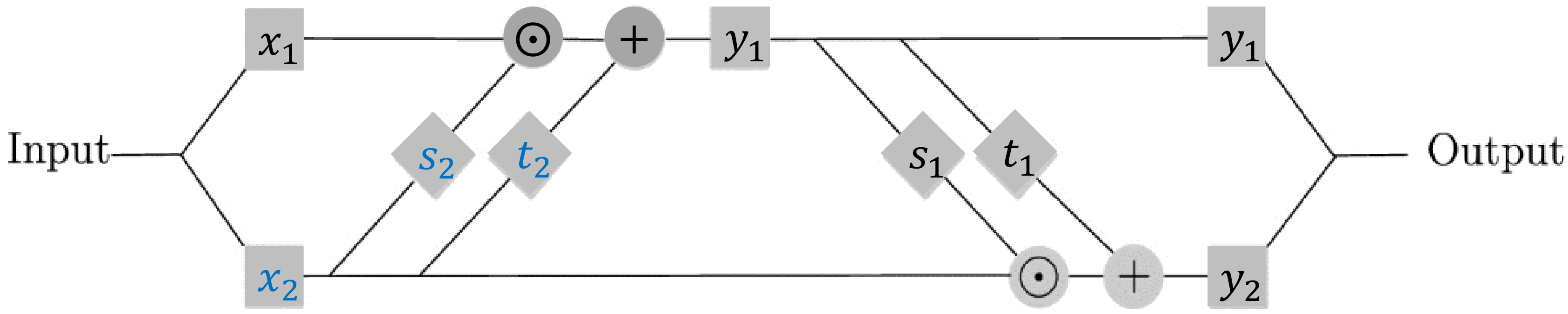
⊙ - elementwise multiplication



Ardizzone et al. (2018 – arXiv)
Zhang & Curtis (2021 – J. Geophys. Res.)

Invertible neural networks

$\odot$ - elementwise multiplication



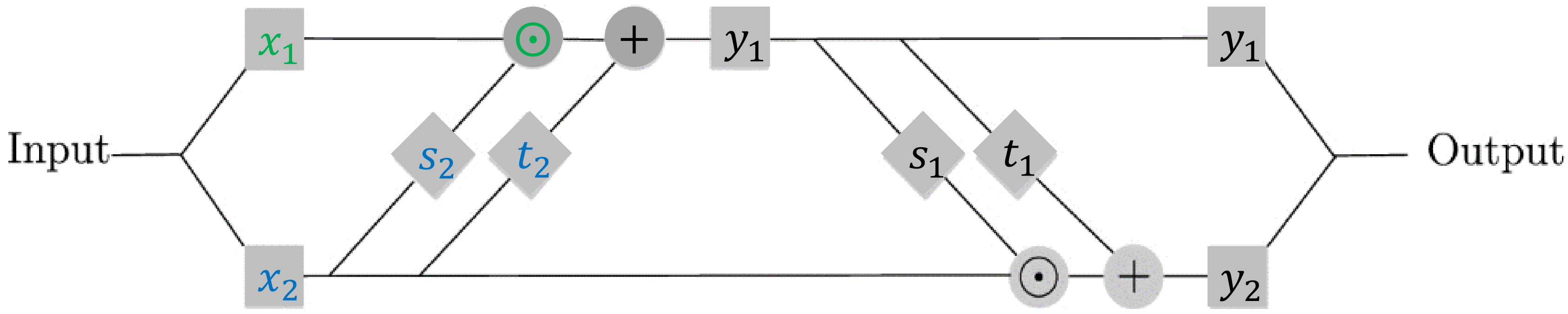
Any functions

$s_2(x_2)$ $t_2(x_2)$

Ardizzone et al. (2018 – arXiv)
Zhang & Curtis (2021 – J. Geophys. Res.)

Invertible neural networks

$\odot$ - elementwise multiplication

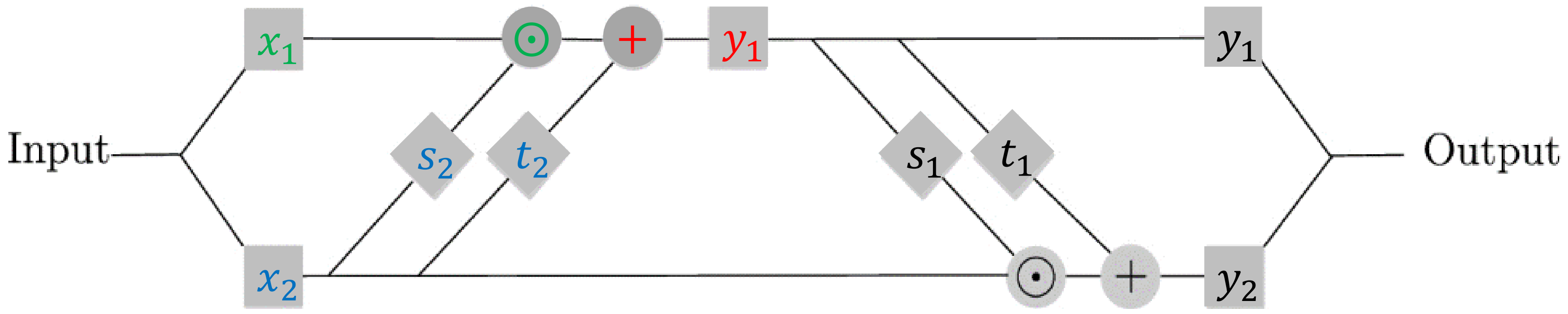


$x_1 \odot \exp(s_2(x_2))$ $\quad$ $t_2(x_2)$

Ardizzone et al. (2018 – arXiv)
Zhang & Curtis (2021 – J. Geophys. Res.)

Invertible neural networks

$\odot$ - elementwise multiplication

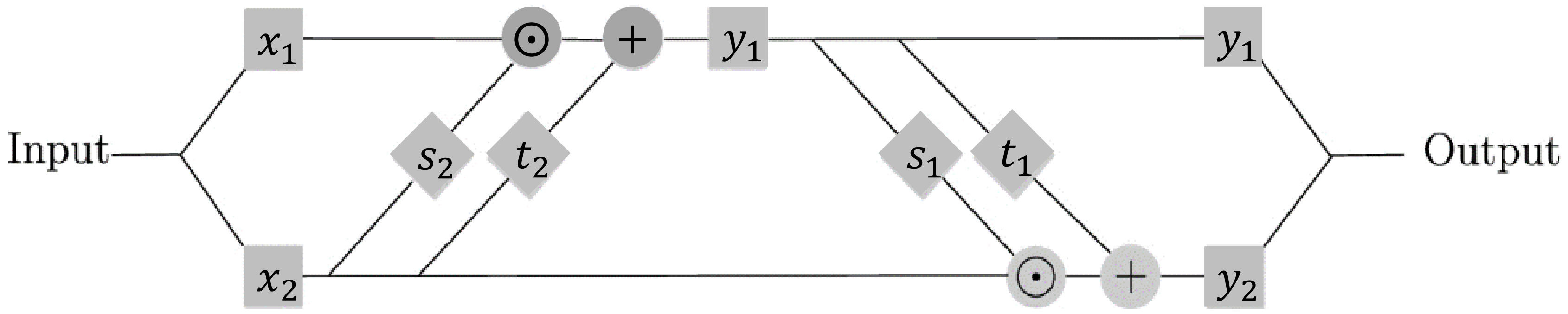


$$y_1 = x_1 \odot \exp(s_2(x_2)) + t_2(x_2)$$

Ardizzone et al. (2018 – arXiv)
Zhang & Curtis (2021 – J. Geophys. Res.)

Invertible neural networks

$\odot$ - elementwise multiplication

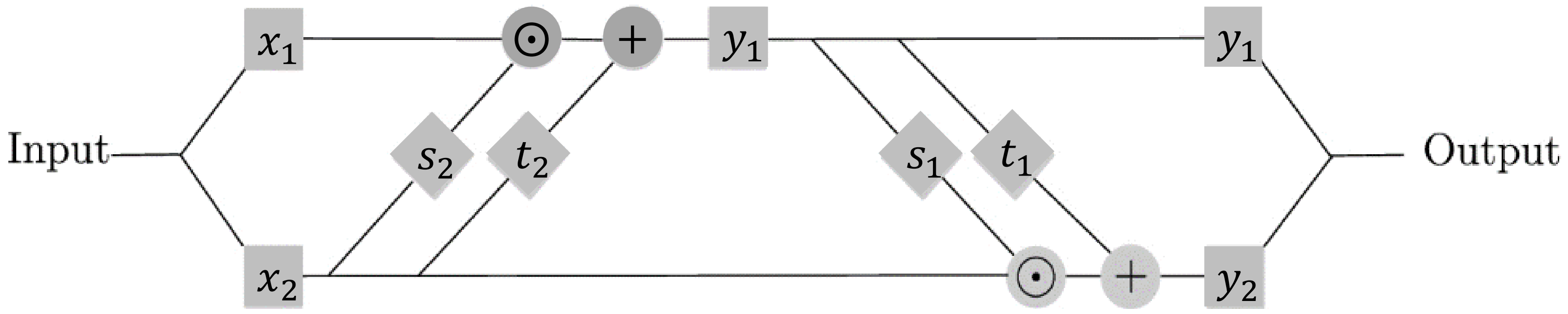


$$y_1 = x_1 \odot \exp(s_2(x_2)) + t_2(x_2)$$
$$y_2 = x_2 \odot \exp(s_1(y_1)) + t_1(y_1)$$

Ardizzone et al. (2018 – arXiv)
Zhang & Curtis (2021 – J. Geophys. Res.)

Invertible neural networks

$\odot$ - elementwise multiplication



$$y_1 = x_1 \odot \exp(s_2(x_2)) + t_2(x_2)$$
$$y_2 = x_2 \odot \exp(s_1(y_1)) + t_1(y_1)$$

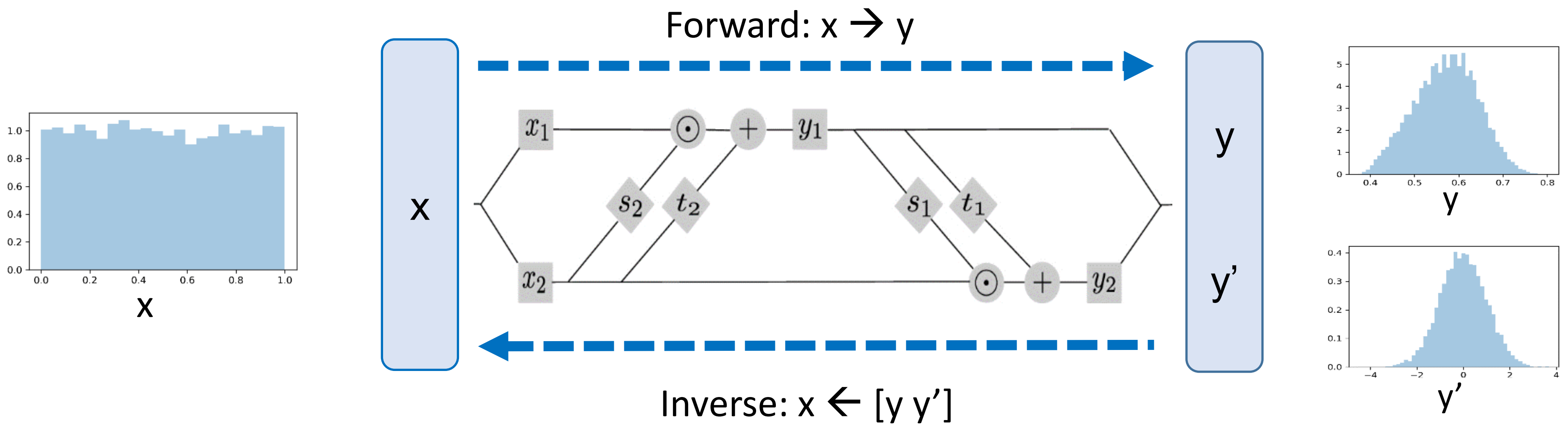
$$x_1 = (y_1 - t_2(x_2)) \odot \exp(-s_2(x_2))$$
$$x_2 = \left(y_2 - t_1(y_1)\right) \odot \exp\left(-s_1(y_1)\right)$$

s_1, s_2, t_1, t_2 can be any functions, e.g., fully connected NN, CNN, PINN, …

Coupling Structure

Ardizzone et al. (2018 – arXiv)
Zhang & Curtis (2021 – J. Geophys. Res.)

Invertible neural networks



auxiliary variable to introduce randomness to the output of Inverse

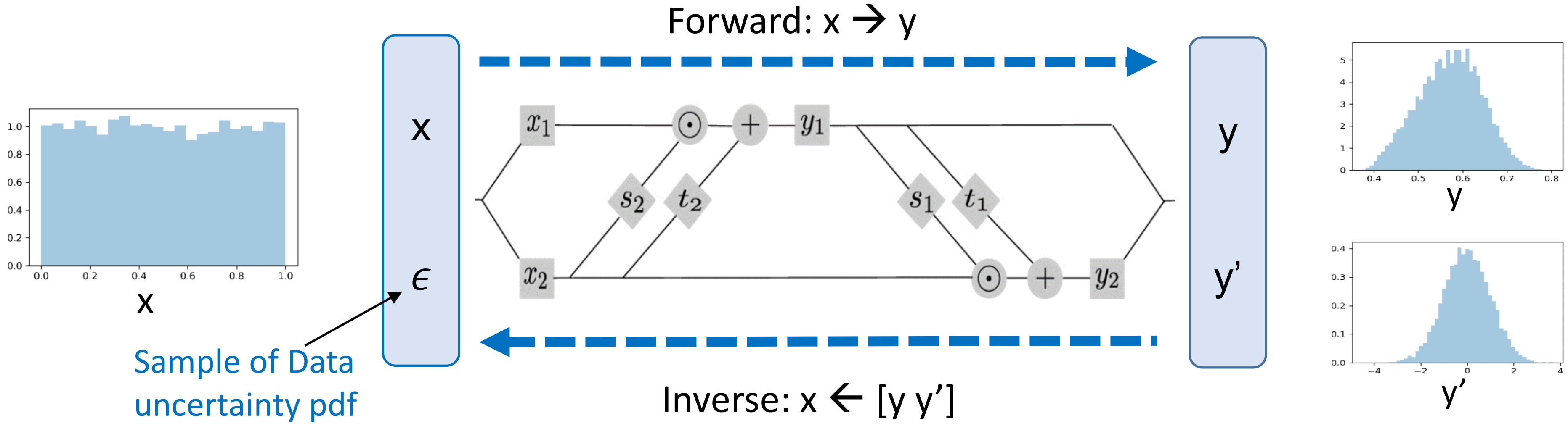
$$\text{Min: } \|y - nn(x)\| + MMD[p(y, y'), nn(p(x))] \quad \text{where } y' \sim N(0,1)$$

- **Maximum Mean Discrepancy (MMD)**: a measure of difference between two distributions

$$MMD(P,Q) = E_{X,X'\sim P}[k(X,X')] + E_{Y,Y'\sim Q}[k(Y,Y')] - 2E_{X\sim P,Y\sim Q}[k(X,Y)]$$

Ardizzone et al. (2018 – arXiv)
Zhang & Curtis (2021 – J. Geophys. Res.)

Invertible neural networks



auxiliary variable to introduce randomness to the output of Inverse

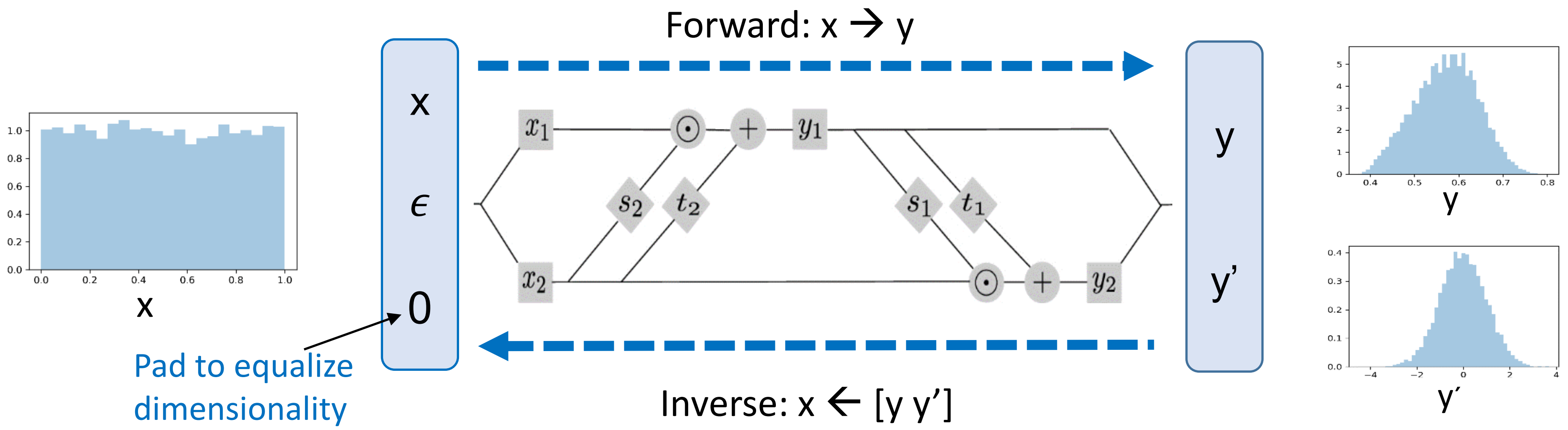
Min: $\|y - nn(x,\epsilon)\| + MMD[p(y,y'), nn(p(x,\epsilon))]$ where $y' \sim N(0,1)$

- **Maximum Mean Discrepancy (MMD)**: a measure of difference between two distributions

$$MMD(P,Q) = E_{X,X' \sim P}[k(X,X')] + E_{Y,Y' \sim Q}[k(Y,Y')] - 2E_{X\sim P, Y\sim Q}[k(X,Y)]$$

Ardizzone et al. (2018 – arXiv)
Zhang & Curtis (2021 – J. Geophys. Res.)

Invertible neural networks



auxiliary variable to introduce randomness to the output of Inverse

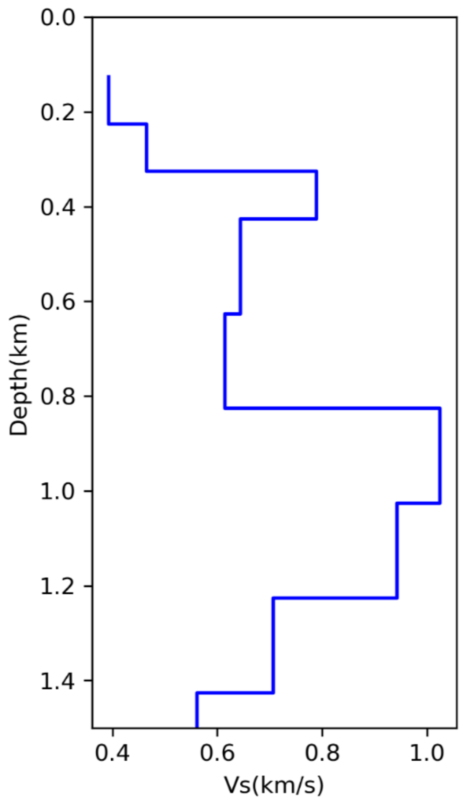
$$\text{Min: } \|y - nn(x,\epsilon)\| + MMD[p(y,y'), nn(p(x,\epsilon))] \quad \text{where } y' \sim N(0,1)$$

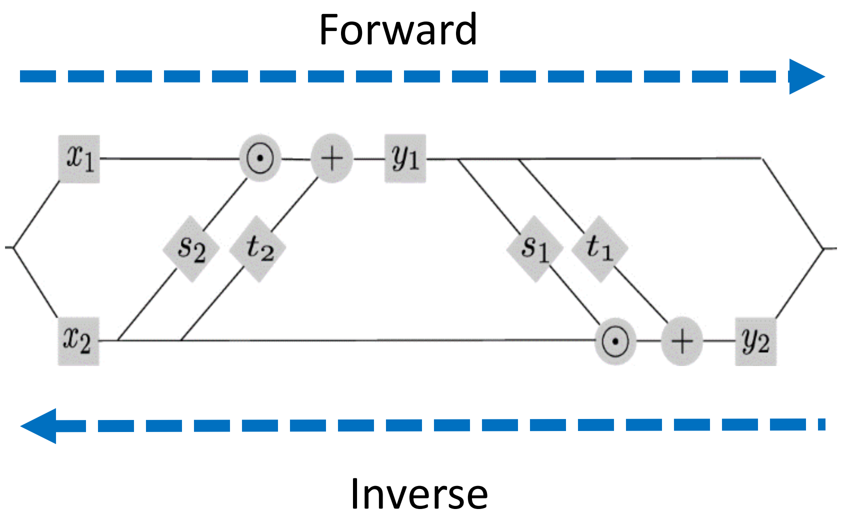
- **Maximum Mean Discrepancy (MMD)**: a measure of difference between two distributions

$$MMD(P,Q) = E_{X,X' \sim P}[k(X,X')] + E_{Y,Y' \sim Q}[k(Y,Y')] - 2E_{X\sim P, Y\sim Q}[k(X,Y)]$$

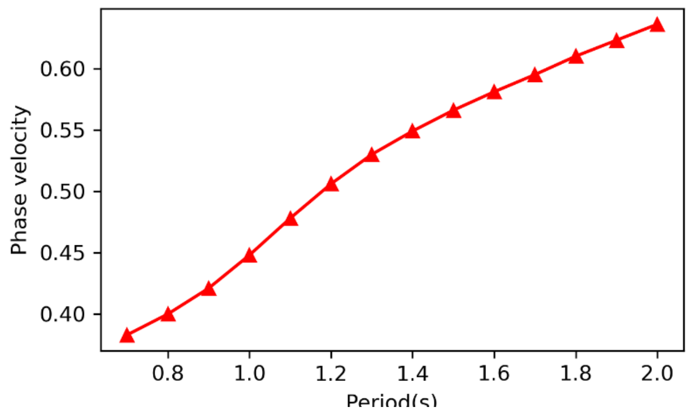
Ardizzone et al. (2018 – arXiv)
Zhang & Curtis (2021 – J. Geophys. Res.)

Surface wave dispersion inversion



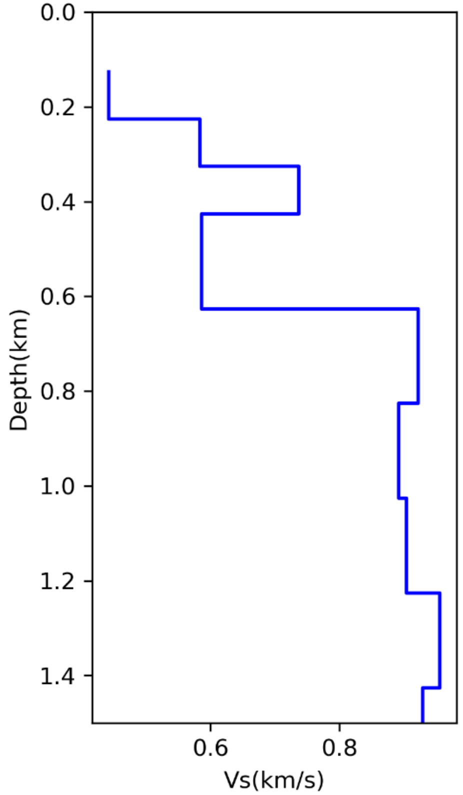


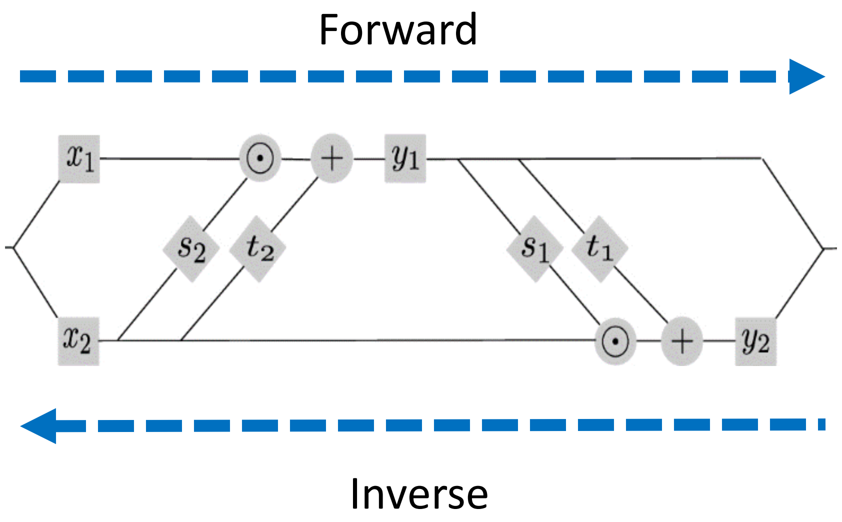
Phase velocity

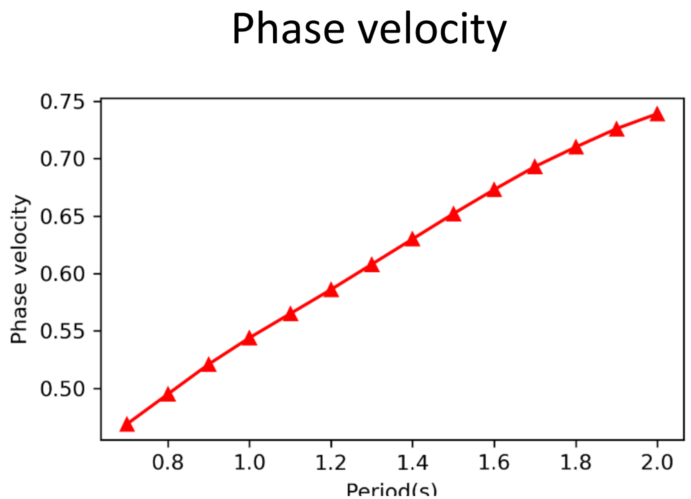


Zhang & Curtis (2021 – J. Geophys. Res.)

Surface wave dispersion inversion

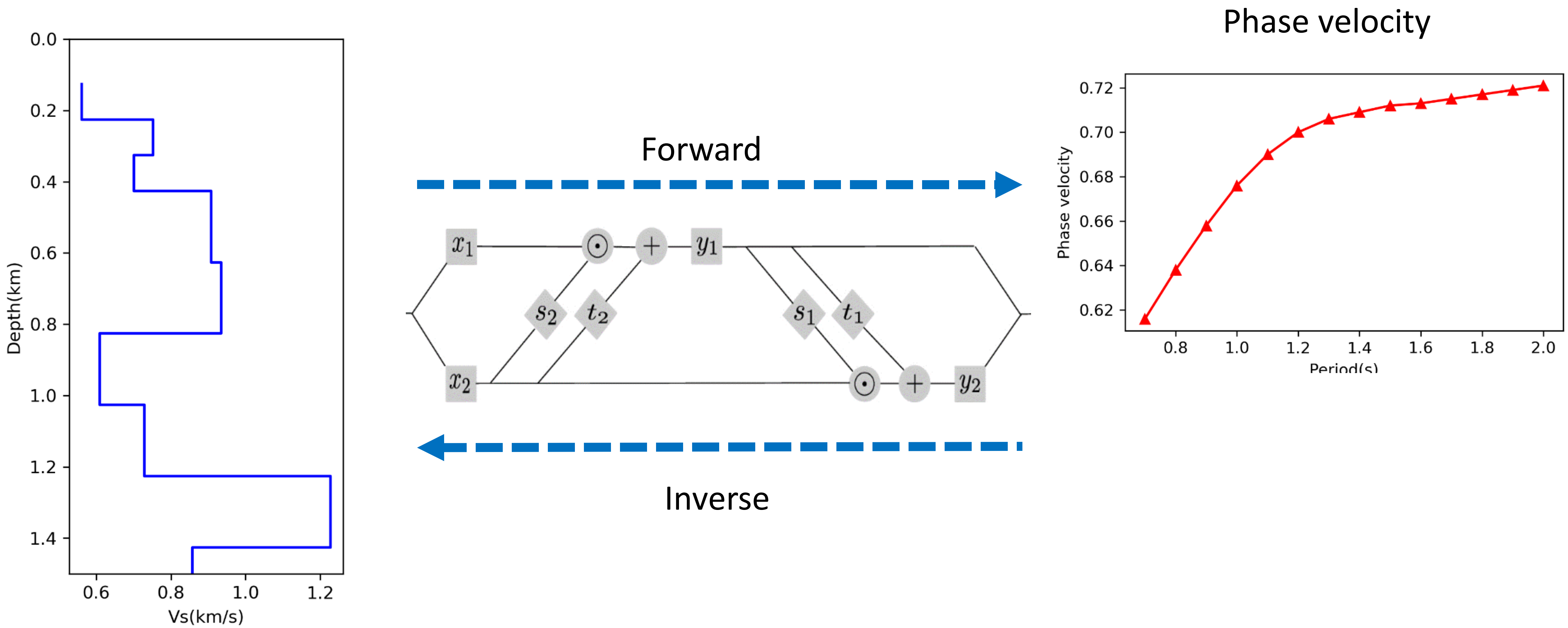






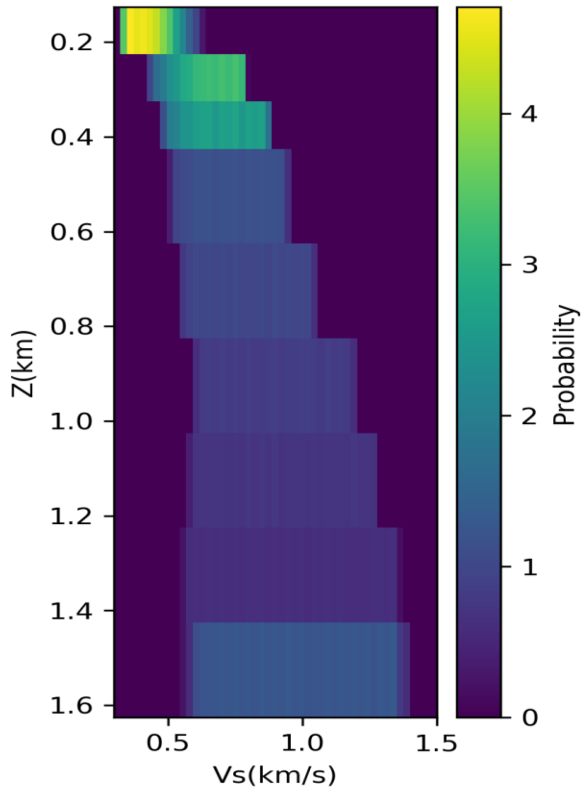
Zhang & Curtis (2021 – J. Geophys. Res.)

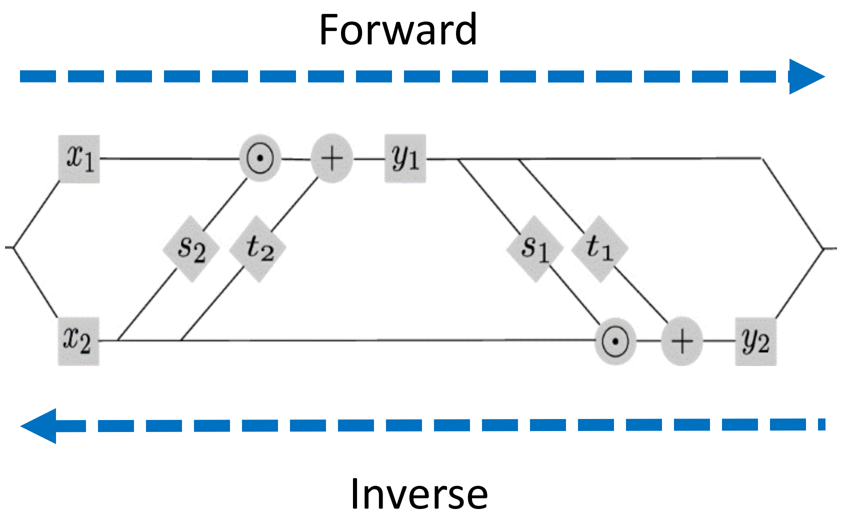
Surface wave dispersion inversion

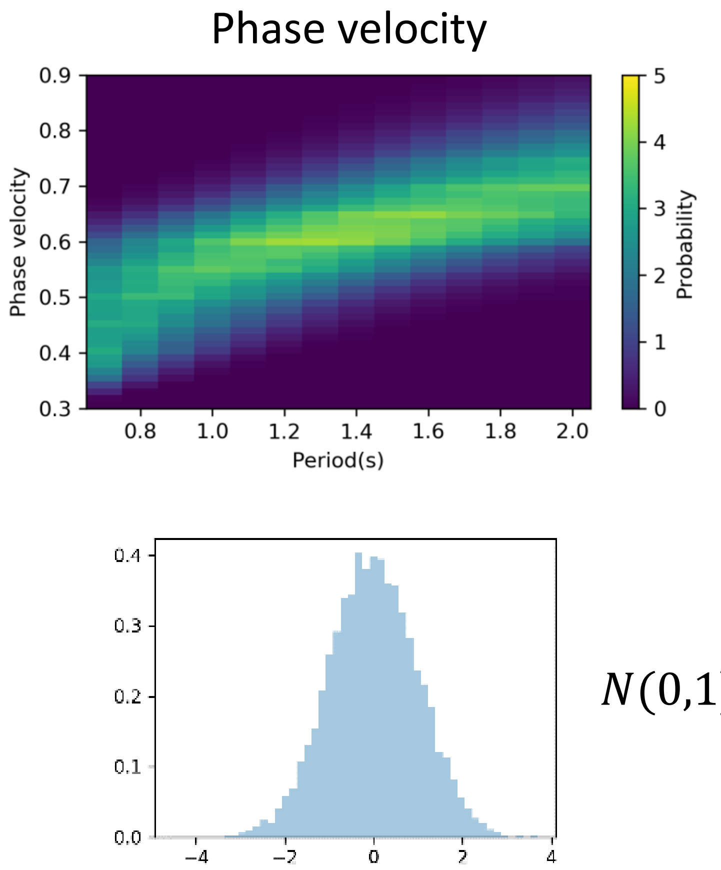


Zhang & Curtis (2021 – J. Geophys. Res.)

Surface wave dispersion inversion – **Training Set**







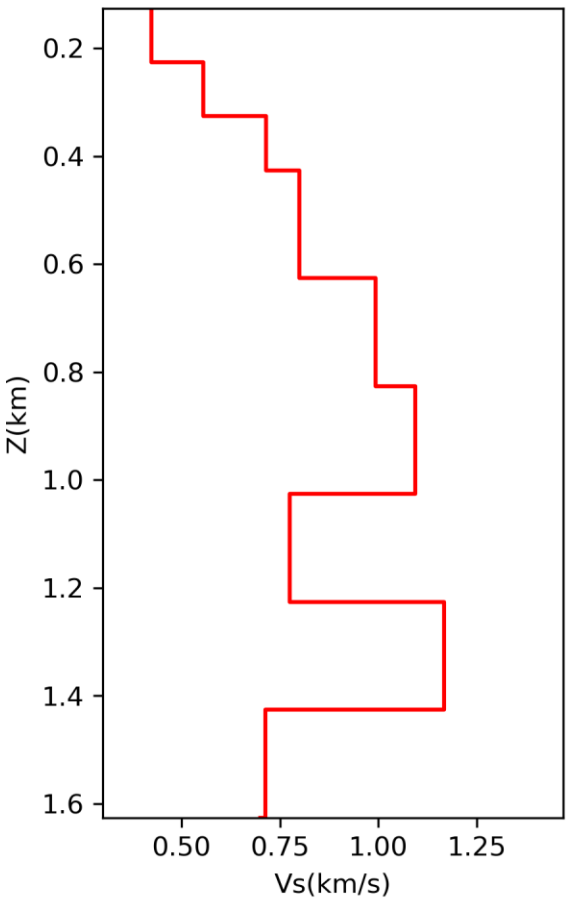
$N(0,1)$

- 100,000 training data
- Fully connected sub-networks (s1,t1,s2,t2)

Zhang & Curtis (2021 – J. Geophys. Res.)

Examples

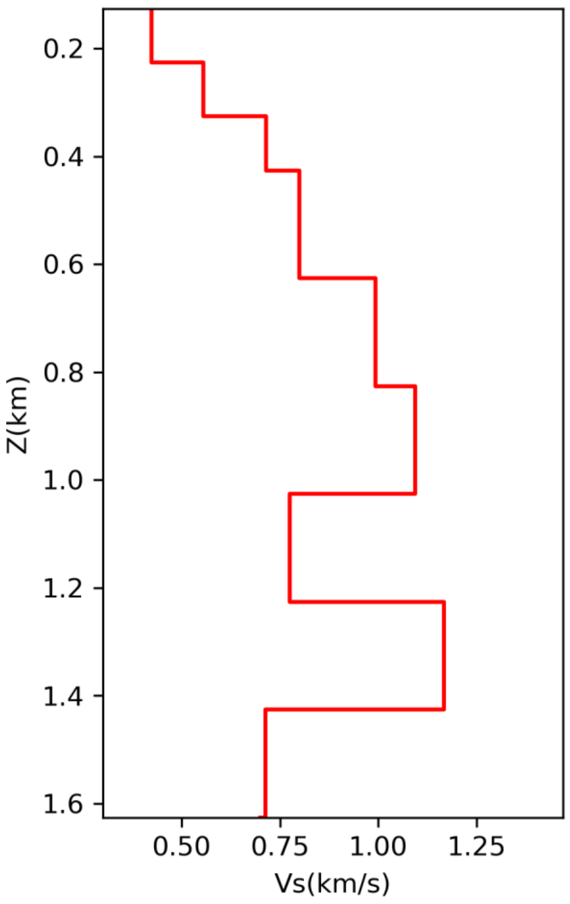
True model



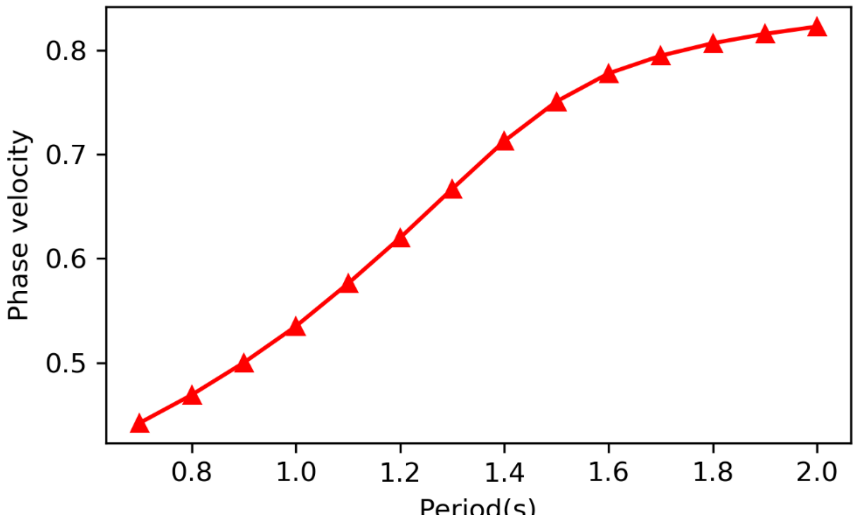
Zhang & Curtis (2021 – J. Geophys. Res.)

Examples

True model



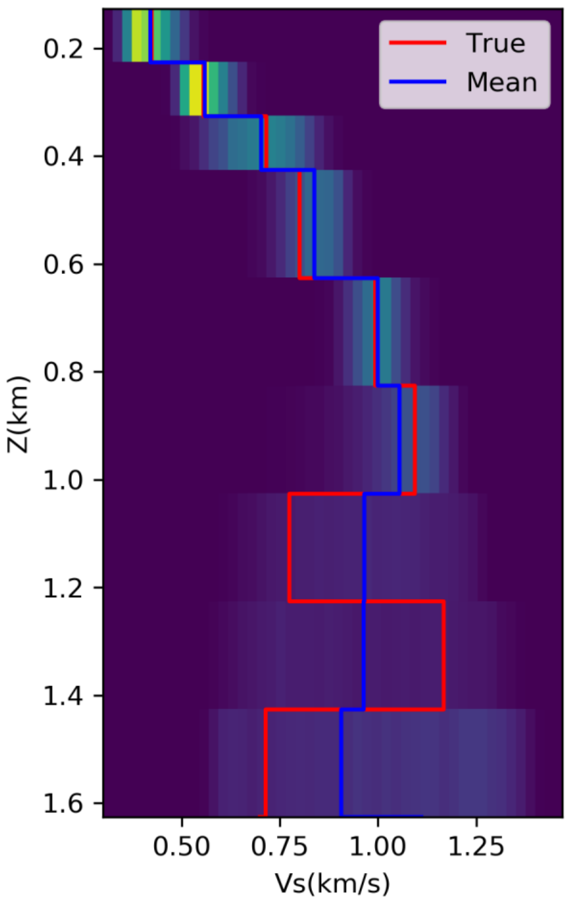
Phase velocity



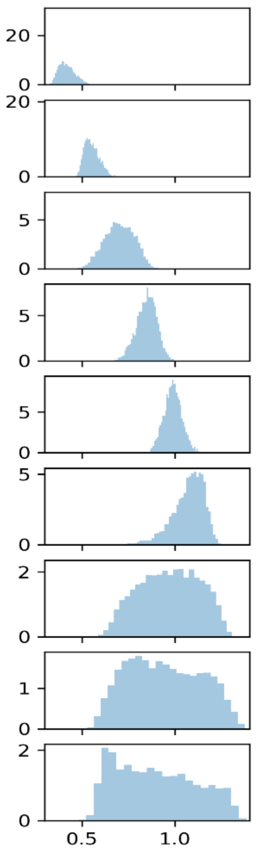
Zhang & Curtis (2021 – J. Geophys. Res.)

Posterior pdfs

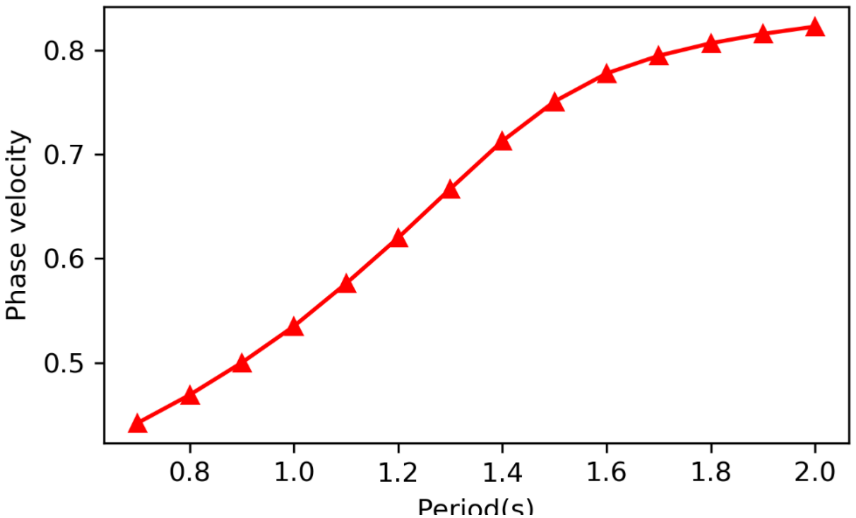
INN



INN



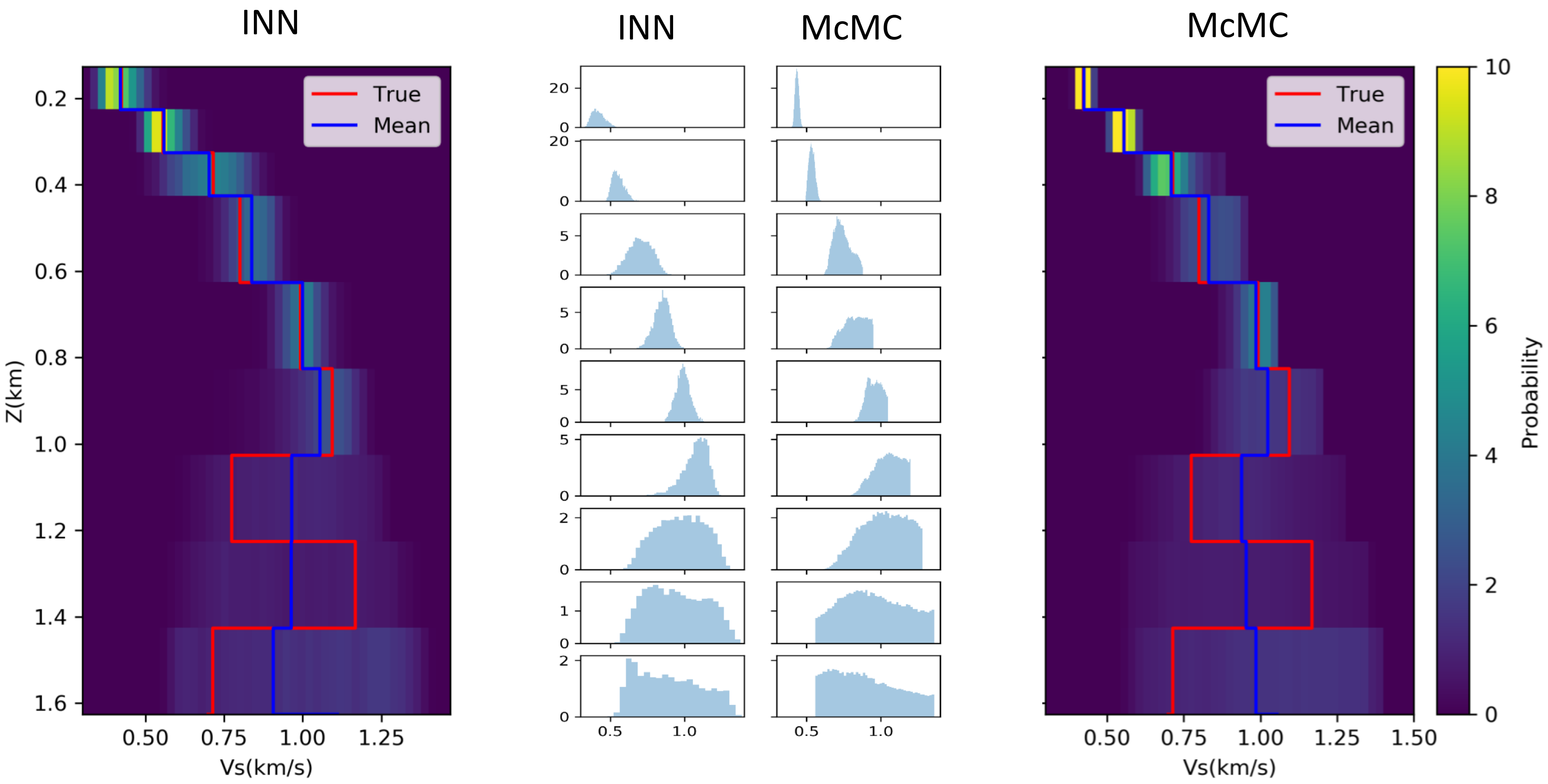
Phase velocity



2 seconds

Zhang & Curtis (2021 – J. Geophys. Res.)

Posterior pdfs

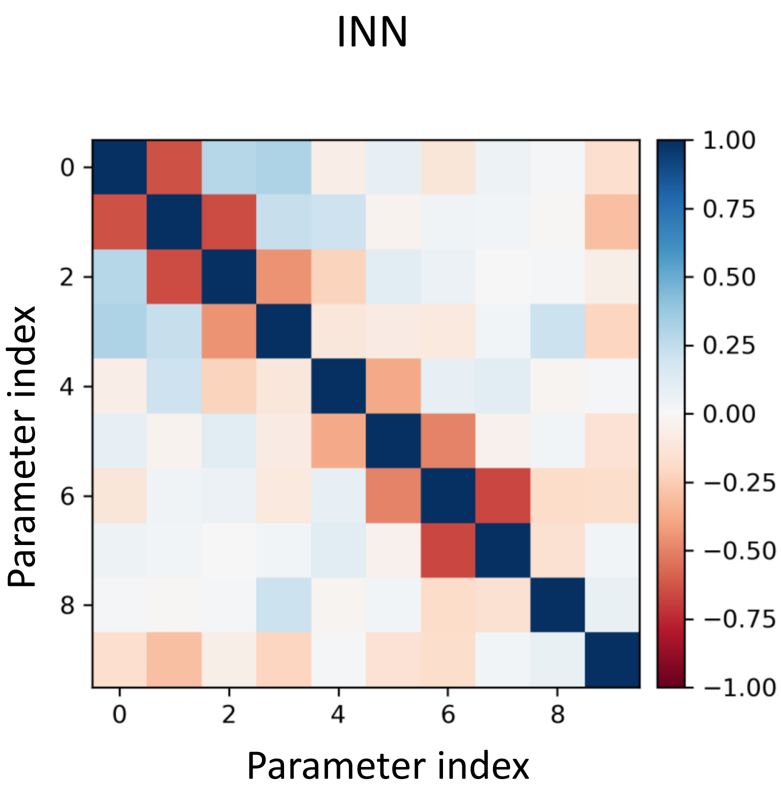


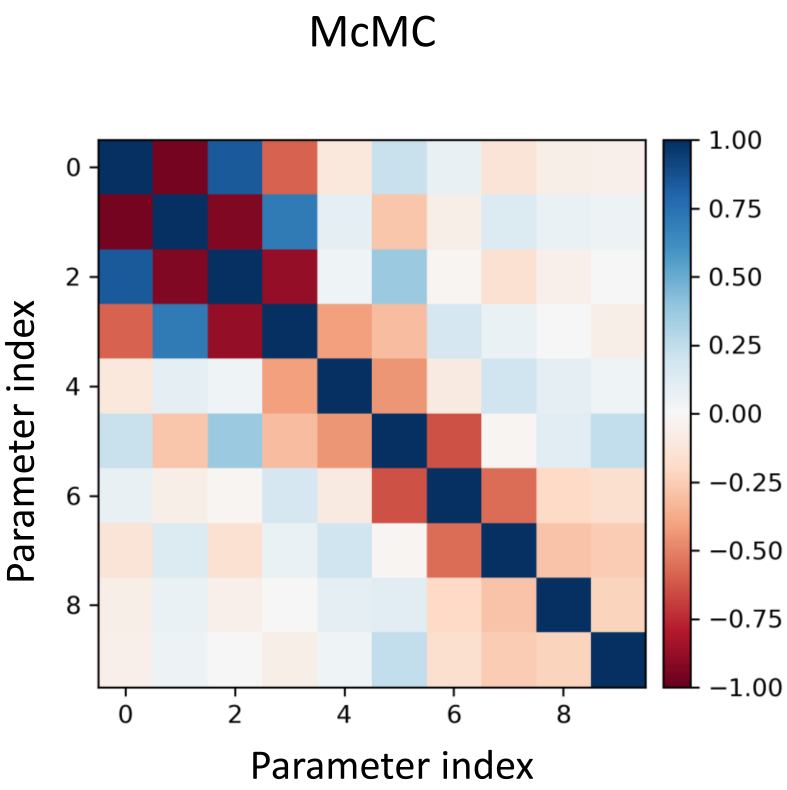
2 seconds

2 hours

Zhang & Curtis (2021 – J. Geophys. Res.)

Covariance matrix

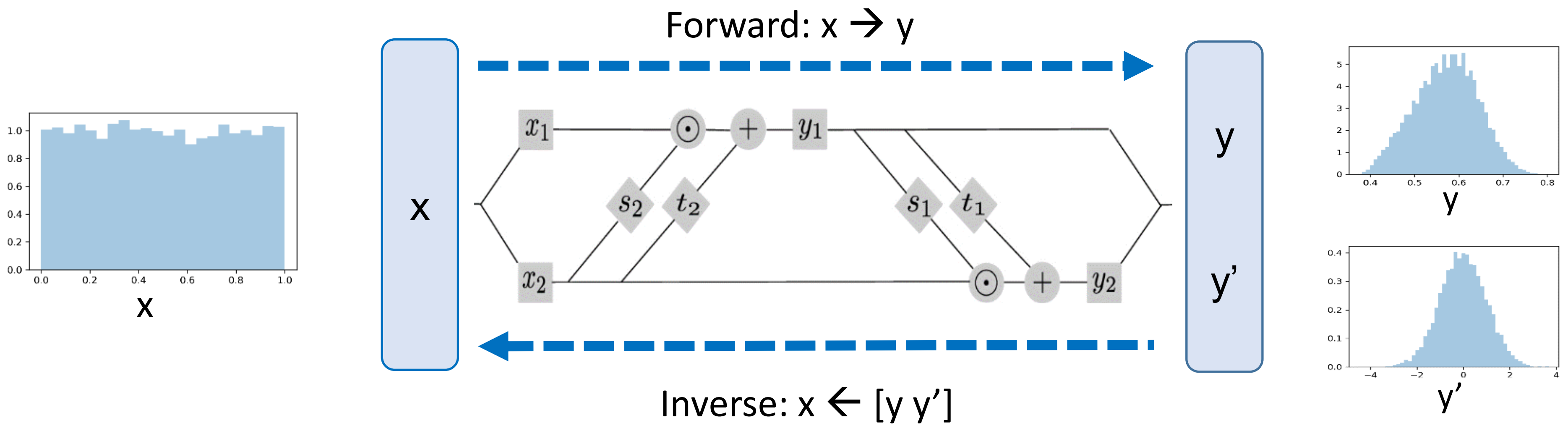




Hard to obtain from MDNs

Zhang & Curtis (2021 – J. Geophys. Res.)

Invertible neural networks

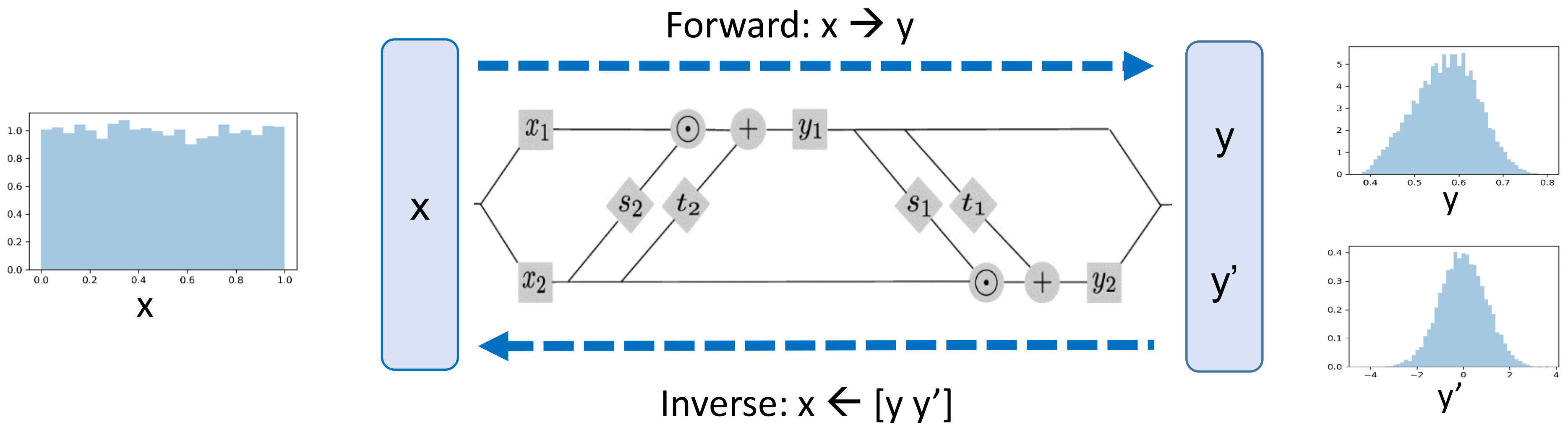


$$\text{Min } \|y - nn(x)\| + MMD[p(y,y'), nn(p(x))] \quad \text{where } y' \sim N(0,1)$$

- Does not work well in high dimensionality because of the MMD measure

Zhang & Curtis (2021 – J. Geophys. Res.)

Invertible neural networks

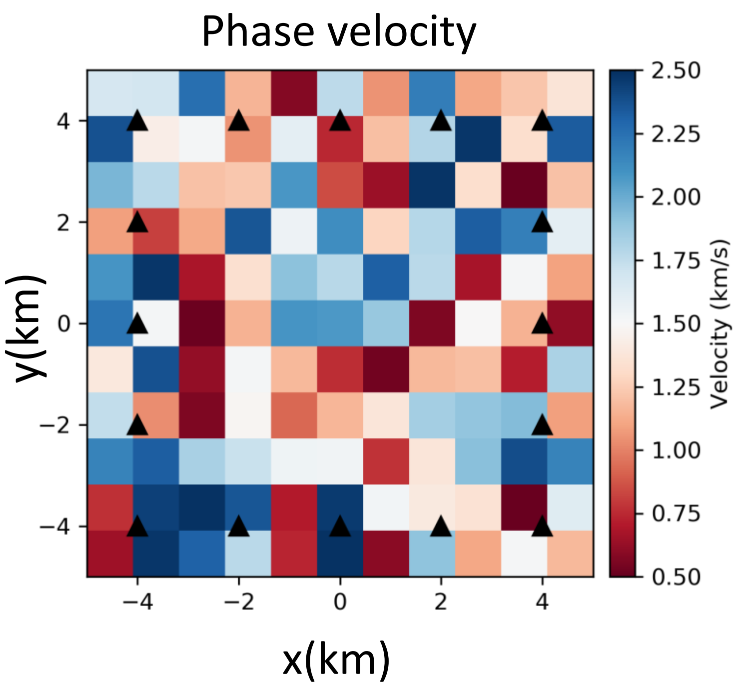


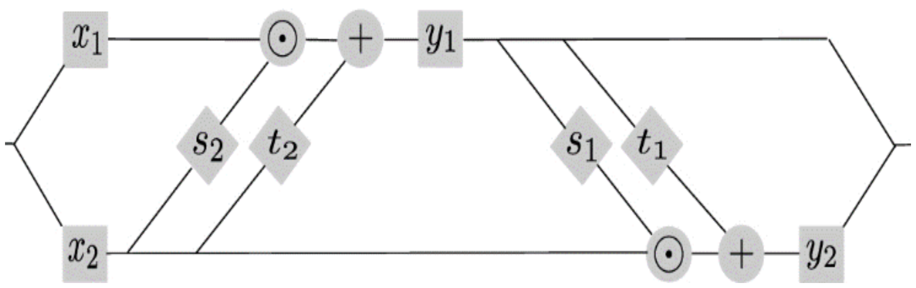
$$\text{Min } \|y - nn(x)\| + \left(-p[nn^{-1}(y,y')]\,|det\mathbf{J}_{nn^{-1}}(y,y')|\right) \text{ where } y' \sim N(0,1)$$

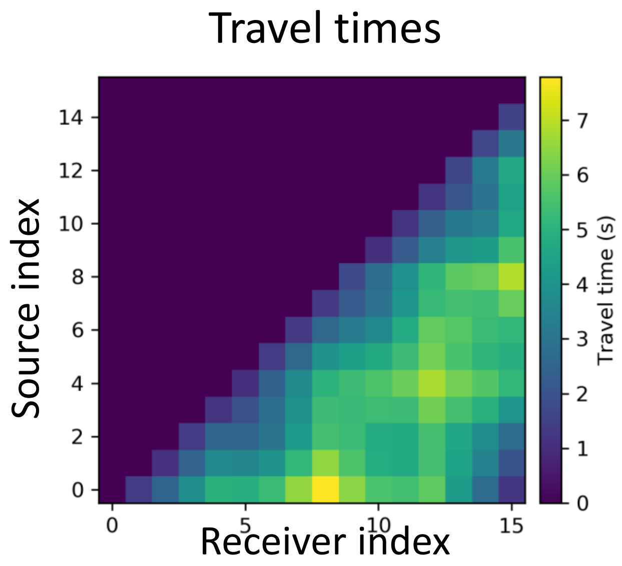
- Minimize MMD → Maximize prior likelihood of (y,y')

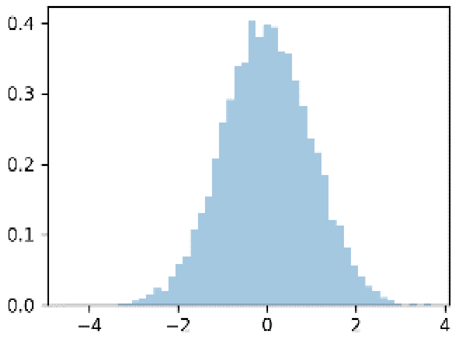
Zhang & Curtis (2021 – J. Geophys. Res.)

2D Travel time Tomography







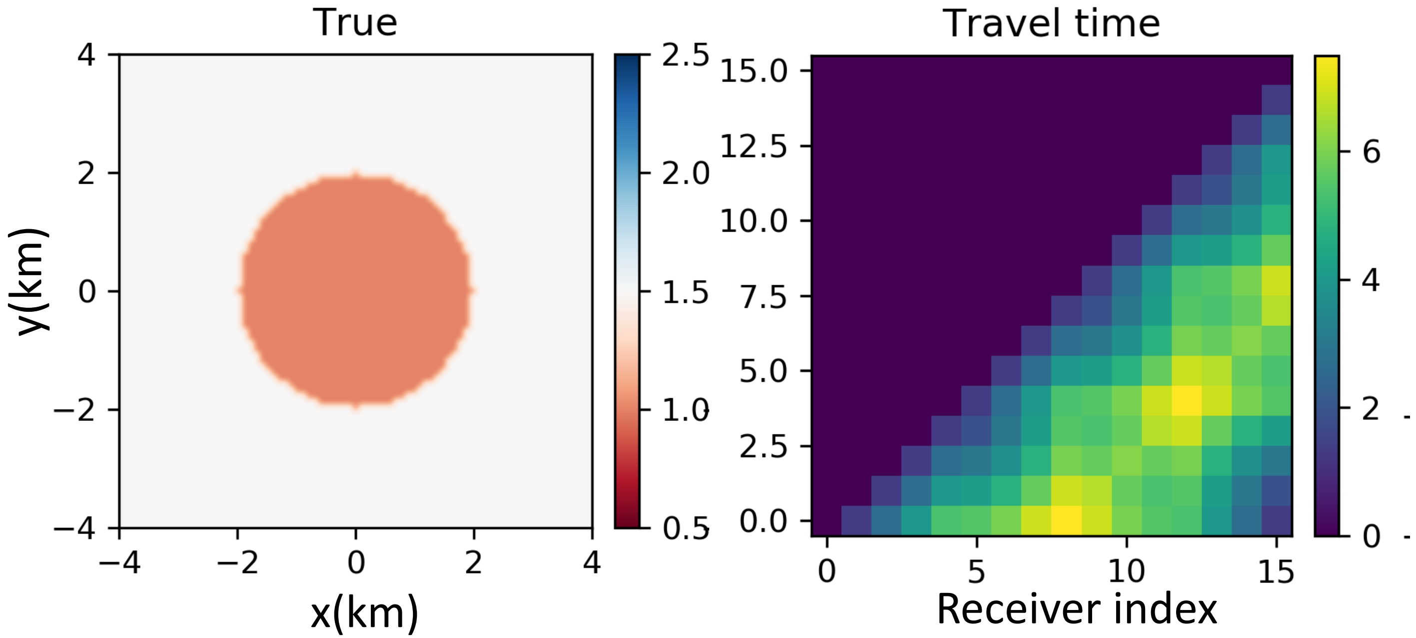


$N(0,1)$

- Uniform prior: 0.5 – 2.5 km/s
- 100,000 training data
- fully connected subnetwork (s1, t1, s2, t2)

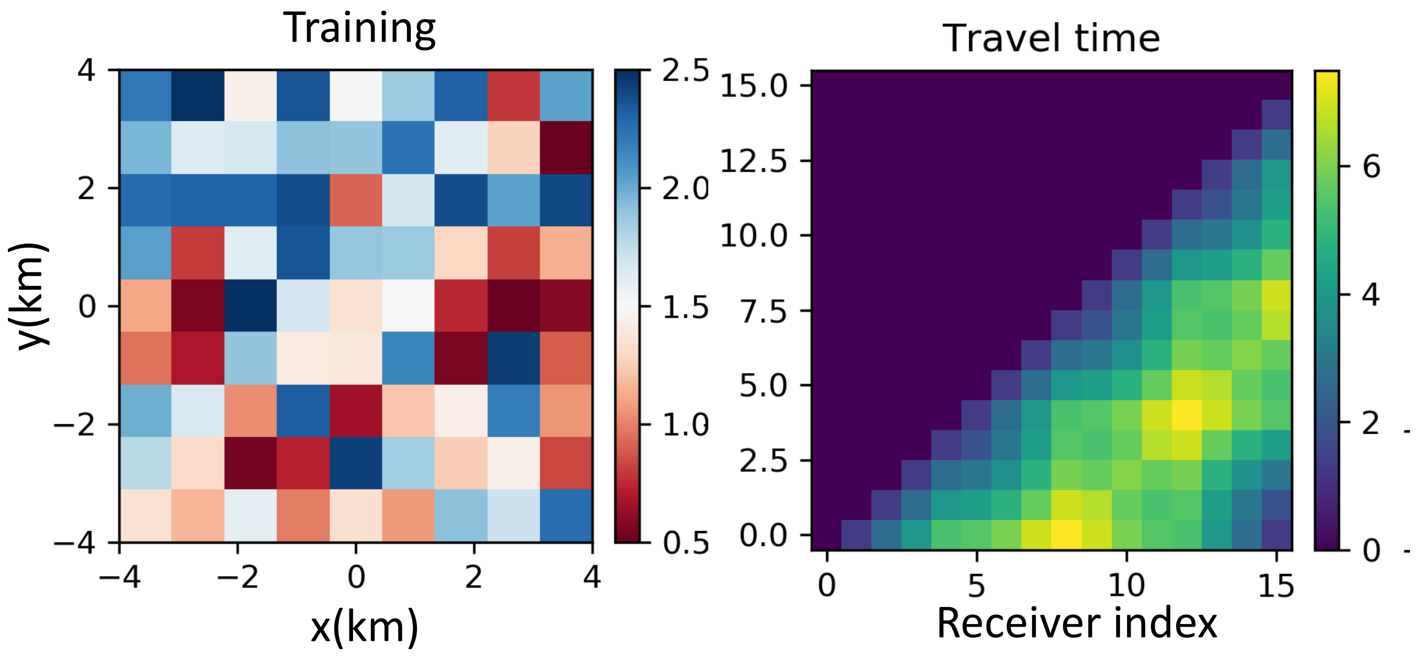
Zhang & Curtis (2021 – J. Geophys. Res.)

Examples



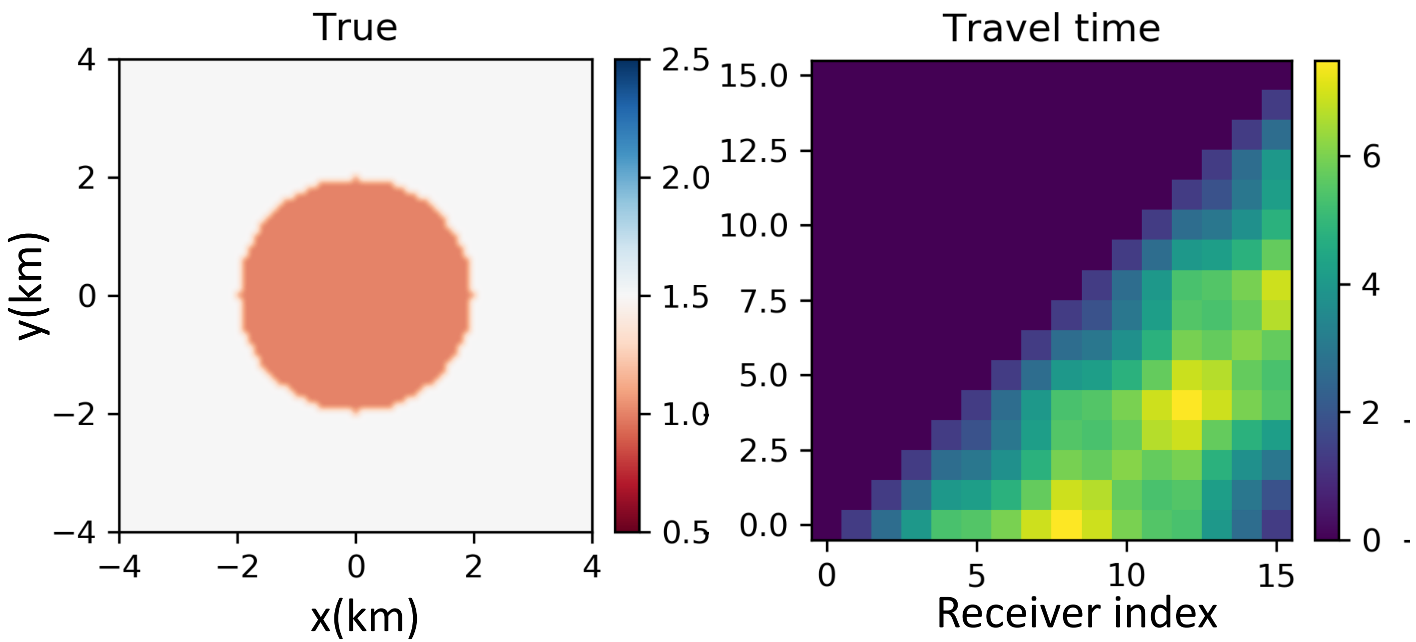
Zhang & Curtis (2021 – J. Geophys. Res.)

Examples



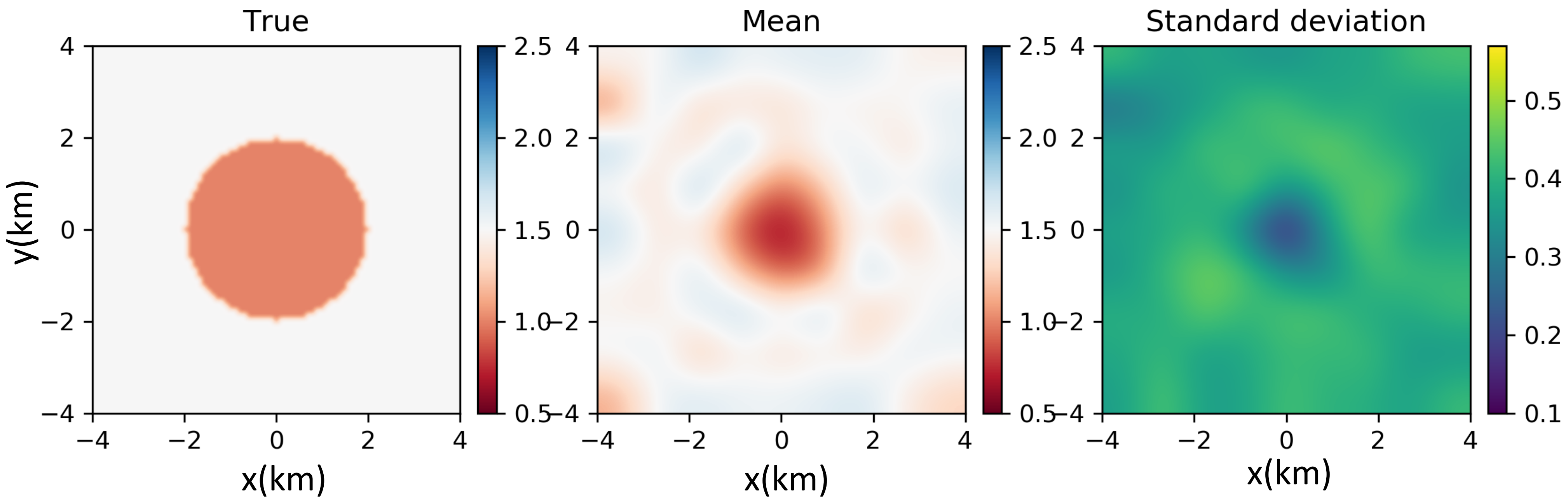
Zhang & Curtis (2021 – J. Geophys. Res.)

Examples



Zhang & Curtis (2021 – J. Geophys. Res.)

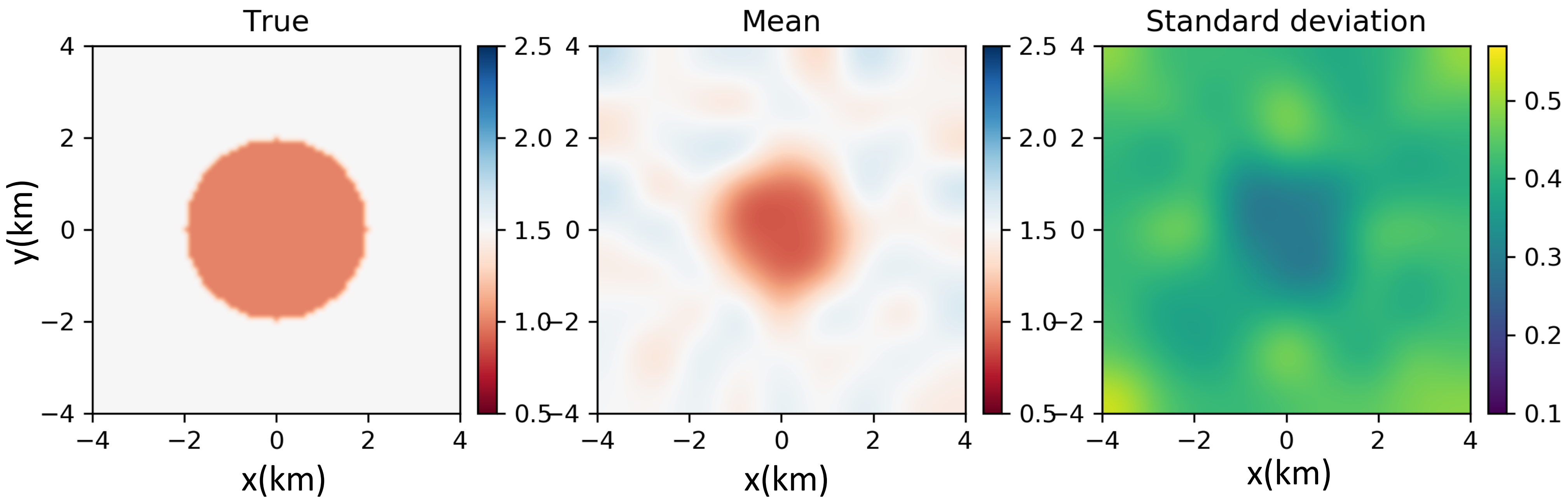
Examples - INN



2 seconds

Zhang & Curtis (2021 – J. Geophys. Res.)

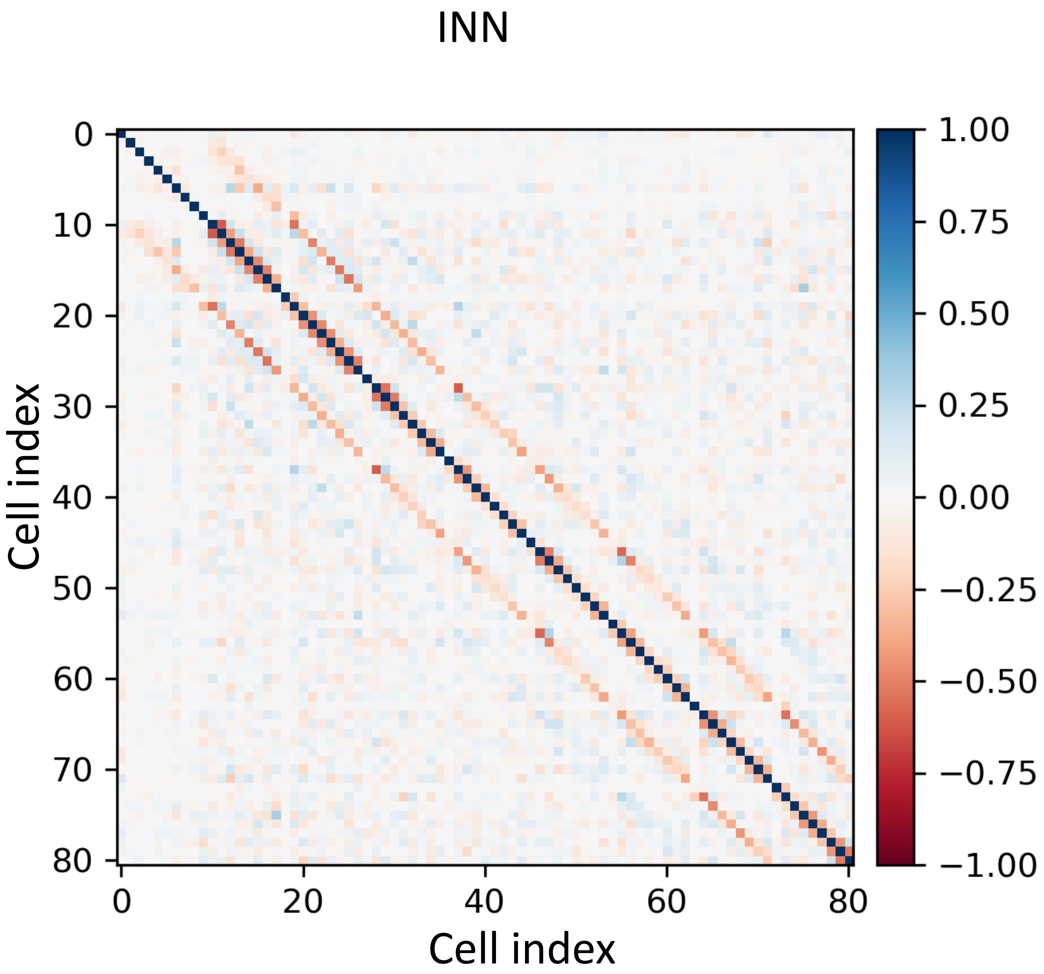
Examples - McMC

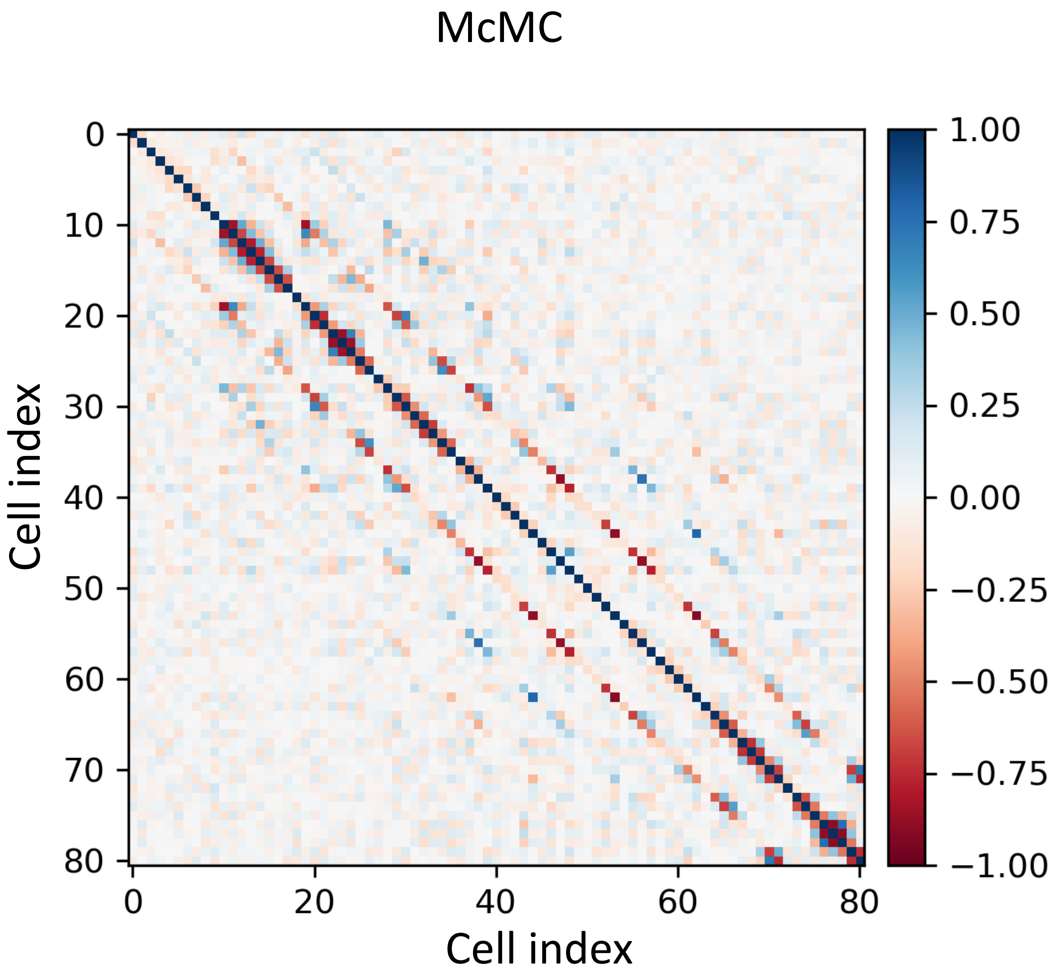


3 days

Zhang & Curtis (2021 – J. Geophys. Res.)

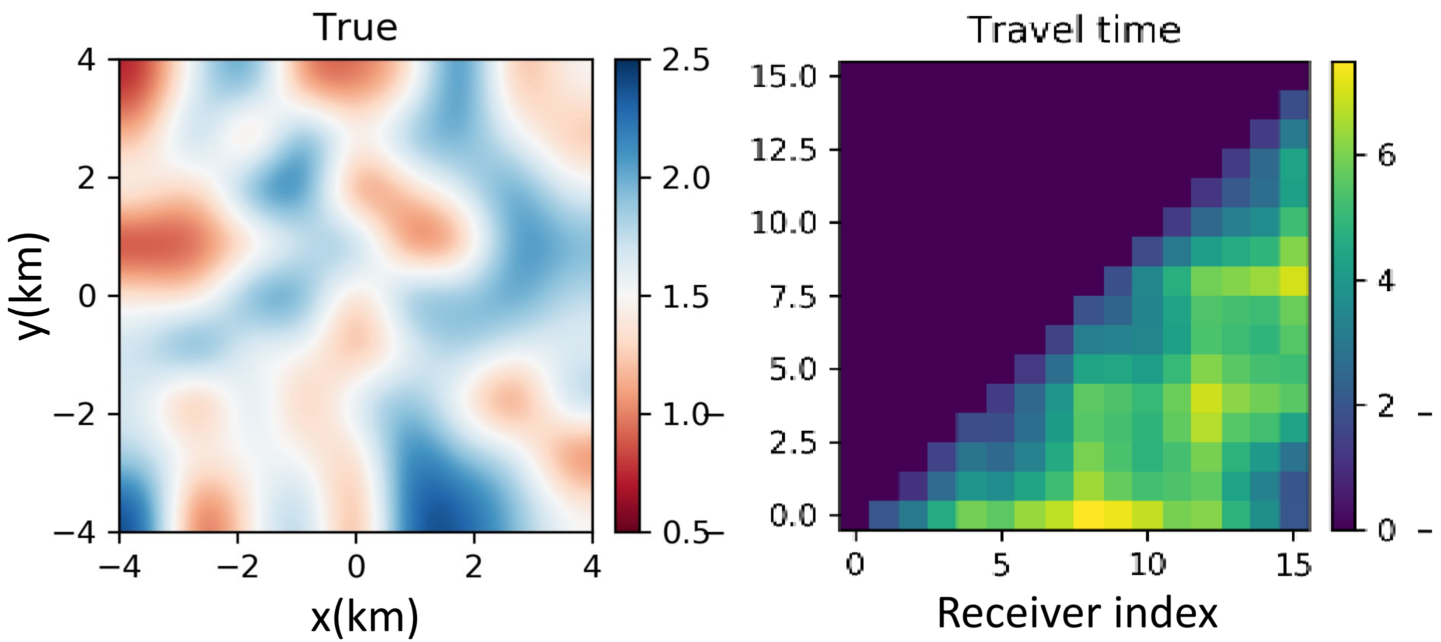
Covariance matrix





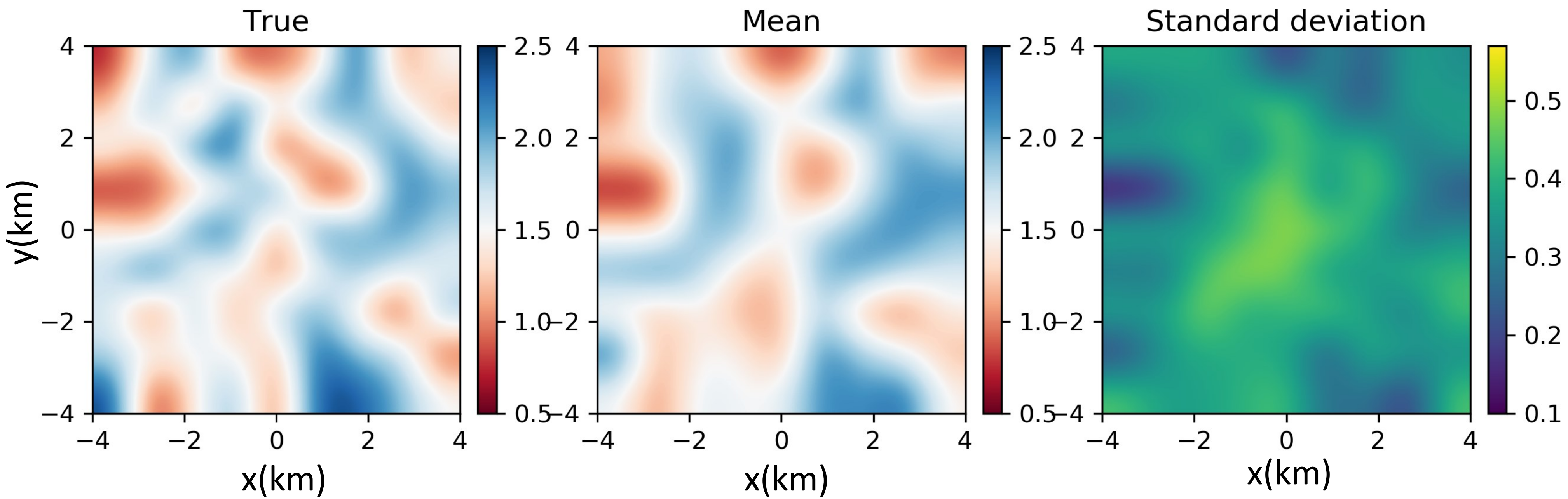
Zhang & Curtis (2021 – J. Geophys. Res.)

Examples



Zhang & Curtis (2021 – J. Geophys. Res.)

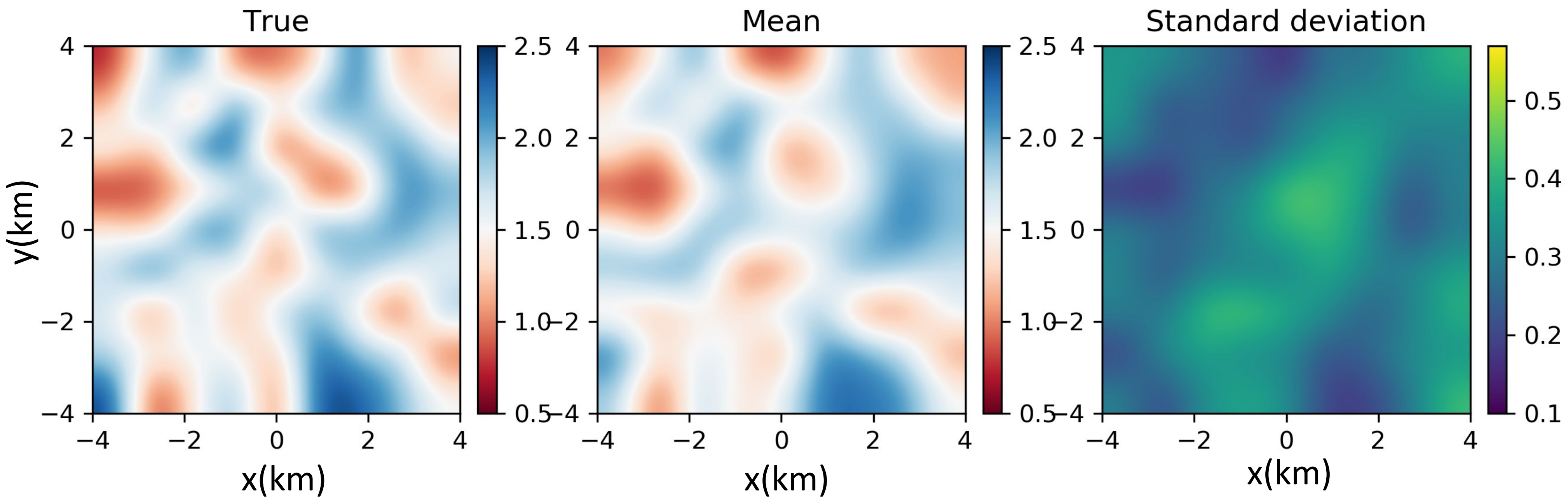
Examples - INN



2 seconds

Zhang & Curtis (2021 – J. Geophys. Res.)

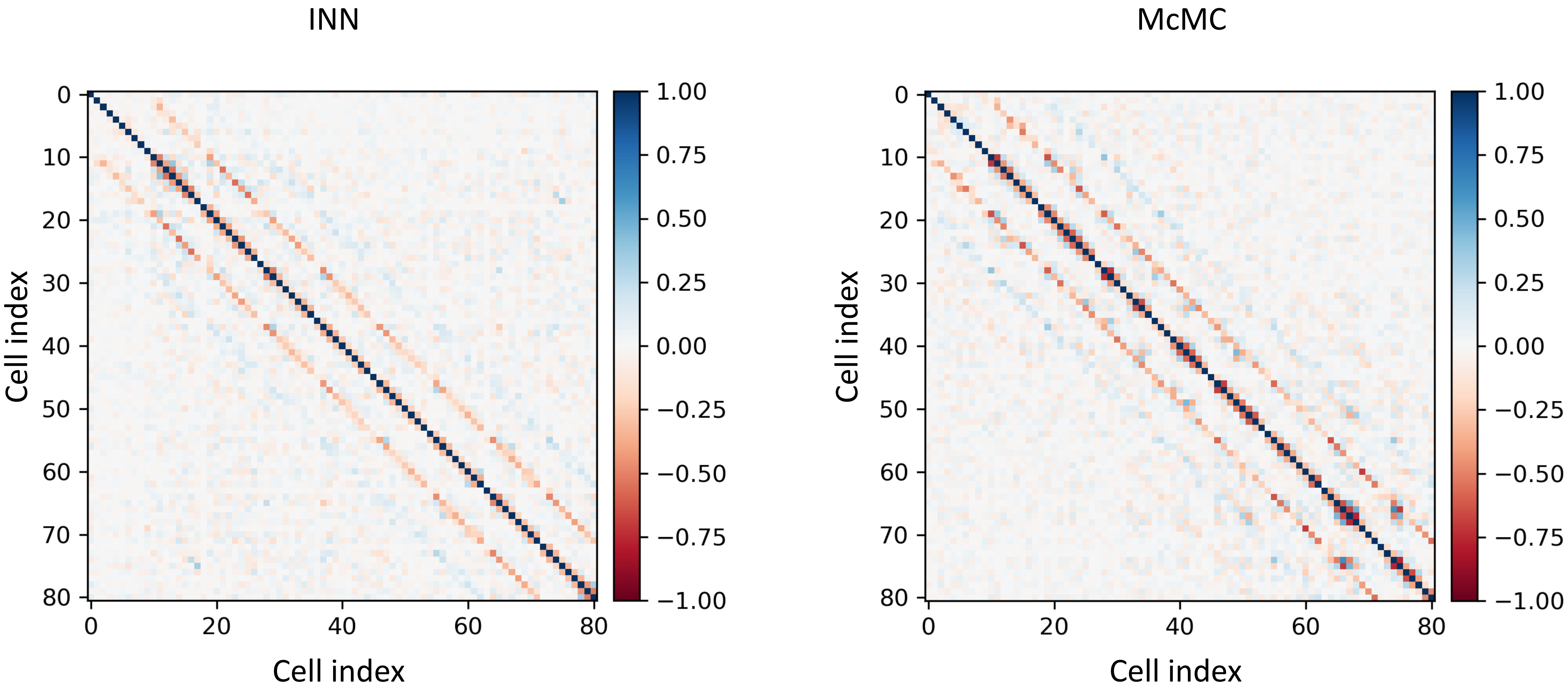
Examples - McMC



3 days

Zhang & Curtis (2021 – J. Geophys. Res.)

Covariance matrix



Zhang & Curtis (2021 – J. Geophys. Res.)

Summary – INN’s

- Invertible Neural Networks + **Max. Likelihood Loss Function** estimate **posterior pdf’s**
    - Zhang & Curtis, 2021. “Geophysical Inversion using Invertible Neural Networks”, *J. Geophys. Res.*

- Solve multiple similar problems **very** rapidly **post-training**

- 2D tomography: **Similar training time to running Monte Carlo once**

- **Both the strength, and the weakness comes from training over the entire prior pdf**

Variational Inference using Normalizing Flows

Rezende & Mohamed (2015 – arXiv)
Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

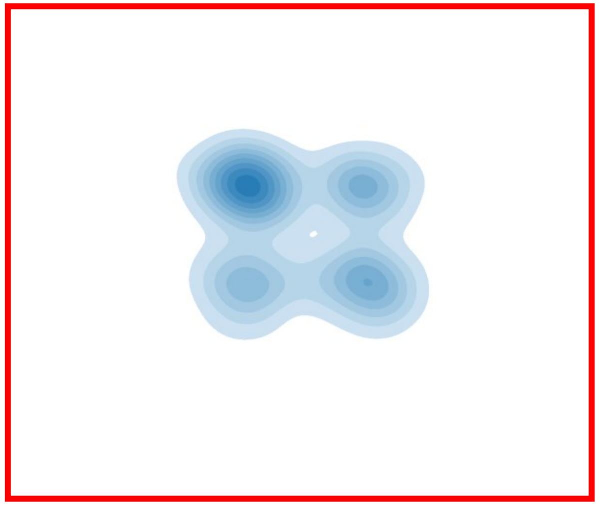
Variational Inference

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

Variational Inference approximates $p(\mathbf{m}|\mathbf{d}_{obs})$ by simpler pdf $q(\mathbf{m})$ in family $Q(\mathbf{m})$:

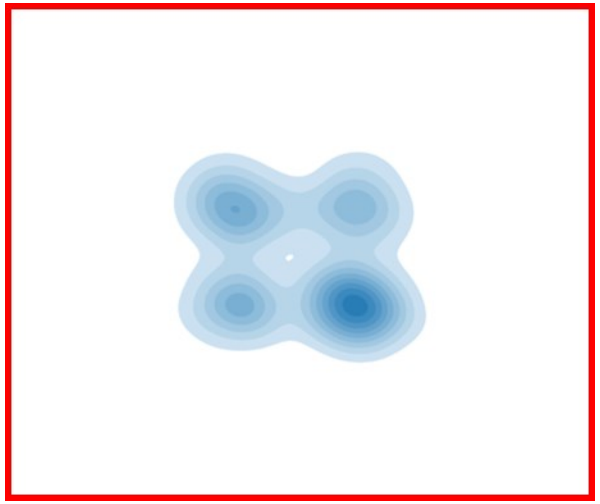
Variational Inference

Posterior pdf $p(\mathbf{m}|\mathbf{d}_{obs})$



$\mathrm{KL}[q||p]$

Variational pdfs $q(\mathbf{m})$ in $Q(\mathbf{m})$



KL (Kullback-Leibler) divergence measures difference between q and p:

Minimize $\mathrm{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{obs})] = \mathrm{E}_q\left[\log\left(q(\mathbf{m})/p(\mathbf{m}|\mathbf{d}_{obs})\right)\right]$

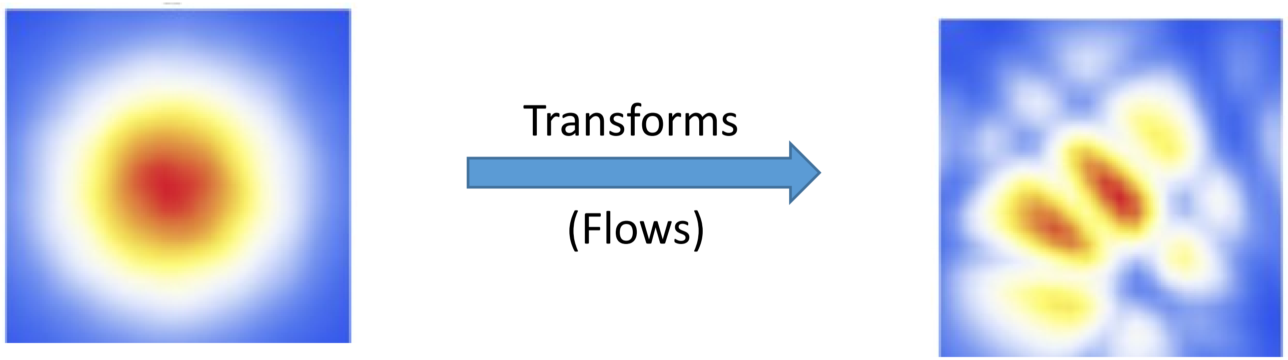
$\Leftrightarrow$ $\boxed{\text{Maximize}}$ $\boxed{\text{ELBO}}(q) = \boxed{\mathrm{E}_q[\log p(\mathbf{m}, \mathbf{d}_{obs})] - \mathrm{E}_q[\log q(\mathbf{m})]}$

Evidence lower bound

Normalizing Flows

Apply K invertible transforms to convert $q_0(\mathbf{m}_0)$ to target pdf $q_K(\mathbf{m}_K)$

Initial pdf $q_0(\mathbf{m}_0)$ Target pdf $q_K(\mathbf{m}_K)$



$$\mathbf{m}_K = F(\mathbf{m}_0) = f_K \circ f_{K-1} \circ \cdots \circ f_2 \circ f_1(\mathbf{m}_0)$$

$$q_K(\mathbf{m}_K) = q_0(\mathbf{m}_0)\boxed{\left|\det\frac{\partial F}{\partial \mathbf{m}_0}\right|^{-1}}$$

Volume change

$$= q_0(\mathbf{m}_0)\prod_{i=1}^{K}\boxed{\left|\det\frac{\partial f_i(\mathbf{m}_{i-1})}{\partial \mathbf{m}_{i-1}}\right|^{-1}}$$

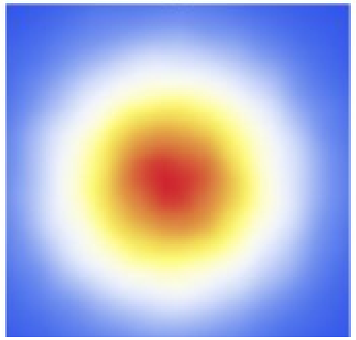
Every flow must be invertible

Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

Normalizing Flows

Apply K invertible transforms to convert $q_0(\mathbf{m}_0)$ to target pdf $q_K(\mathbf{m}_K)$

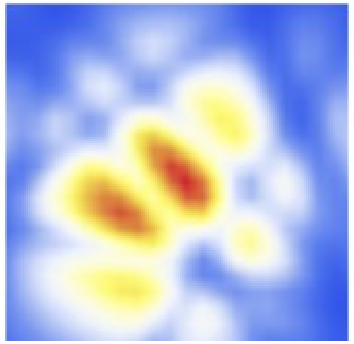
Initial pdf $q_0(\mathbf{m}_0)$



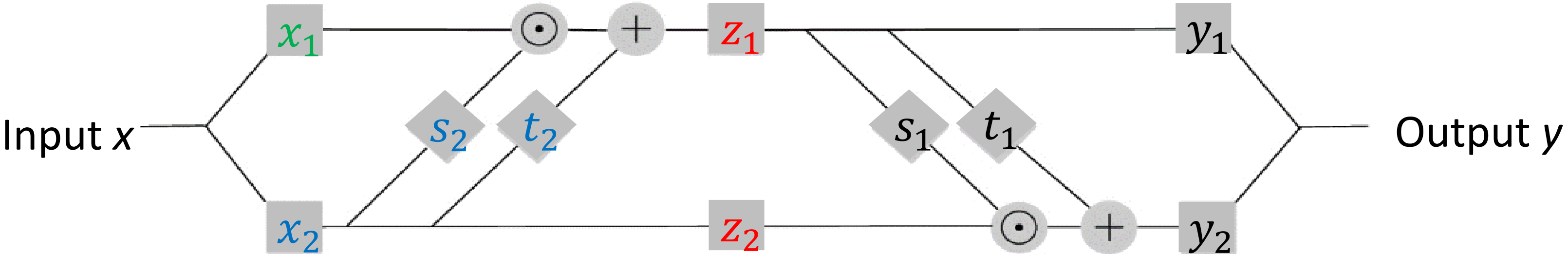
Transforms

(Flows)

Target pdf $q_K(\mathbf{m}_K)$



$$\mathbf{m}_K = F(\mathbf{m}_0) = f_K \circ f_{K-1} \circ \cdots \circ f_2 \circ f_1(\mathbf{m}_0)$$



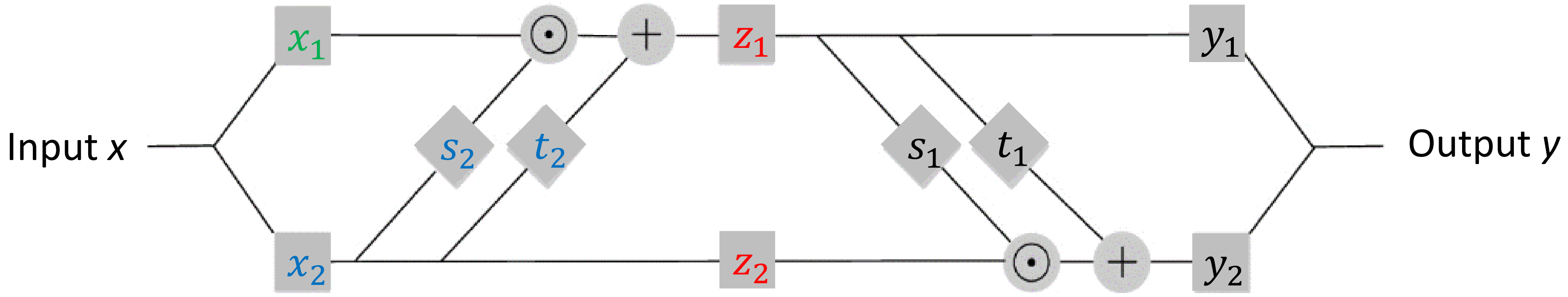
Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

INN versus Flows:

INN

Learn relationship between <u>all</u> of the data and model spaces

Loss function: min. $\|y - nn(x)\| + MMD[p(y, y'), nn(p(x))]$ or use *Max. Prior Likelihood form*



Flows

Find an approximation to the posterior pdf for <u>one</u> specific d_{obs}

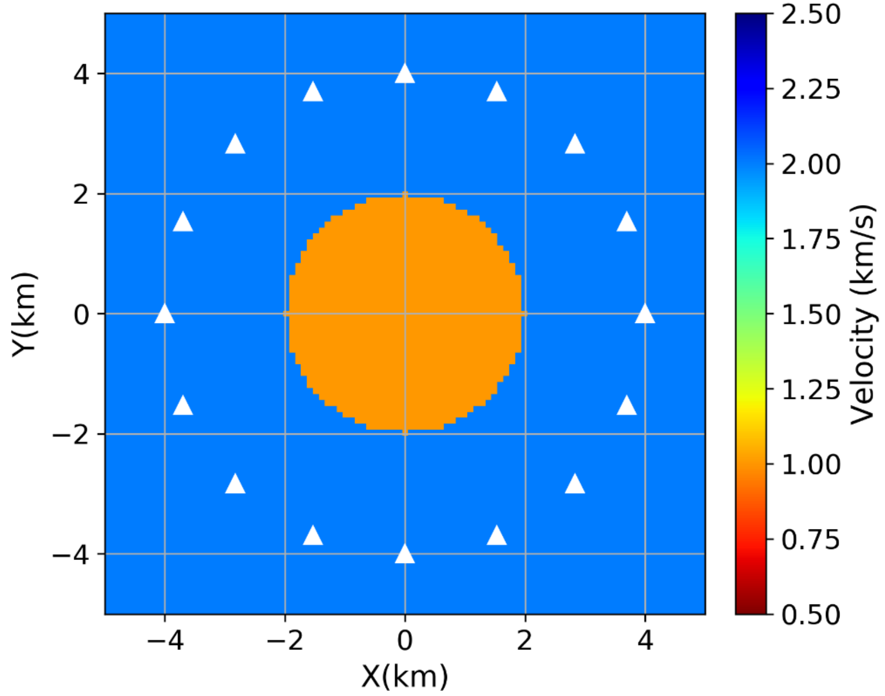
Loss function: min. $\mathrm{KL}(q(\mathbf{m})||p(\mathbf{m}|\mathbf{d_{obs}}))$

Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

Synthetic Test of travel time tomography

Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

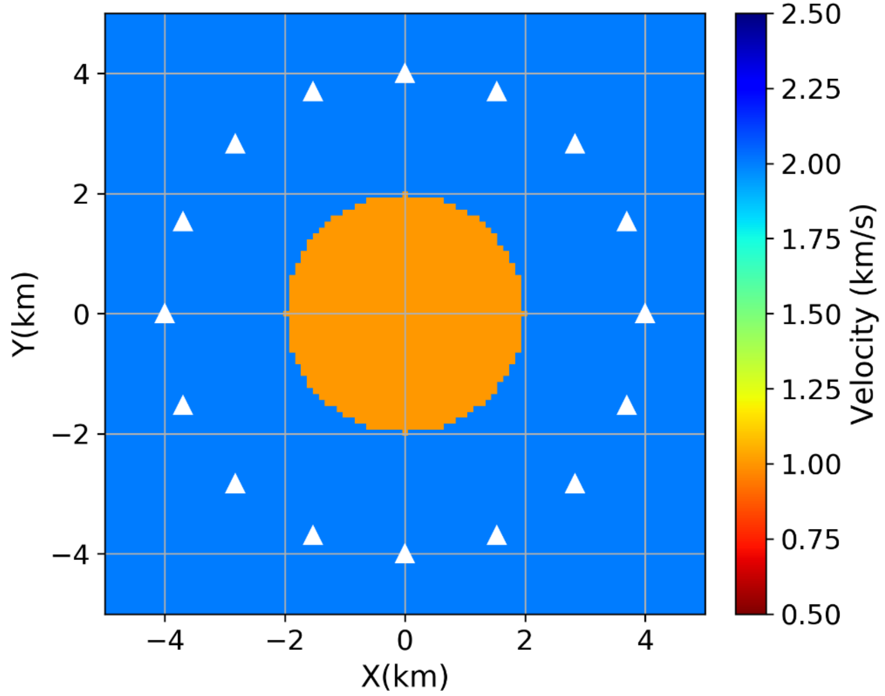
Travel Time Tomography



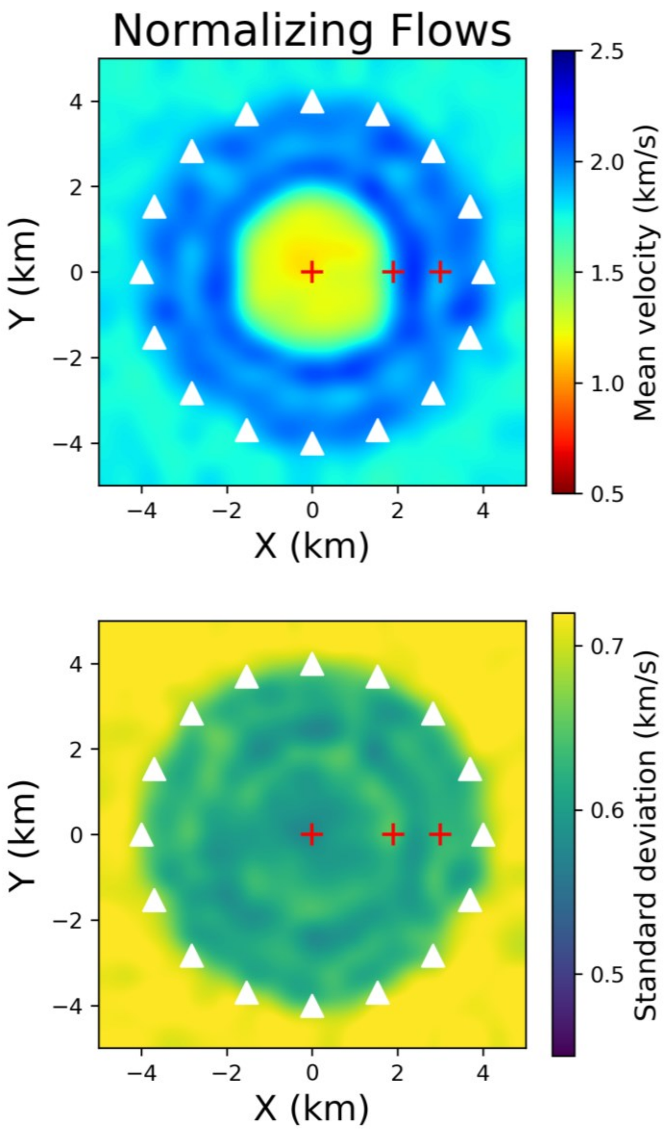
True Velocity Model

Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

Travel Time Tomography

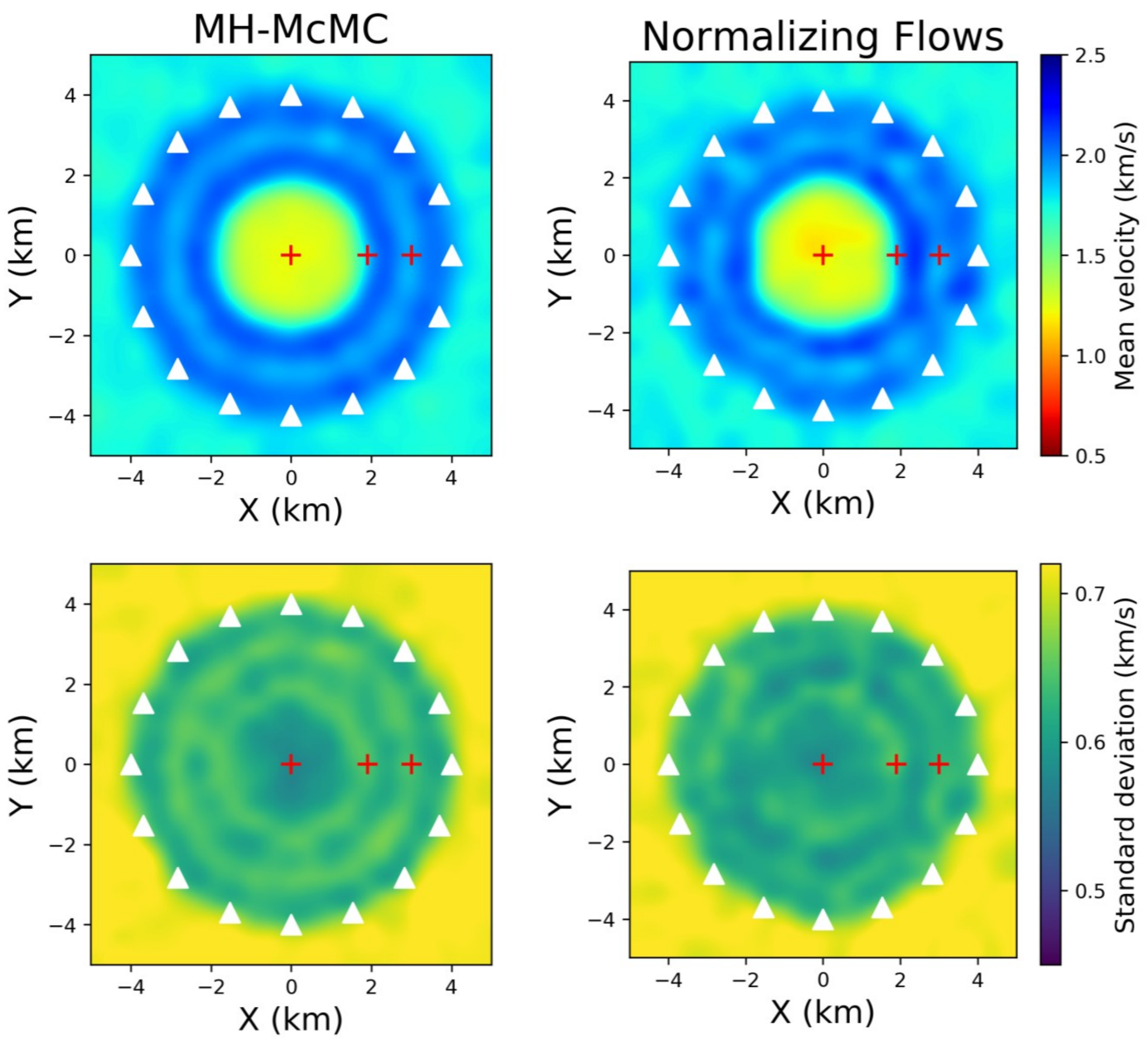


True Velocity Model



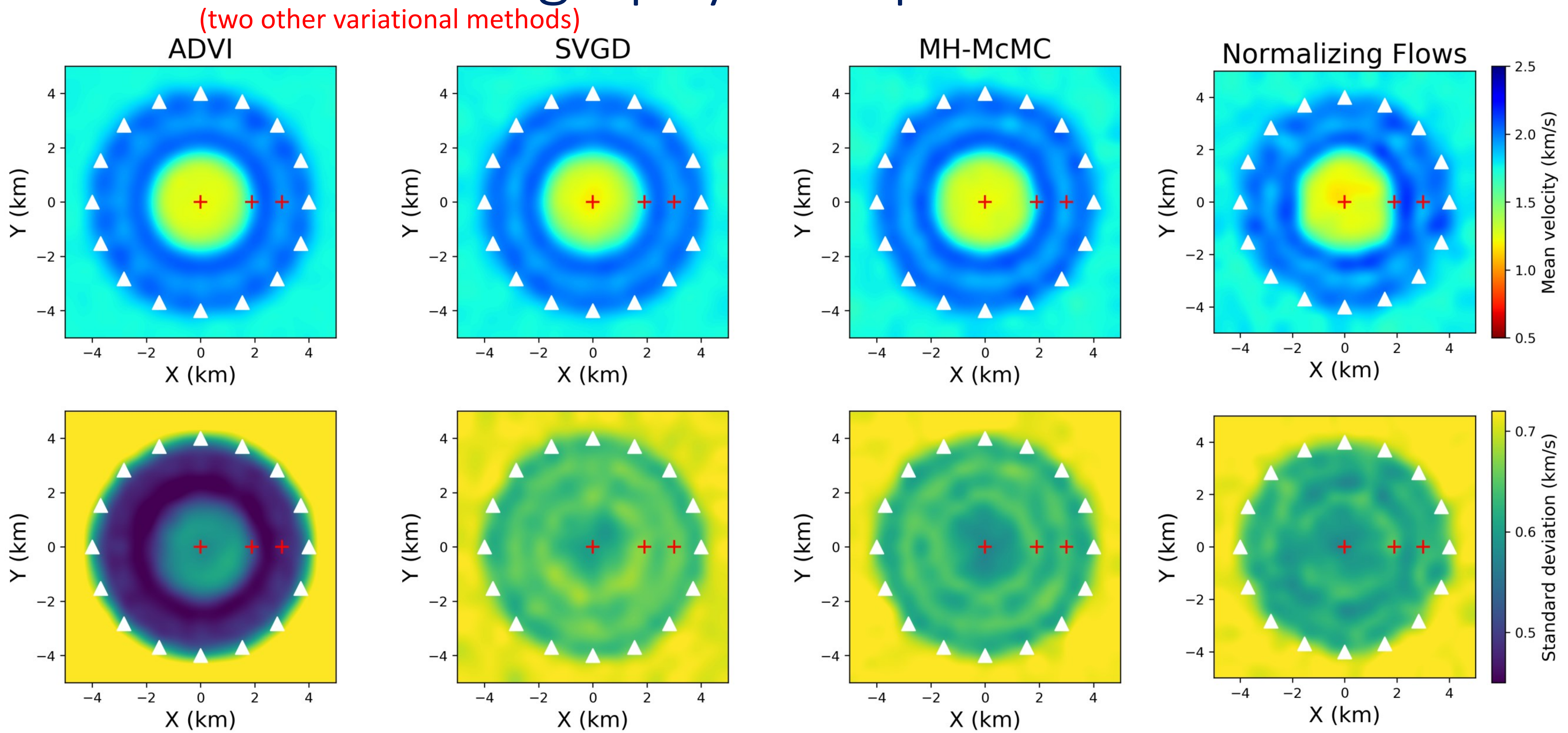
Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

Travel Time Tomography: comparison



Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

Travel Time Tomography: comparison



Results of ADVI, SVGD and McMC from Zhang & Curtis (2020a,b)

Computational cost

Methods	Number of simulations
ADVI	10,000
Normalizing Flows	30,000
SVGD	400,000
Rj-McMC	3,000,000
MH-McMC	12,000,000

Cheapest,
But Incorrect Uncertainty Results

Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

Computational cost

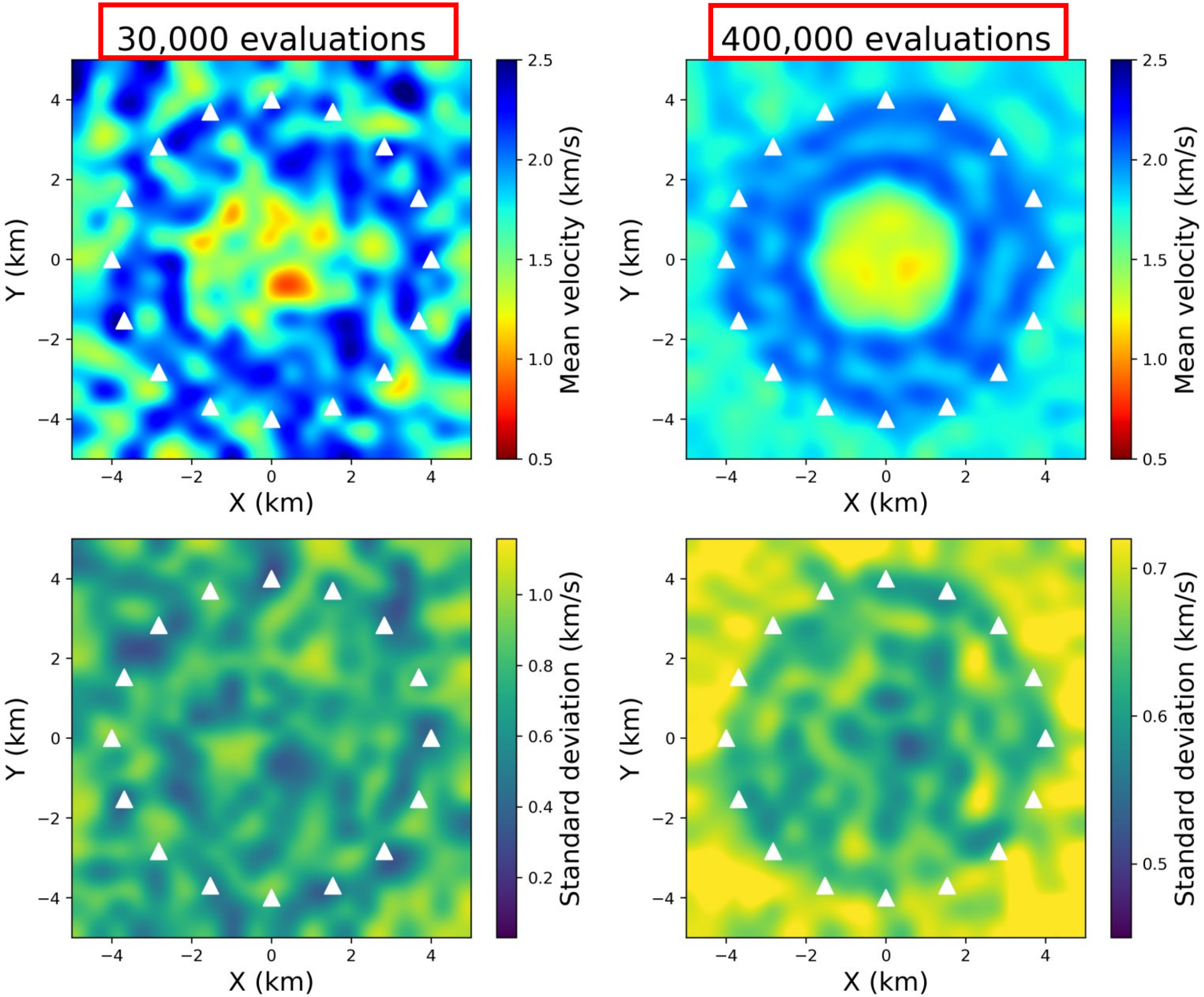
Methods	Number of simulations
ADVI	10,000
Normalizing Flows	30,000
SVGD	400,000
Rj-McMC	3,000,000
MH-McMC	12,000,000

Cheapest,
But Incorrect Uncertainty Results

Detecting convergence of McMC is subjective

Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

Travel Time Tomography: MH-McMC



Computational cost

Methods	Number of simulations
ADVI	10,000
Normalizing Flows	30,000
SVGD	400,000
Rj-McMC	3,000,000
MH-McMC	12,000,000

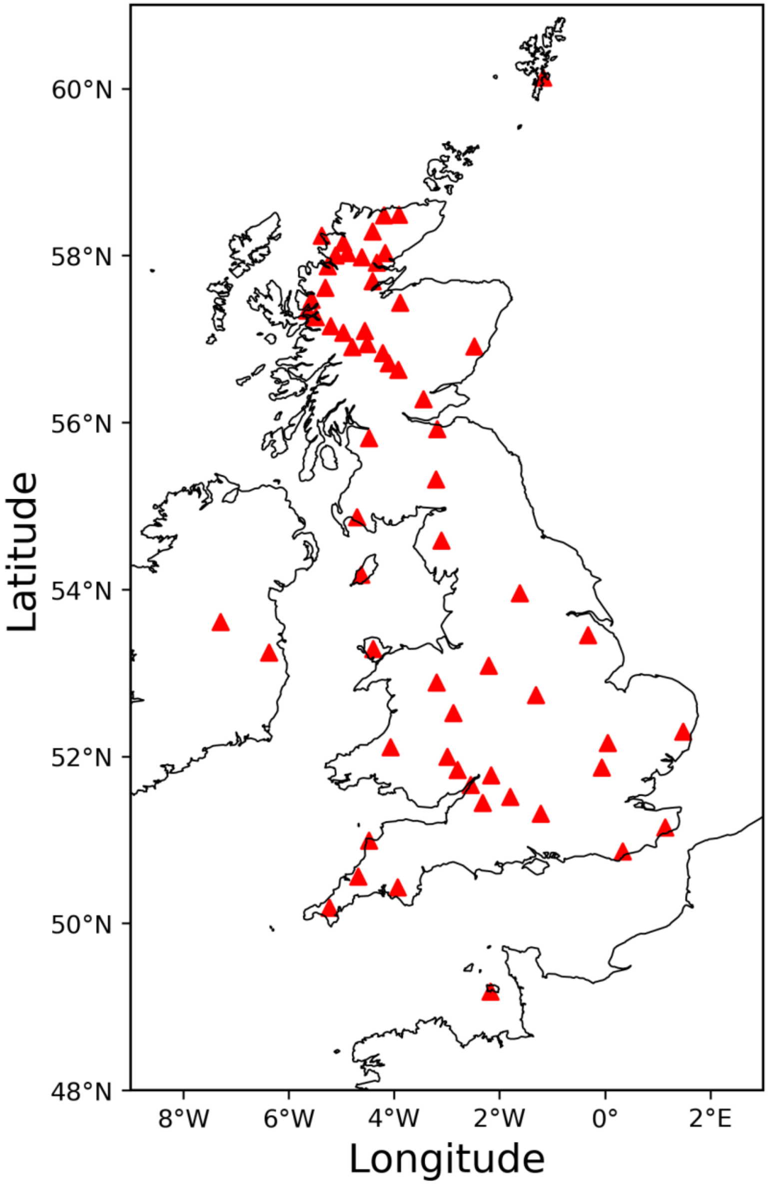
Cheapest,
But Incorrect Uncertainty Results!

Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

Love Wave Tomography of the British Isles

Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

Receiver stations

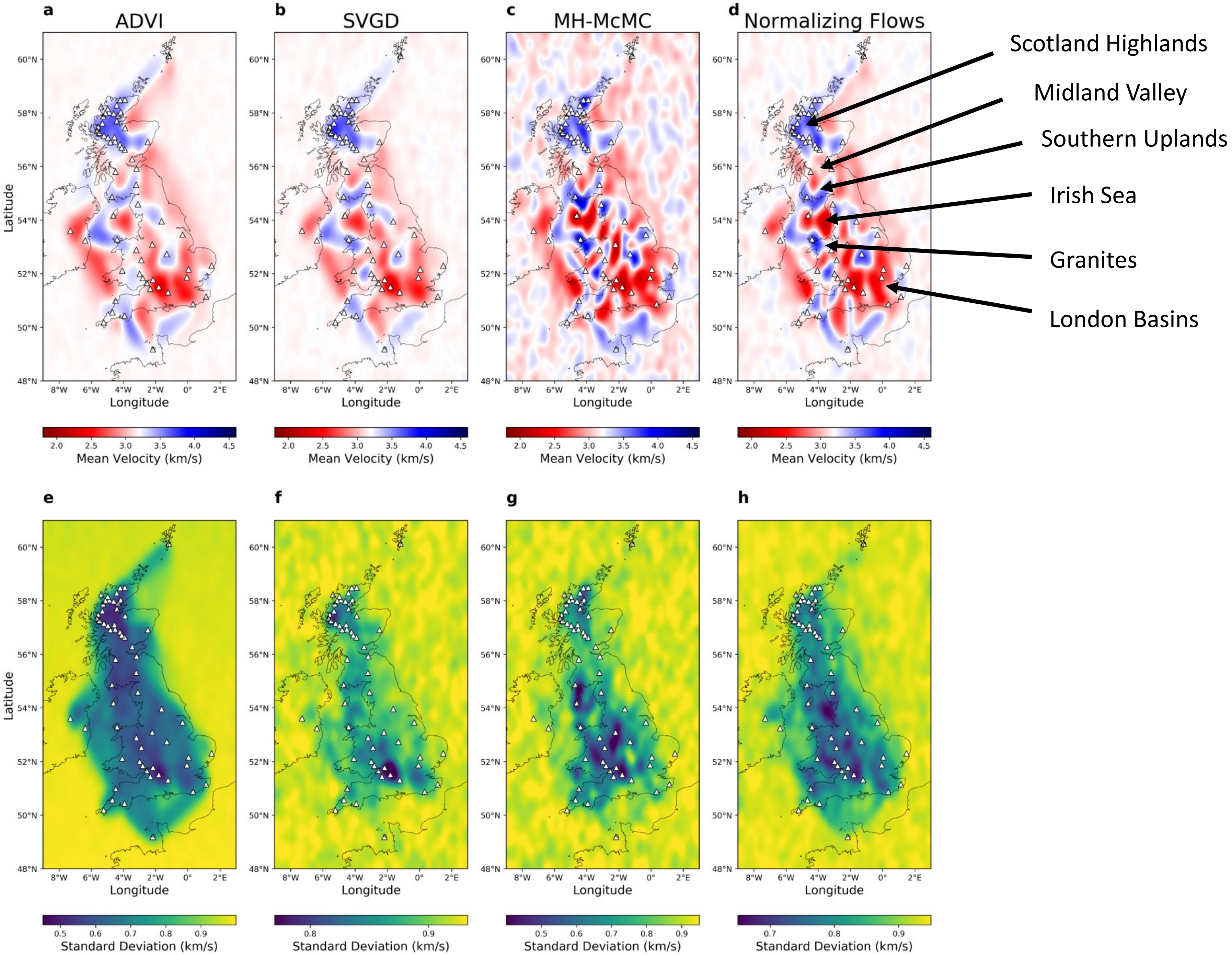


Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

Results

Different resolution in mean models

Uncertainty



Computational cost

Methods	Forward Evaluations	Elapsed Time (Hours)	Parallelization (Cores)
ADVI	10,000	6.95	1
Normalizing Flows	100,000	7.83	10
SVGD	600,000	31.71	10
RJ-McMC	48,000,000	~720	16
MH-McMC	30,000,000	660	20

From Galetti et al., 2017 (same dataset)

Zhao, Curtis, Zhang (2021 – Geophys. J. Int.)

Conclusions

1. ***Normalizing flows***: chain of invertible transforms → one approximate posterior pdf
2. One particular form of flow is an ***Invertible Neural Network (INN)***
3. INN's provide uncertainties including correlations for **all** (~small) problems

Try it! – Code package:

VIP: Variational Inversion Package (Zhang & Curtis, 2024: *Seismica*)

All papers are available: https://blogs.ed.ac.uk/curtis/publications Andrew.Curtis@ed.ac.uk

Zhang & Curtis, 2020: *"Seismic tomography using variational inference methods"*, J. Geophys. Res.

Zhang & Curtis, 2024: *"VIP – Variational Inversion Package with example implementations of Bayesian tomographic imaging"*, Seismica

Zhang et al., 2023: *"3D Bayesian Variational Full Waveform Inversion"*, Geophys. J. Int.

Zhao & Curtis, 2024a: *"Physically Structured Variational Inference for Bayesian Full Waveform Inversion"*, J. Geophys. Res.

Zhao & Curtis, 2024b: *"Variational prior replacement in Bayesian inference and inversion"*, Geophys. J. Int

Zhao & Curtis, 2024c: *"Efficient Bayesian Full Waveform Inversion and Analysis of Prior Hypotheses in 3D"*, arXiv

Conclusions

1. Reasonable 3D Bayesian FWI results → **One extra order of computation c.f. linearised FWI**

2. Discriminate ~10 different prior hypotheses → **Same order of computation as linearised FWI**

Try it! – Code package:

VIP: Variational Inversion Package (Zhang & Curtis, 2024: *Seismica*)

All papers are available: https://blogs.ed.ac.uk/curtis/publications Andrew.Curtis@ed.ac.uk

Zhang & Curtis, 2020: *"Seismic tomography using variational inference methods"*, J. Geophys. Res.

Zhang & Curtis, 2024: *"VIP – Variational Inversion Package with example implementations of Bayesian tomographic imaging"*, Seismica

Zhang et al., 2023: *"3D Bayesian Variational Full Waveform Inversion"*, Geophys. J. Int.

Zhao & Curtis, 2024a: *"Physically Structured Variational Inference for Bayesian Full Waveform Inversion"*, J. Geophys. Res.

Zhao & Curtis, 2024b: *"Variational prior replacement in Bayesian inference and inversion"*, Geophys. J. Int

Zhao & Curtis, 2024c: *"Efficient Bayesian Full Waveform Inversion and Analysis of Prior Hypotheses in 3D"*, arXiv