Normalizing Flows for Bayesian Experimental Design in Imaging Applications

Rafael Orozco[1], Abhinav Gahlot[1], Peng Chen[1] , Mathias Louboutin[1*] and Felix J. Herrmann[1]

1 Georgia Tech College of Computing
School of Computational Science and Engineering

* now at Devito Codes

Seismic Laboratory for Imaging and Modeling (SLIM)
Georgia Institute of Technology

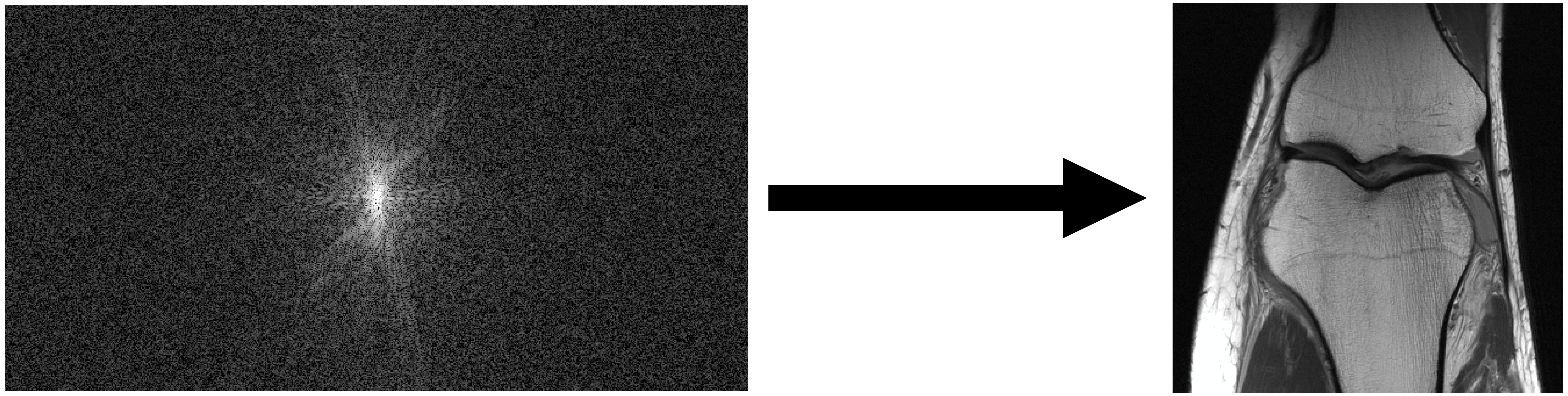
Presentation in one sentence:

Gradient optimization of designs w.r.t a normalizing flow loss enables experimental design in problems with…

Presentation in one sentence:

Gradient optimization of designs w.r.t a normalizing flow loss enables experimental design in problems with…

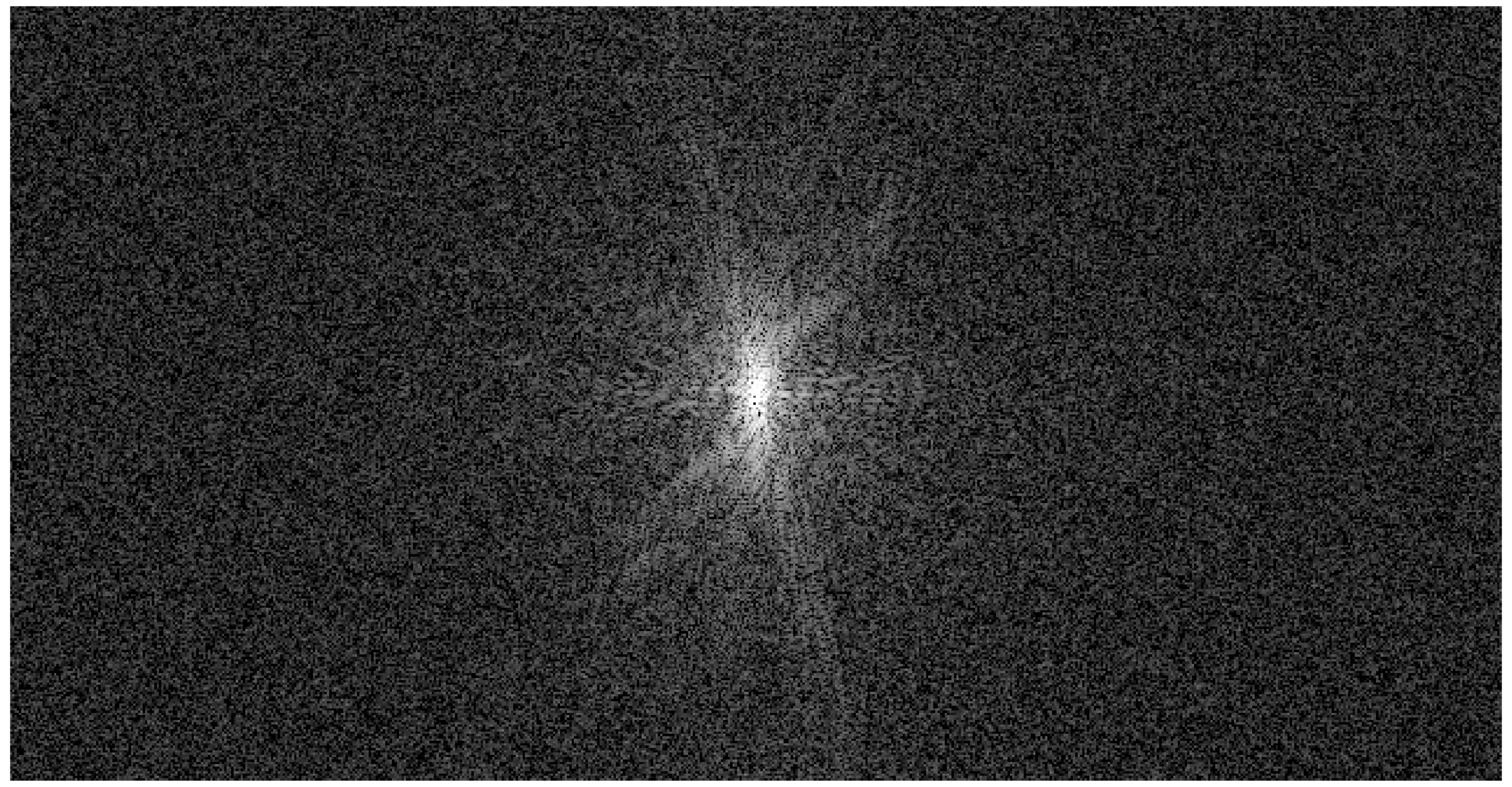
- large parameter designs (200,000 for medical imaging)



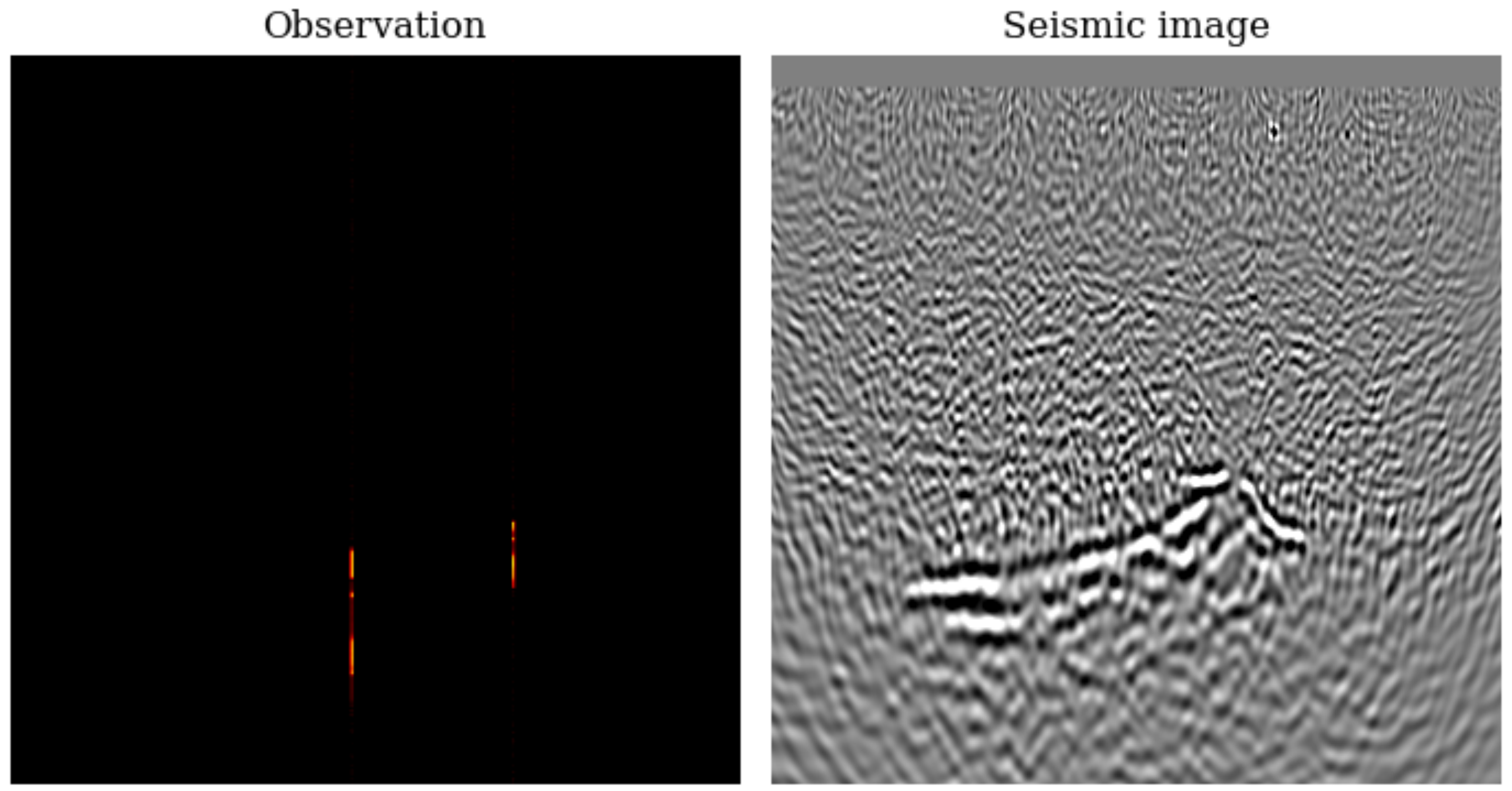
Presentation in one sentence:

Gradient optimization of designs w.r.t a normalizing flow loss enables experimental design in problems with…

- large parameter designs (200,000 for medical imaging)



- non-linear, expensive forward operators (wave equation for $CO2$ monitoring).



Takeaways from presentation

1. Exact likelihood evaluation keeps normalizing flows relevant in this diffusion era.

SORA WHO????

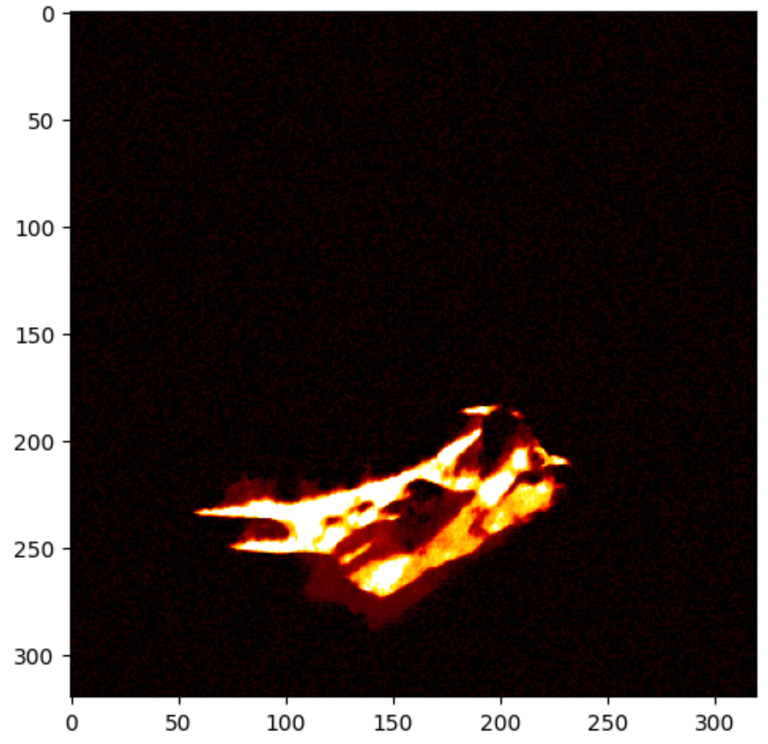
Takeaways from presentation

1. Exact likelihood evaluation keeps normalizing flows relevant in this diffusion era.

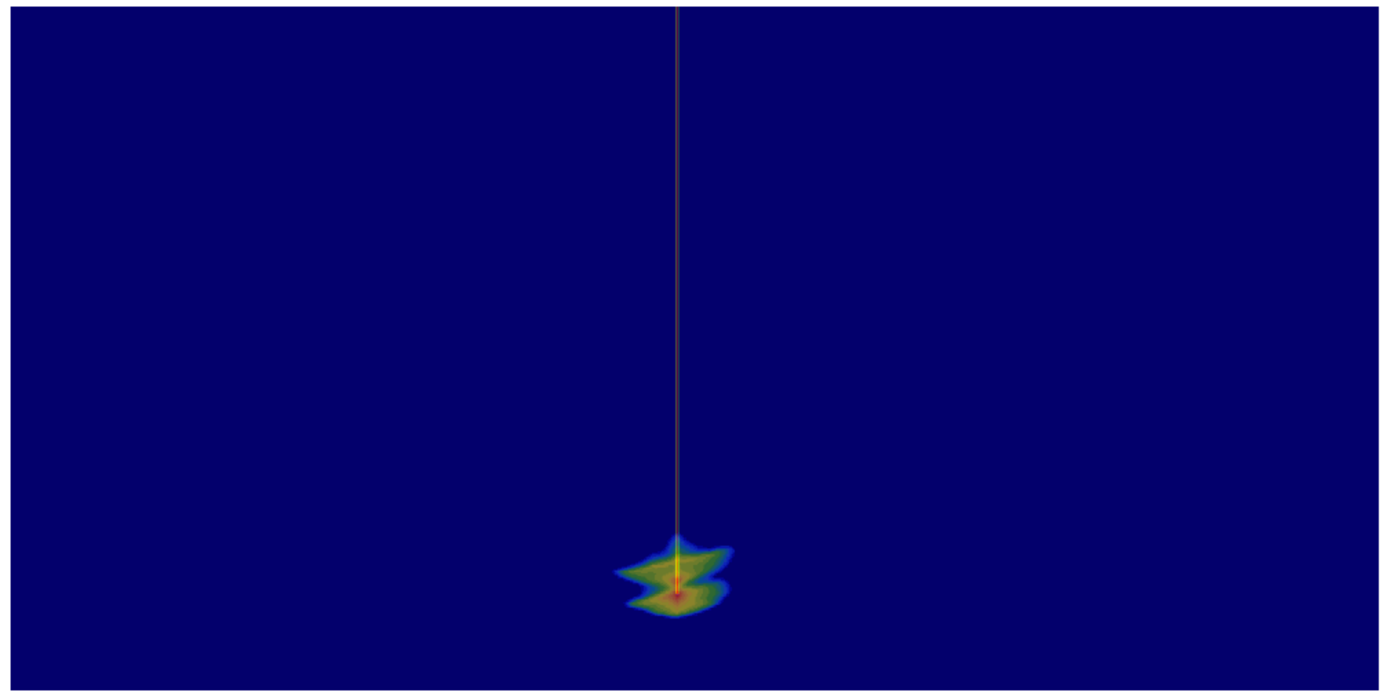
SORA WHO????

2. Simulation based inference is a general framework for Bayesian inference **and downstream tasks i.e. experimental design.**

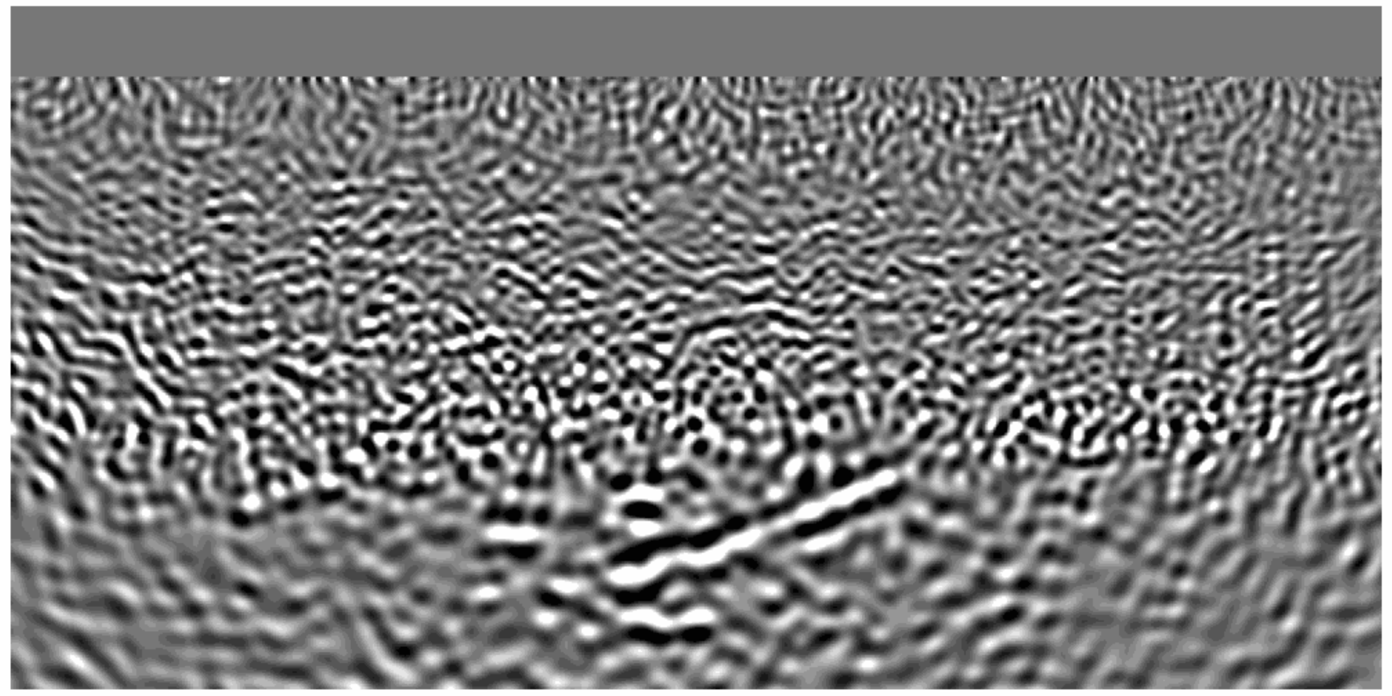
prior samples



fluid simulator



simulated observations



Bayesian experimental design

How should we collect data $\mathbf{y}$ over observable $\mathbf{u}$ to inform inference?

$$\mathbf{y} = \mathbf{M}(\mathbf{u})$$

Go, Jinwoo, and Tobin Isaac. "Robust expected information gain for optimal Bayesian experimental design using ambiguity sets." *Uncertainty in Artificial Intelligence*. PMLR, 2022.

Bayesian experimental design

How should we collect data $\mathbf{y}$ over observable $\mathbf{u}$ to inform inference?

$$\mathbf{y} = \mathbf{M}(\mathbf{u})$$

Bayesians have a powerful answer: "Collect the data that maximizes the information gained" - where information gain is quantified by Kullback-Leibler divergence:

$$\max_{\mathbf{M}} D_{KL}(p(\mathbf{x}\,|\,\mathbf{y})\,||\,p(\mathbf{x}))$$

Go, Jinwoo, and Tobin Isaac. "Robust expected information gain for optimal Bayesian experimental design using ambiguity sets." *Uncertainty in Artificial Intelligence*. PMLR, 2022.

Bayesian experimental design

How should we collect data $\mathbf{y}$ over observable $\mathbf{u}$ to inform inference?

$$\mathbf{y} = \mathbf{M}(\mathbf{u})$$

Bayesians have a powerful answer: “Collect the data that maximizes the information gained” - where information gain is quantified by Kullback-Leibler divergence:

$$\max_{\mathbf{M}} D_{KL}(p(\mathbf{x}\,|\,\mathbf{y})\,||\,p(\mathbf{x}))$$

Expected information gain (EIG) *averages* over *all* possible $\mathbf{y}$

$$\max_{\mathbf{M}} EIG(\mathbf{M}) = \mathbb{E}_{p(\mathbf{y}|\mathbf{M})}\left[D_{KL}(p(\mathbf{x}\,|\,\mathbf{y})\,||\,p(\mathbf{x}))\right]$$

Go, Jinwoo, and Tobin Isaac. "Robust expected information gain for optimal Bayesian experimental design using ambiguity sets." *Uncertainty in Artificial Intelligence*. PMLR, 2022.

Relation between EIG and posterior likelihood

SLIM

Maximizing the expected information gain is equivalent to maximizing the expected posterior likelihood

$$\max_{\mathbf{M}} \quad EIG(\mathbf{M}) = \mathbb{E}_{p(\mathbf{y}|\mathbf{M})} \left[D_{KL}(p_\theta(\mathbf{x}\,|\,\mathbf{y})\,||\,p(\mathbf{x})) \right]$$

$$= \mathbb{E}_{p(\mathbf{y}|\mathbf{M})} \left[\mathbb{E}_{p(\mathbf{x}|\mathbf{y})} \left[\log p_\theta(\mathbf{x}\,|\,\mathbf{y}) - \log p(\mathbf{x}) \right] \right]$$

$$= \mathbb{E}_{p(\mathbf{y}|\mathbf{M})} \left[\mathbb{E}_{p(\mathbf{x}|\mathbf{y})} \left[\log p_\theta(\mathbf{x}\,|\,\mathbf{y}) \right] \right]$$

$$= \mathbb{E}_{p(\mathbf{x},\mathbf{y}|\mathbf{M})} \left[\log p_\theta(\mathbf{x}\,|\,\mathbf{y}) \right]$$

same as neural posterior objective!

Foster, Adam, et al. "A unified stochastic gradient approach to designing bayesian-optimal experiments." PMLR, 2020.
Hoffmann, Till, and Jukka-Pekka Onnela. "Minimizing the Expected Posterior Entropy Yields Optimal Summary Statistics." arXiv preprint arXiv:2206.02340 (2022).

Relation between EIG and posterior likelihood

Maximizing the expected information gain is equivalent to maximizing the expected posterior likelihood

$$\max_{\mathbf{M}} \; EIG(\mathbf{M}) = \mathbb{E}_{p(\mathbf{y}|\mathbf{M})} \left[D_{KL}(p_\theta(\mathbf{x}\,|\,\mathbf{y}) \,||\, p(\mathbf{x})) \right]$$

$$= \mathbb{E}_{p(\mathbf{y}|\mathbf{M})} \left[\mathbb{E}_{p(\mathbf{x}|\mathbf{y})} \left[\log p_\theta(\mathbf{x}\,|\,\mathbf{y}) - \log p(\mathbf{x}) \right] \right]$$

$$= \mathbb{E}_{p(\mathbf{y}|\mathbf{M})} \left[\mathbb{E}_{p(\mathbf{x}|\mathbf{y})} \left[\log p_\theta(\mathbf{x}\,|\,\mathbf{y}) \right] \right]$$

$$= \mathbb{E}_{p(\mathbf{x},\mathbf{y}|\mathbf{M})} \left[\log p_\theta(\mathbf{x}\,|\,\mathbf{y}) \right]$$

same as neural posterior objective!

Thus *optimizing* $\mathbf{M}$ under posterior learning objective will *increase* its EIG

Foster, Adam, et al. "A unified stochastic gradient approach to designing bayesian-optimal experiments." PMLR, 2020.
Hoffmann, Till, and Jukka-Pekka Onnela. "Minimizing the Expected Posterior Entropy Yields Optimal Summary Statistics." arXiv preprint arXiv:2206.02340 (2022).

Normalizing flows for posteriors

They learn to sample posterior by maximizing the posterior likelihood under training examples

$$\max_{\theta} \mathbb{E}_{p(\mathbf{x},\mathbf{y})} \left[\log p_{\theta}(\mathbf{x} \mid \mathbf{y}) \right]$$

Radev, Stefan T., et al. "BayesFlow: Learning complex stochastic models with invertible neural networks." *IEEE transactions on neural networks and learning systems* 33.4 (2020):

Normalizing flows for posteriors

They learn to sample posterior by maximizing the posterior likelihood under training examples

$$\max_{\theta} \; \mathbb{E}_{p(\mathbf{x},\mathbf{y})} \left[\log p_{\theta}(\mathbf{x} \mid \mathbf{y}) \right]$$

e.g. Normalizing flows are trained as such

$$\hat{\theta} \; = \arg\max_{\theta} \; \frac{1}{N} \sum_{n=1}^{N} \left(-\|f_{\theta}(\mathbf{x}^{(n)}; \mathbf{y}^{(n)})\|_2^2 + \log \left| \det \mathbf{J}_{f_{\theta}} \right| \right)$$

Radev, Stefan T., et al. "BayesFlow: Learning complex stochastic models with invertible neural networks." *IEEE transactions on neural networks and learning systems* 33.4 (2020):

Proposed method

Prepare posterior learning algorithm as typically:

- use prior samples and forward operator to make training pairs $\{\mathbf{x}^{(n)}, \mathbf{y}^{(n)}\}_{i=1}^{N}$

Instead of optimizing only network parameters:

$$\hat{\theta} \; = \arg\max_{\theta} \frac{1}{N} \sum_{n=1}^{N} \left(-\|f_\theta(\mathbf{x}^{(n)}; \mathbf{y}^{(n)})\|_2^2 + \log \left| \det \mathbf{J}_{f_\theta} \right| \right).$$

Proposed method

Prepare posterior learning algorithm as typically:

- use prior samples and forward operator to make training pairs $\{\mathbf{x}^{(n)}, \mathbf{y}^{(n)}\}_{i=1}^{N}$

Instead of optimizing only network parameters:

$$\hat{\theta} \ = \underset{\theta}{\arg\max} \ \frac{1}{N} \sum_{n=1}^{N} \left(-\| f_\theta(\mathbf{x}^{(n)}; \mathbf{y}^{(n)}) \|_2^2 + \log \left| \det \mathbf{J}_{f_\theta} \right| \right).$$

jointly optimize for design $\mathbf{M}$ as well:

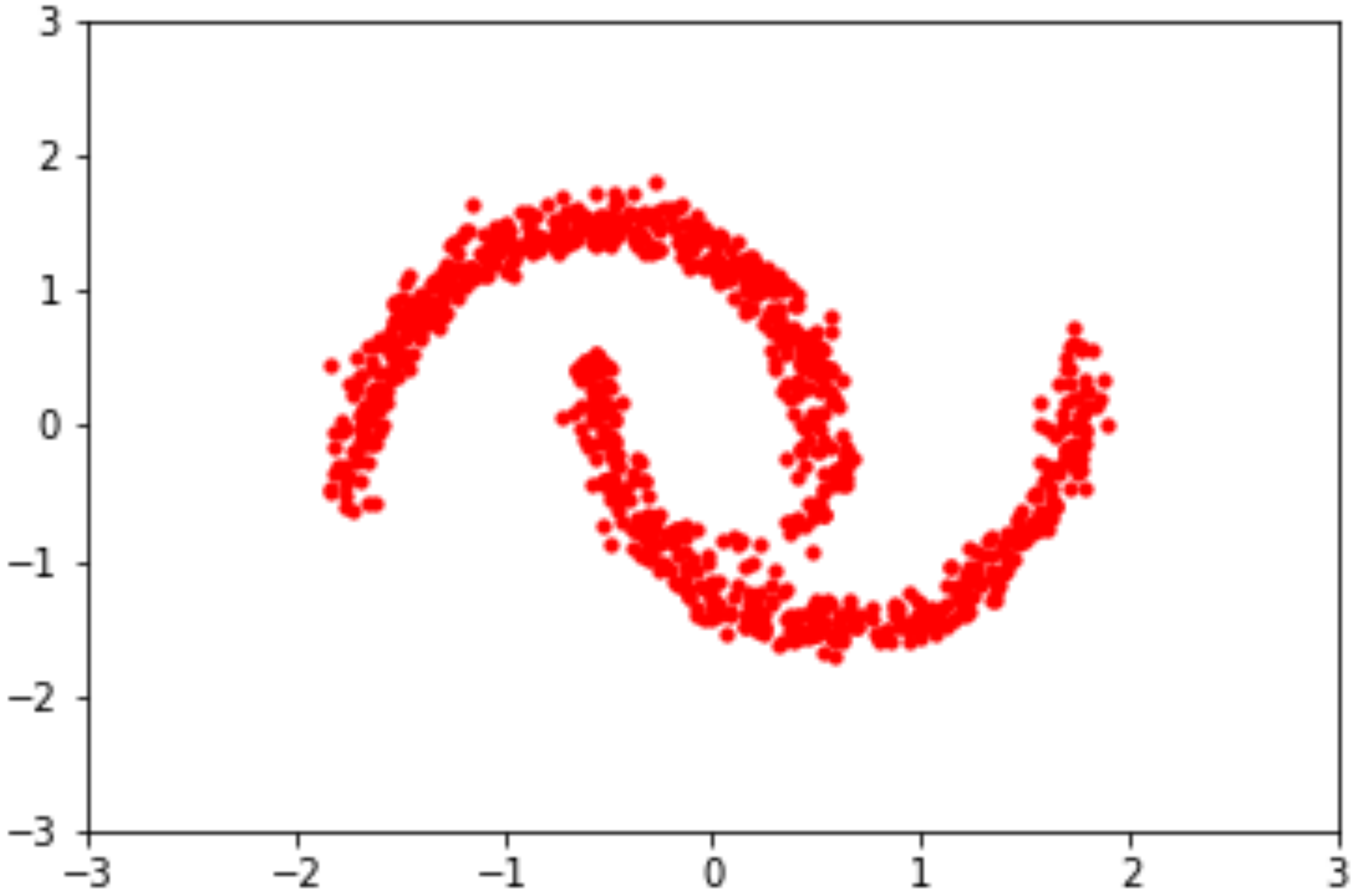
$$\hat{\theta}, \ \hat{\mathbf{M}} = \underset{\theta, \mathbf{M}}{\arg\max} \ \frac{1}{N} \sum_{i=1}^{N} \left(-\| f_\theta(\mathbf{x}^{(n)}; \mathbf{M}(\mathbf{y}^{(n)})) \|_2^2 + \log \left| \det \mathbf{J}_{f_\theta} \right| \right).$$

Normalizing Flows

Normalizing flows

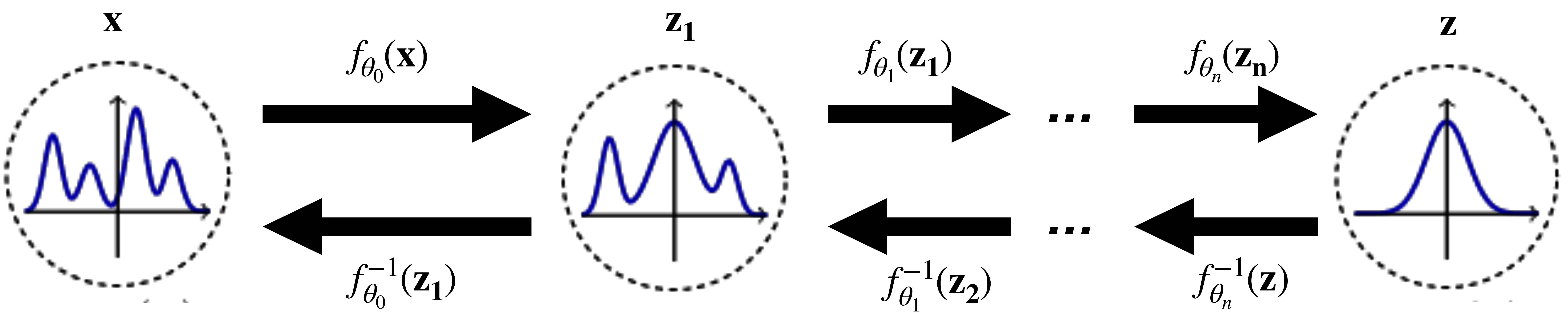
Likelihood-based generative models that:

- have exact likelihood evaluation
- scalable memory usage during training (more on this later)
- fast sampling



Normalizing flows

Learn distribution by mapping samples to simple distribution.



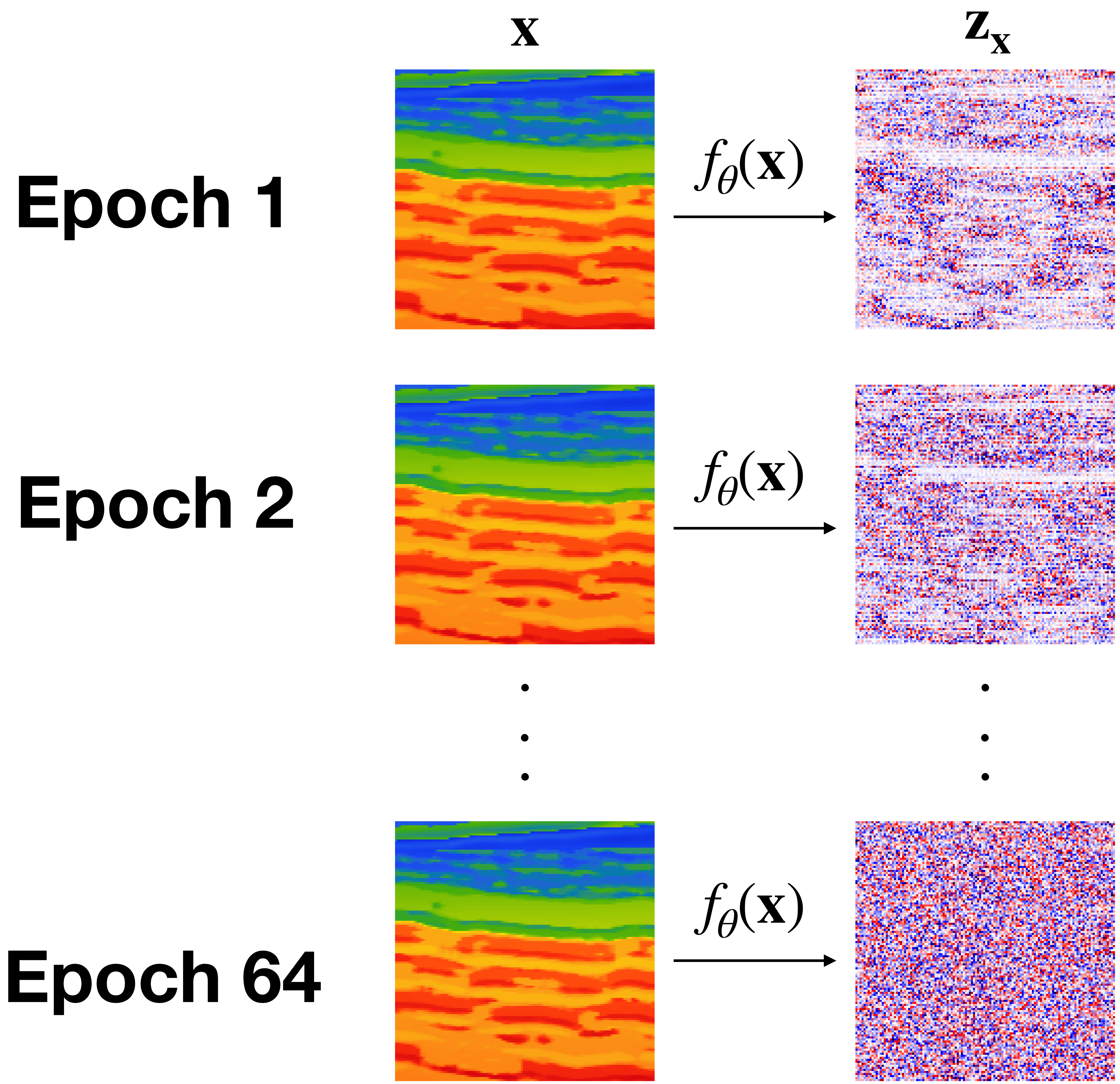
Mapping needs to be

- differentiable
- invertible

Dinh, Laurent, Jascha Sohl-Dickstein, and Samy Bengio. "Density estimation using real nvp." arXiv preprint arXiv:1605.08803 (2016).

Normalizing flow during training

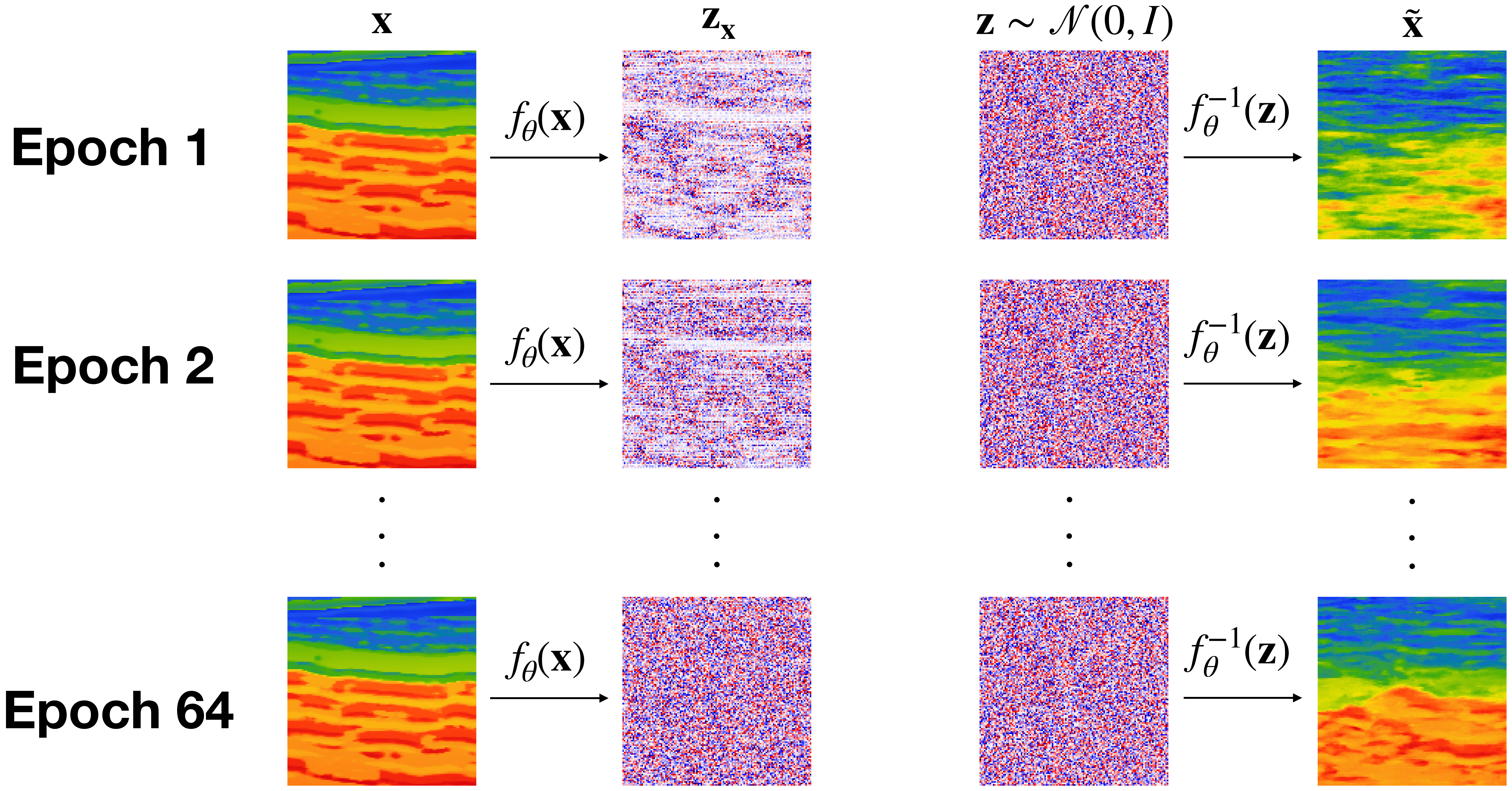
Learn distribution by mapping samples to Normal distribution.



Louboutin, Mathias, et al. "Learned multiphysics inversion with differentiable programming and machine learning." *The Leading Edge* 42.7 (2023): 474-486.

Normalizing flow during training

Learn distribution by mapping samples to Normal distribution.

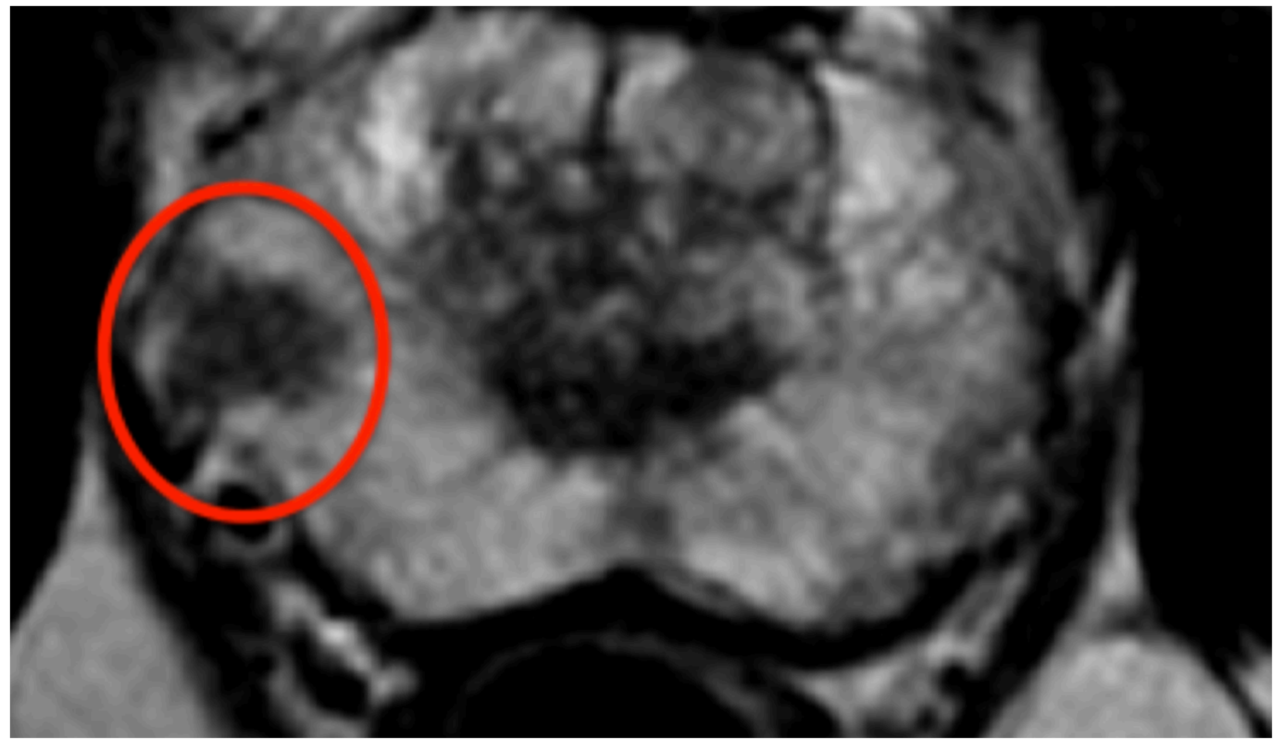


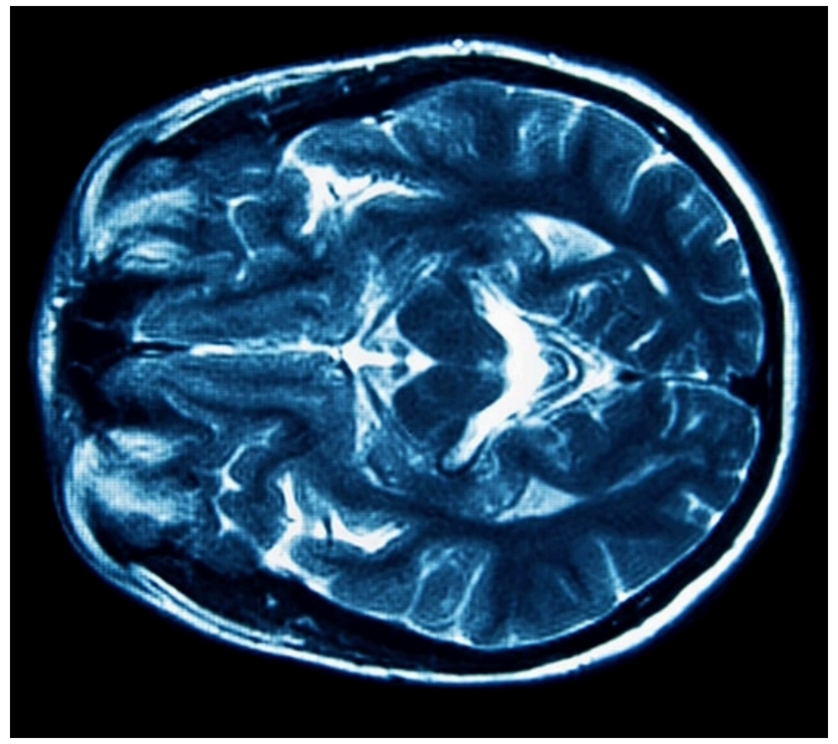
Louboutin, Mathias, et al. "Learned multiphysics inversion with differentiable programming and machine learning." *The Leading Edge* 42.7 (2023): 474-486.

Application: medical imaging

Magnetic Resonance Imaging (MRI)

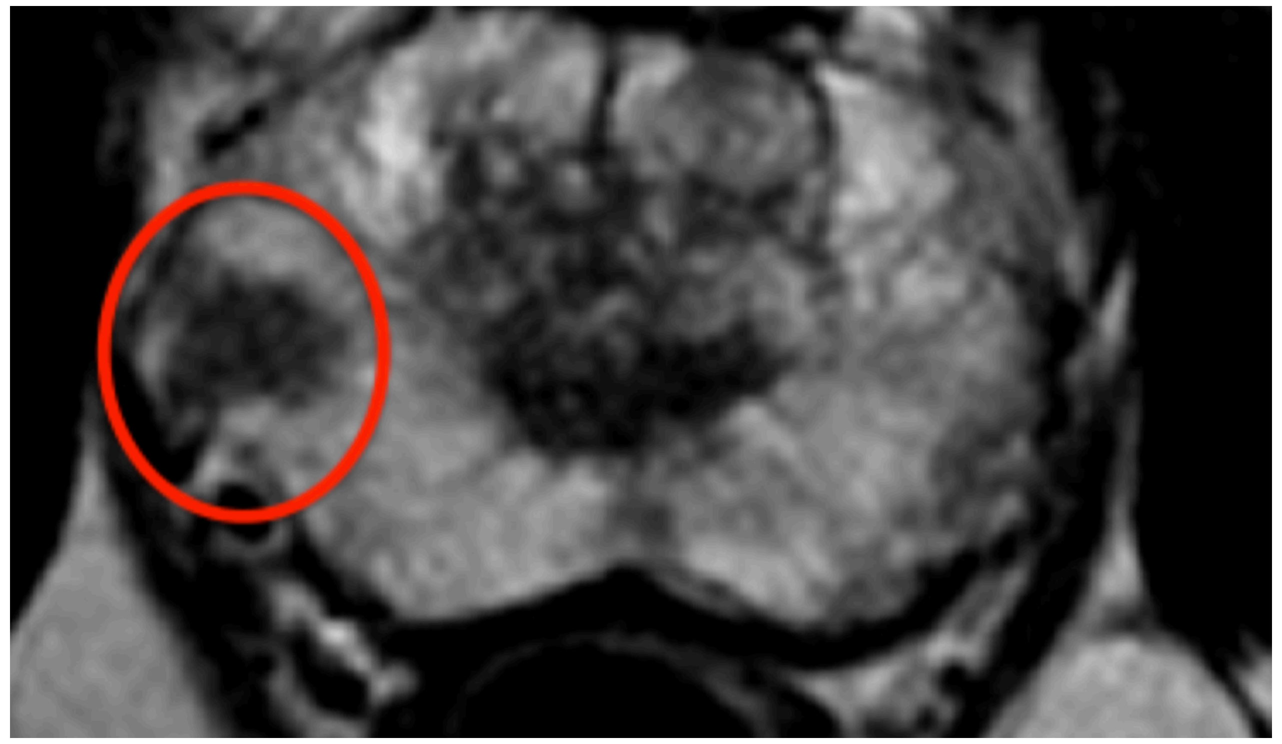
Established imaging modality for diagnosis in oncology, neurology and the muscoloskeletal system.

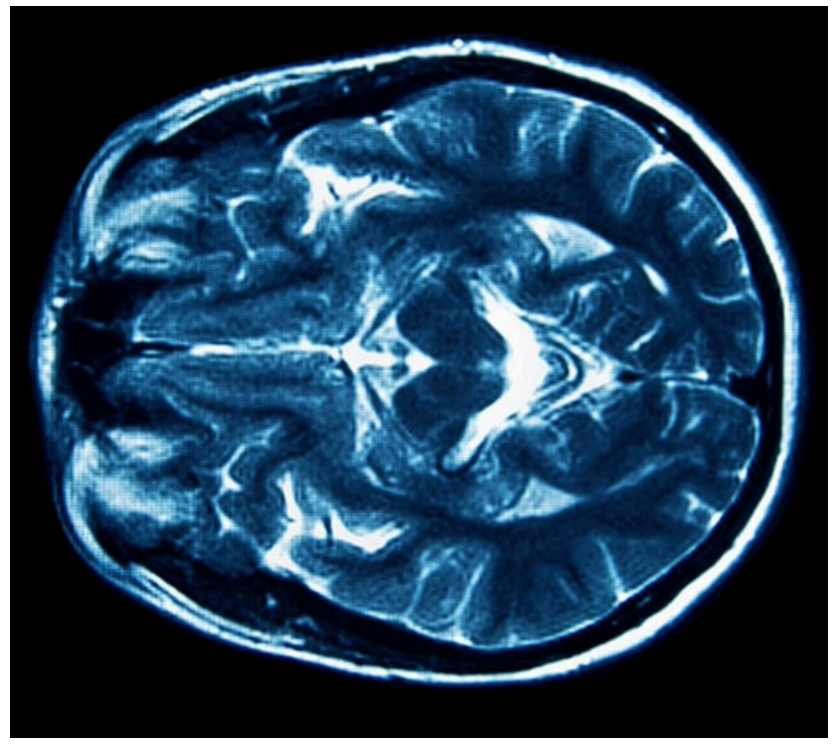




Magnetic Resonance Imaging (MRI)

Established imaging modality for diagnosis in oncology, neurology and the muscoloskeletal system.





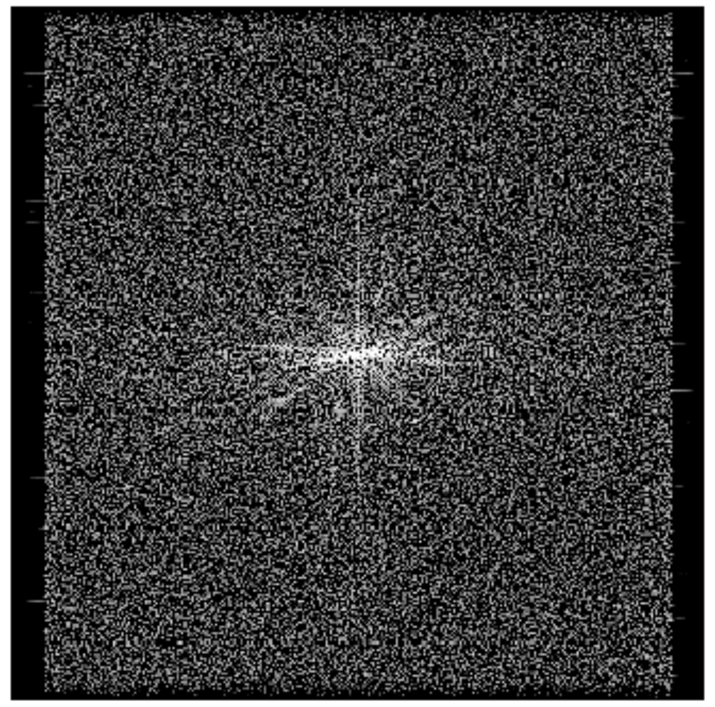
Observation process involves magnetic field that captures the spatial frequency and phase of cross-section through patient tissue:

patient

Magnetic pulses

Fourier transform

observations $\mathbb{C}^{m \times n}$

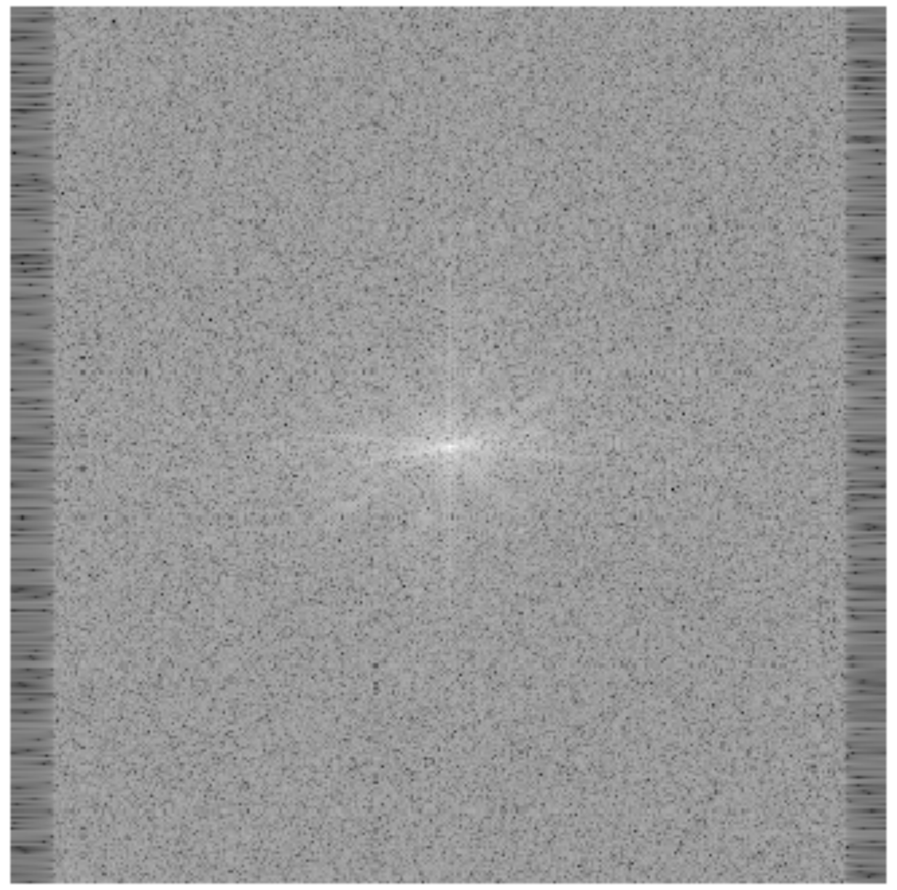


Accelerated Magnetic Resonance Imaging (MRI)

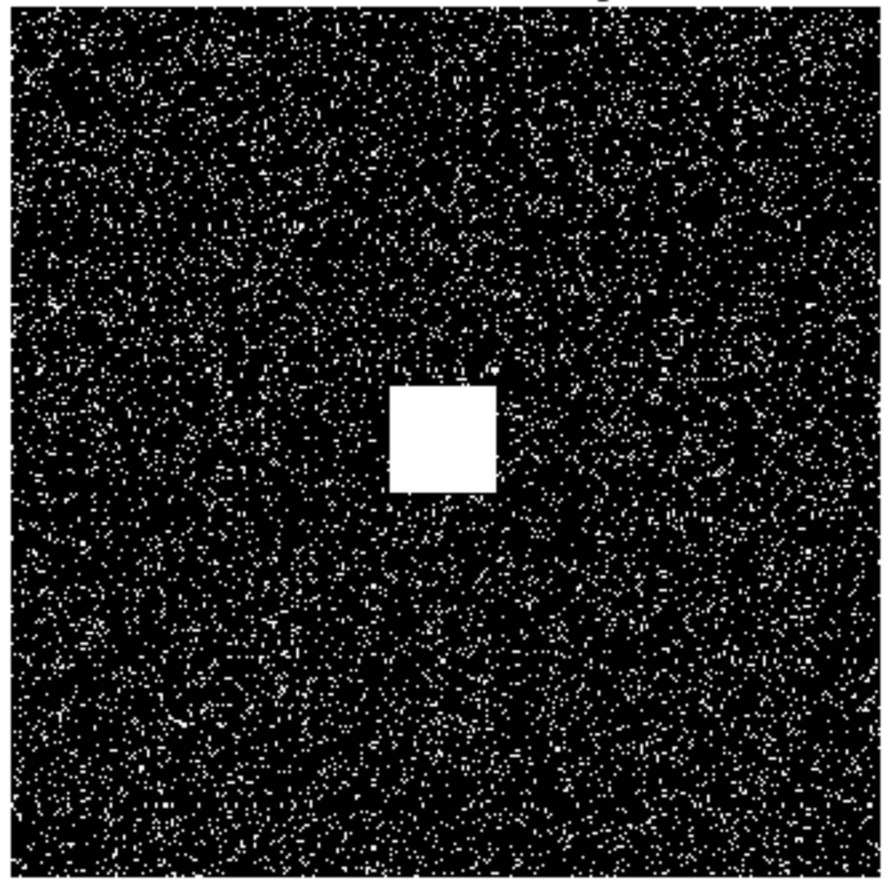
Process is lengthy (easily > 30 min), leads to low patient throughput, problems with patient comfort, artifacts from patient motion, and high exam costs.

Situation: accelerating MRI by subsampling data is important but…

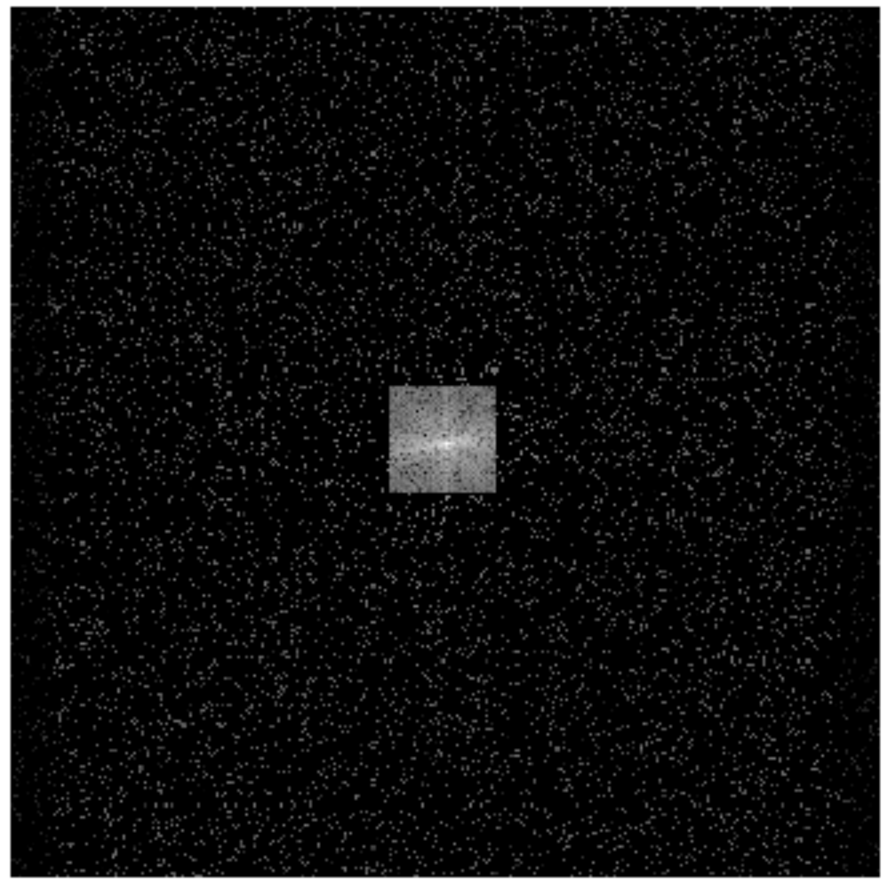
all data



selected locations



accelerated data

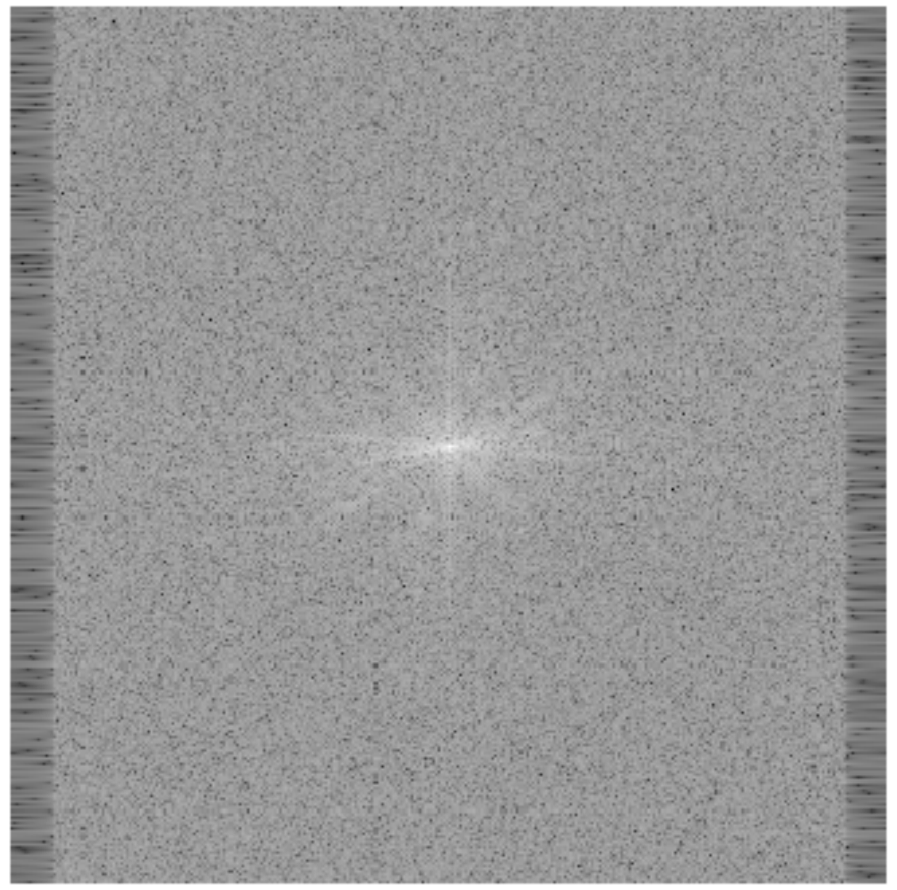


Accelerated Magnetic Resonance Imaging (MRI)

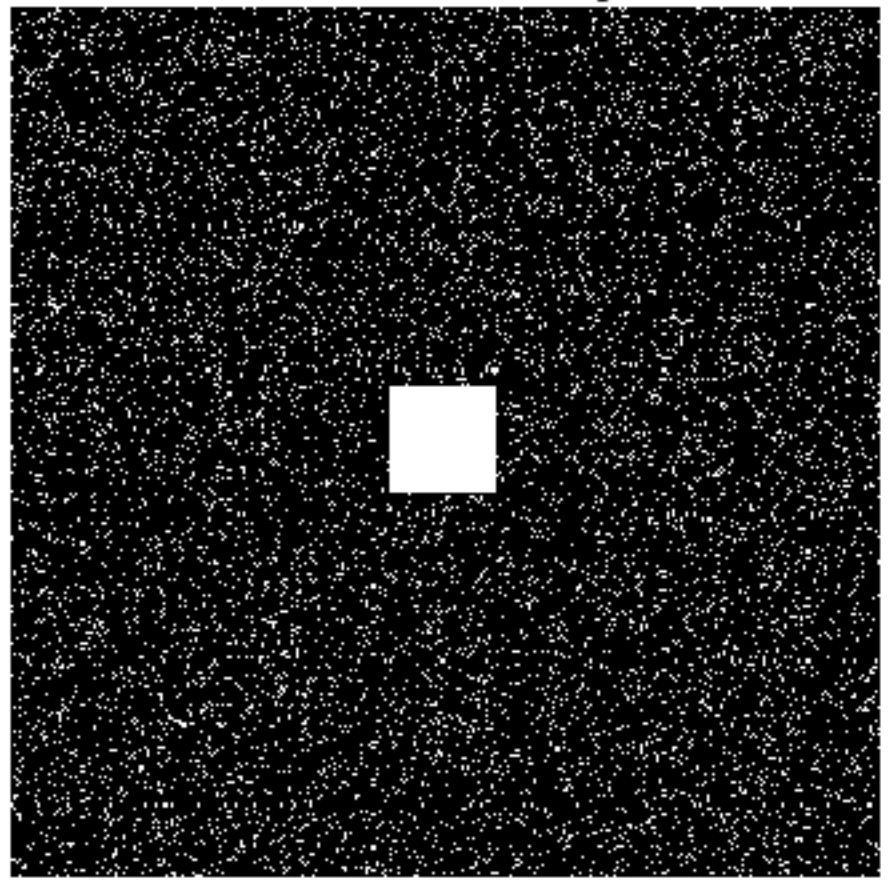
Process is lengthy (easily > 30 min), leads to low patient throughput, problems with patient comfort, artifacts from patient motion, and high exam costs.

Situation: accelerating MRI by subsampling data is important but…

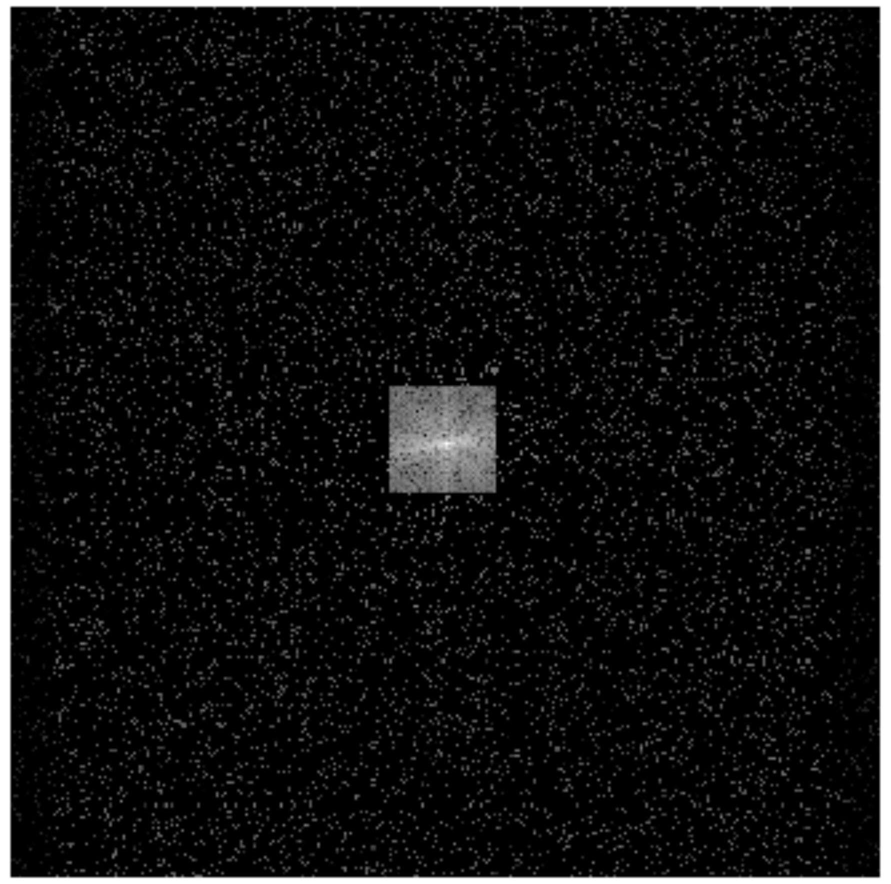
all data



selected locations



accelerated data



Problem: which data points should we measure for best image inference?

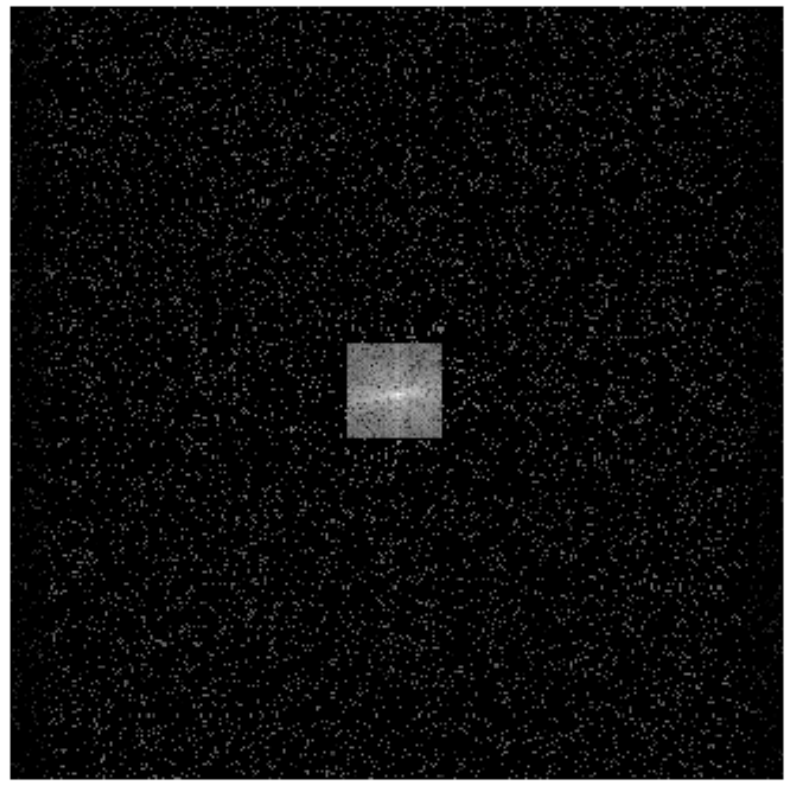
Solution: experimental design

Accelerated Magnetic Resonance Imaging (MRI)

Due to noise and subsampling the imaging is ill-posed thus best solved with Bayesian framework:

given observation $\mathbf{y}$ (acquired w/ our experimental design) the goal is to sample the posterior:

$$\sim p(\mathbf{x}\,|\,\mathbf{y} =$$



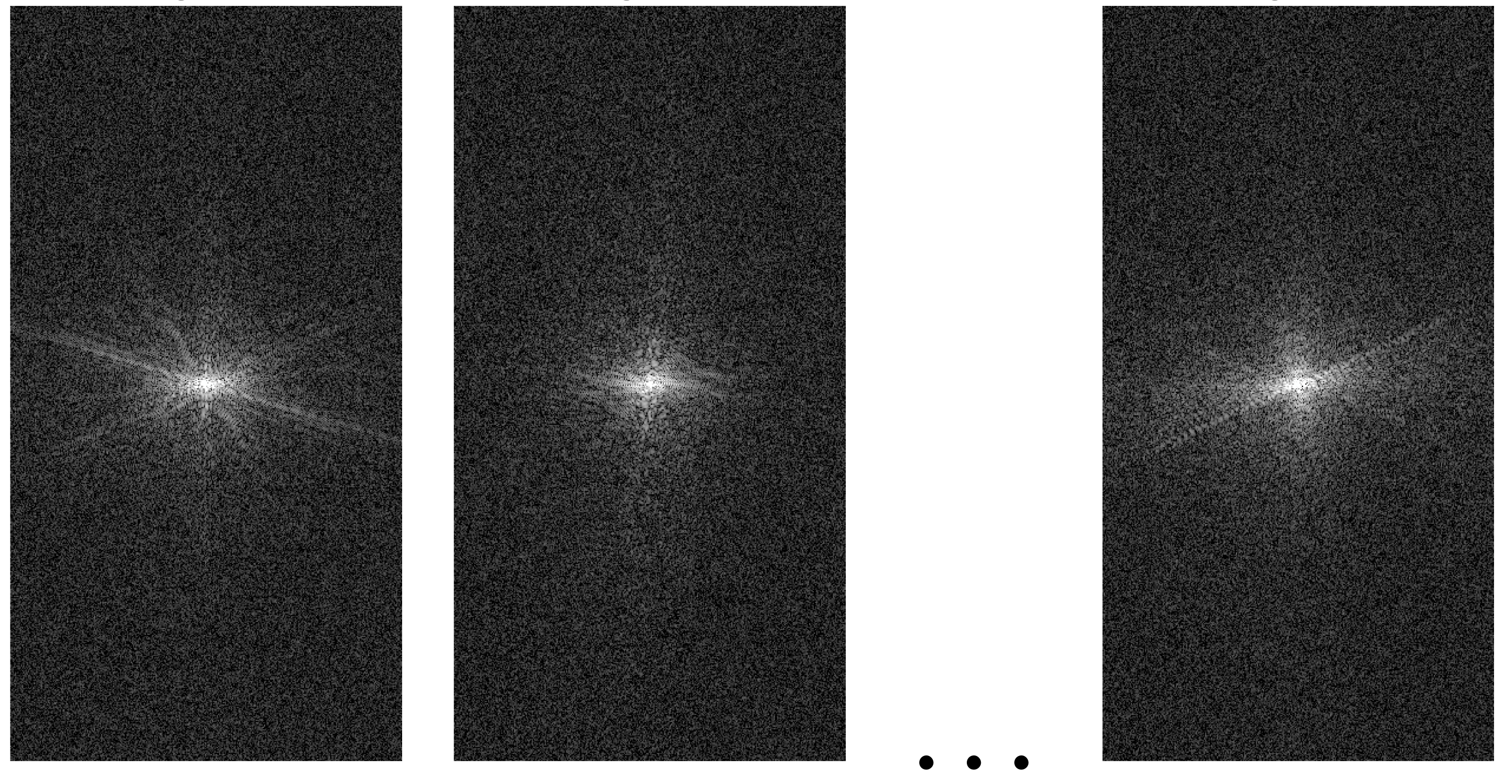
$$)$$

MRI experimental design with normalizing flows

FASTMRI pairs of high quality images $\mathbf{x}^{(i)}$ and fully sampled k-space data $\mathbf{y}^{(i)}$:

$\mathbf{x}^{(1)}$ $\mathbf{x}^{(2)}$ $\mathbf{x}^{(1800)}$

$\mathbf{y}^{(1)}$ $\mathbf{y}^{(2)}$ $\mathbf{y}^{(1800)}$

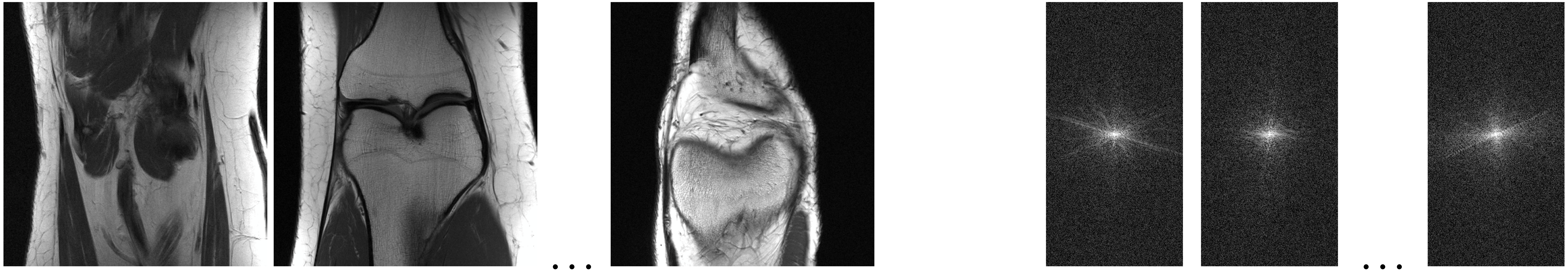


Zbontar, Jure, et al. "fastMRI: An open dataset and benchmarks for accelerated MRI." *arXiv preprint arXiv:1811.08839* (2018).

MRI experimental design with normalizing flows

FASTMRI pairs of high quality images $\mathbf{x}^{(i)}$ and fully sampled k-space data $\mathbf{y}^{(i)}$:

$\mathbf{x}^{(1)}$ $\mathbf{x}^{(2)}$ … $\mathbf{x}^{(1800)}$ $\mathbf{y}^{(1)}$ $\mathbf{y}^{(2)}$ … $\mathbf{y}^{(1800)}$



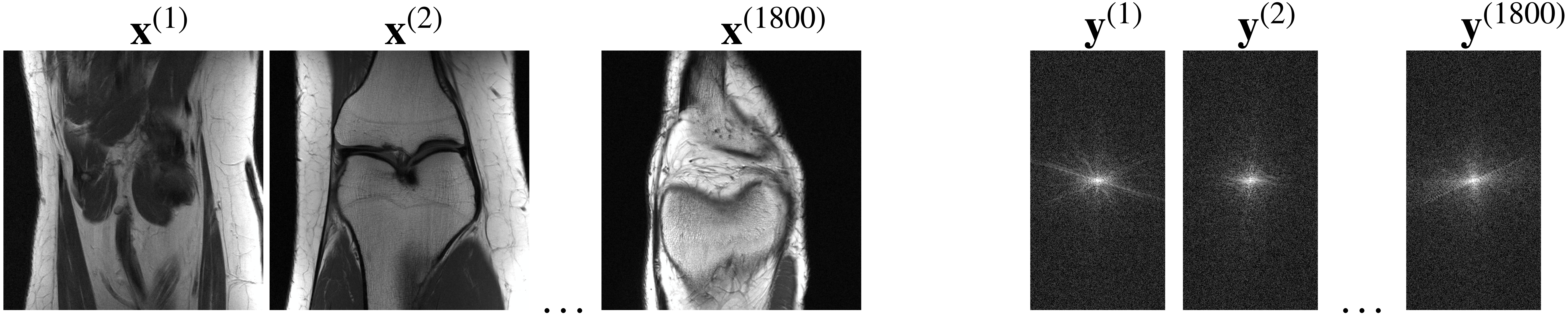
Jointly train normalizing flow and subsampling pattern:

$$\max_{\theta, \mathbf{M}} \frac{1}{N} \sum_{i=1}^{N} \left(-\frac{1}{2} \| f_\theta(\mathbf{x}^{(i)}; \mathbf{A}^\top \mathbf{M} \odot \mathbf{y}^{(i)}) \|_2^2 + \log \left| \det \mathbf{J}_{f_\theta} \right| \right).$$

 Zbontar, Jure, et al. "fastMRI: An open dataset and benchmarks for accelerated MRI." *arXiv preprint arXiv:1811.08839* (2018).

MRI experimental design with normalizing flows

FASTMRI pairs of high quality images $\mathbf{x}^{(i)}$ and fully sampled k-space data $\mathbf{y}^{(i)}$:



Jointly train normalizing flow and subsampling pattern:

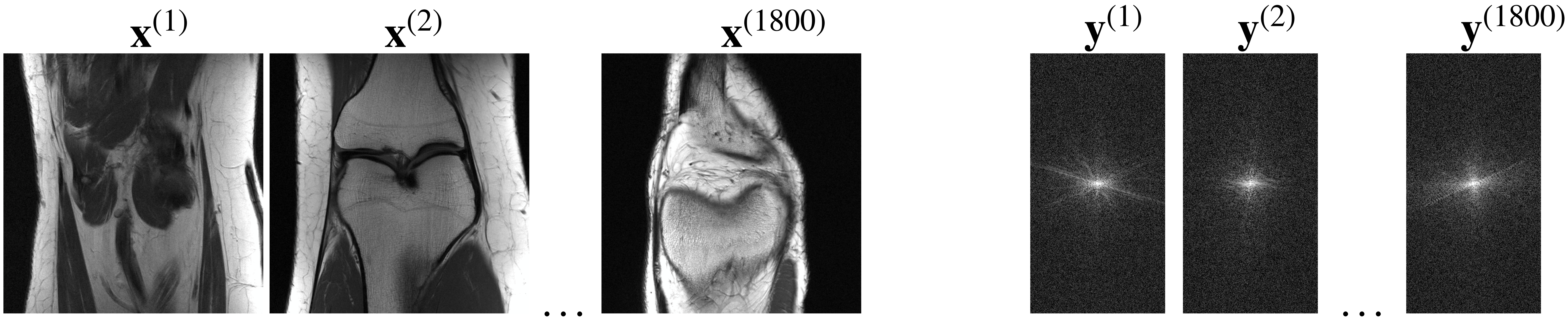
$$\max_{\theta, \mathbf{M}} \frac{1}{N} \sum_{i=1}^{N} \left(-\frac{1}{2} \| f_\theta(\mathbf{x}^{(i)}; \mathbf{A}^\top \mathbf{M} \odot \mathbf{y}^{(i)}) \|_2^2 + \log \left| \det \mathbf{J}_{f_\theta} \right| \right).$$

Problem: how do you optimize binary mask?

Zbontar, Jure, et al. "fastMRI: An open dataset and benchmarks for accelerated MRI." *arXiv preprint arXiv:1811.08839* (2018).

MRI experimental design with normalizing flows

FASTMRI pairs of high quality images $\mathbf{x}^{(i)}$ and fully sampled k-space data $\mathbf{y}^{(i)}$:



Jointly train normalizing flow and subsampling pattern:

$$\max_{\theta, \mathbf{M}} \frac{1}{N} \sum_{i=1}^{N} \left(-\frac{1}{2} \| f_\theta(\mathbf{x}^{(i)}; \mathbf{A}^\top \mathbf{M} \odot \mathbf{y}^{(i)}) \|_2^2 + \log \left| \det \mathbf{J}_{f_\theta} \right| \right).$$

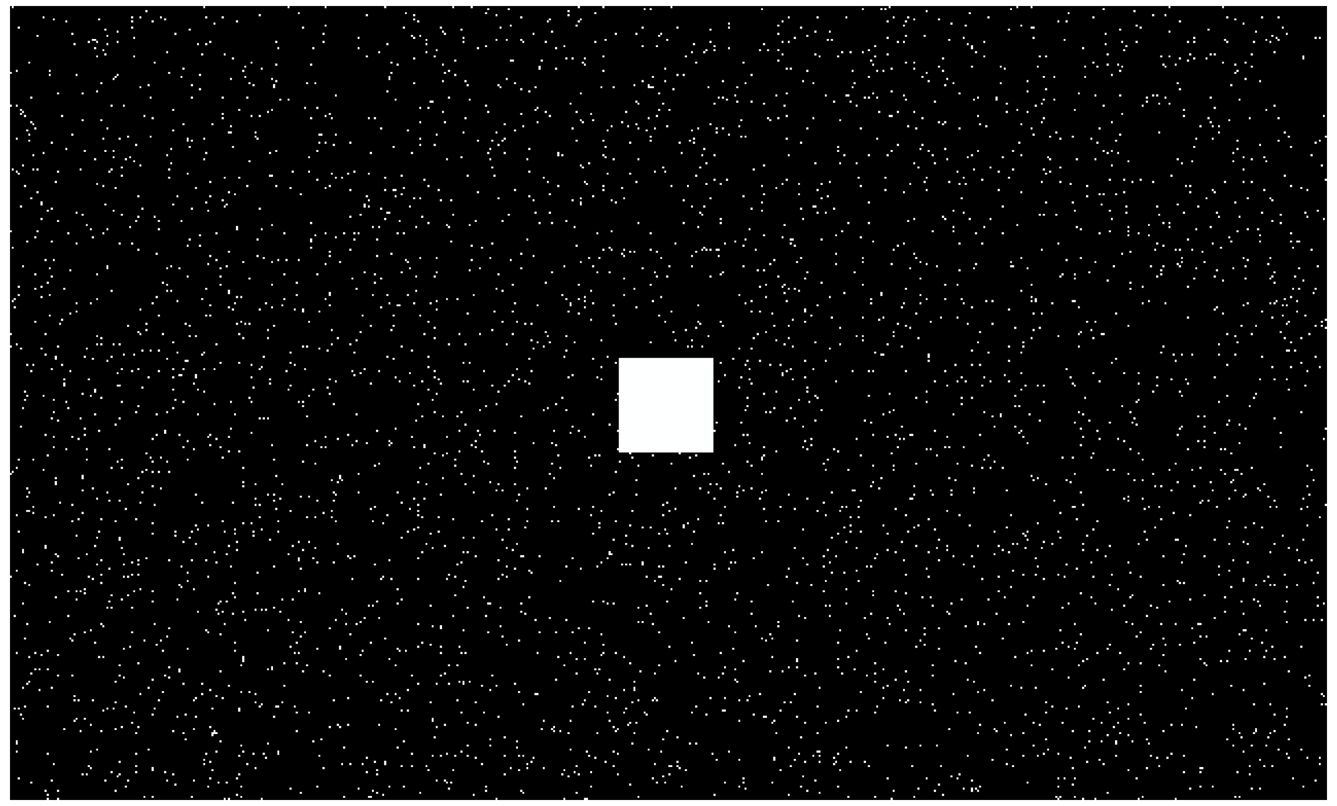
Problem: how do you optimize binary mask?

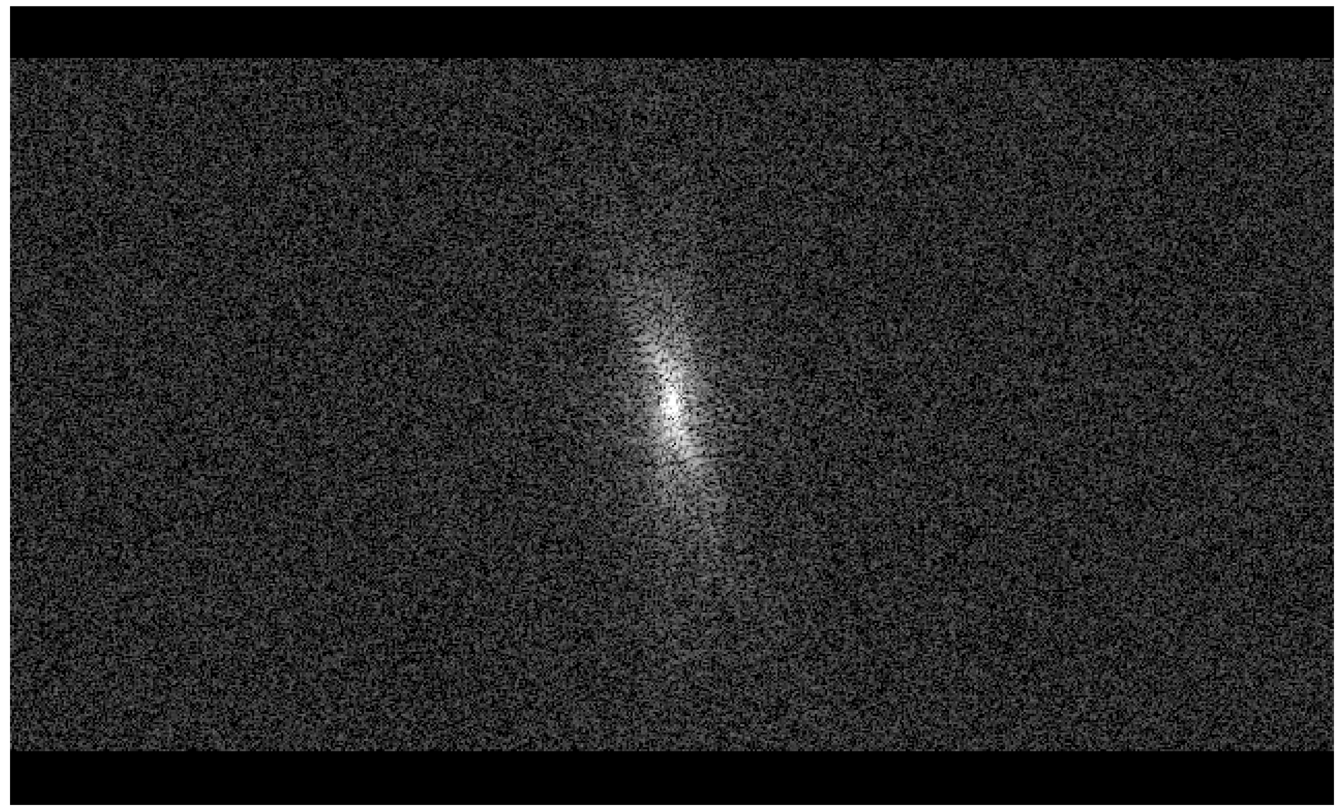
Solution: reinterpret mask as a sampling density.

Zbontar, Jure, et al. "fastMRI: An open dataset and benchmarks for accelerated MRI." *arXiv preprint arXiv:1811.08839* (2018).

Sampling density for receiver placement

Instead of optimizing for binary mask $\mathbf{M} \in \mathbb{Z}^{m \times n}: \ \mathbf{M}_{i,j} = \{0,1\}$



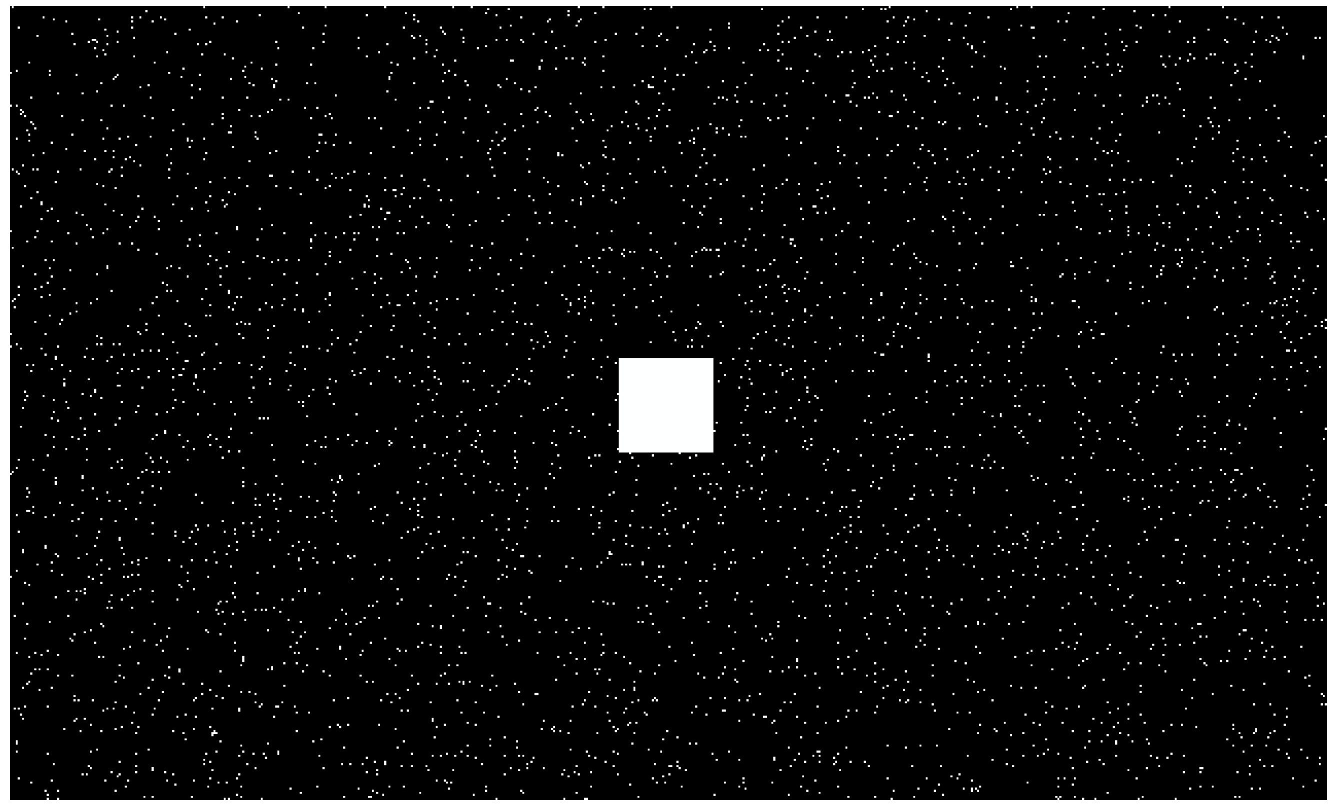


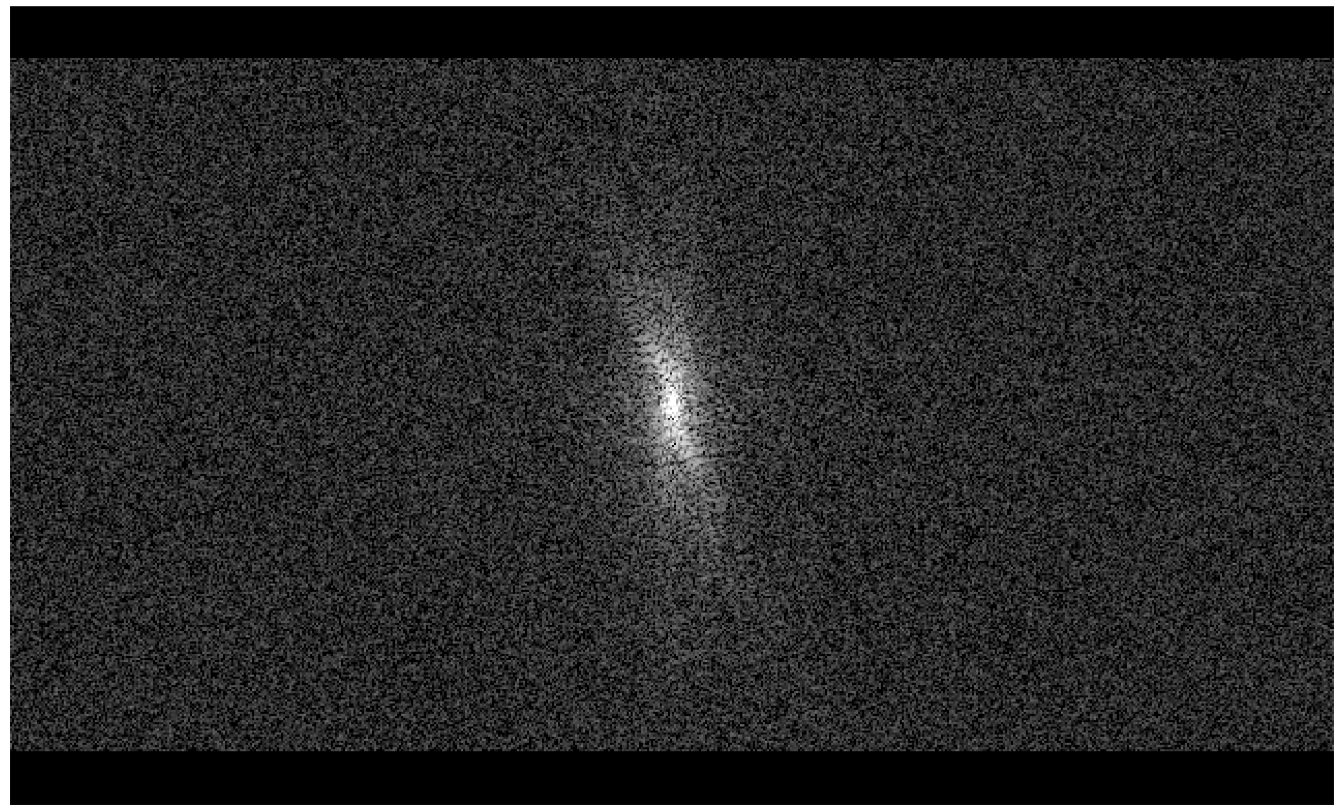
Wu, Sixue, Dirk J. Verschuur, and Gerrit Blacquière. "Automated seismic acquisition geometry design for optimized illumination at the target: A linearized approach." *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021)

Bengio, Yoshua, Nicholas Léonard, and Aaron Courville. "Estimating or propagating gradients through stochastic neurons for conditional computation." *arXiv:1308.3432* (2013).

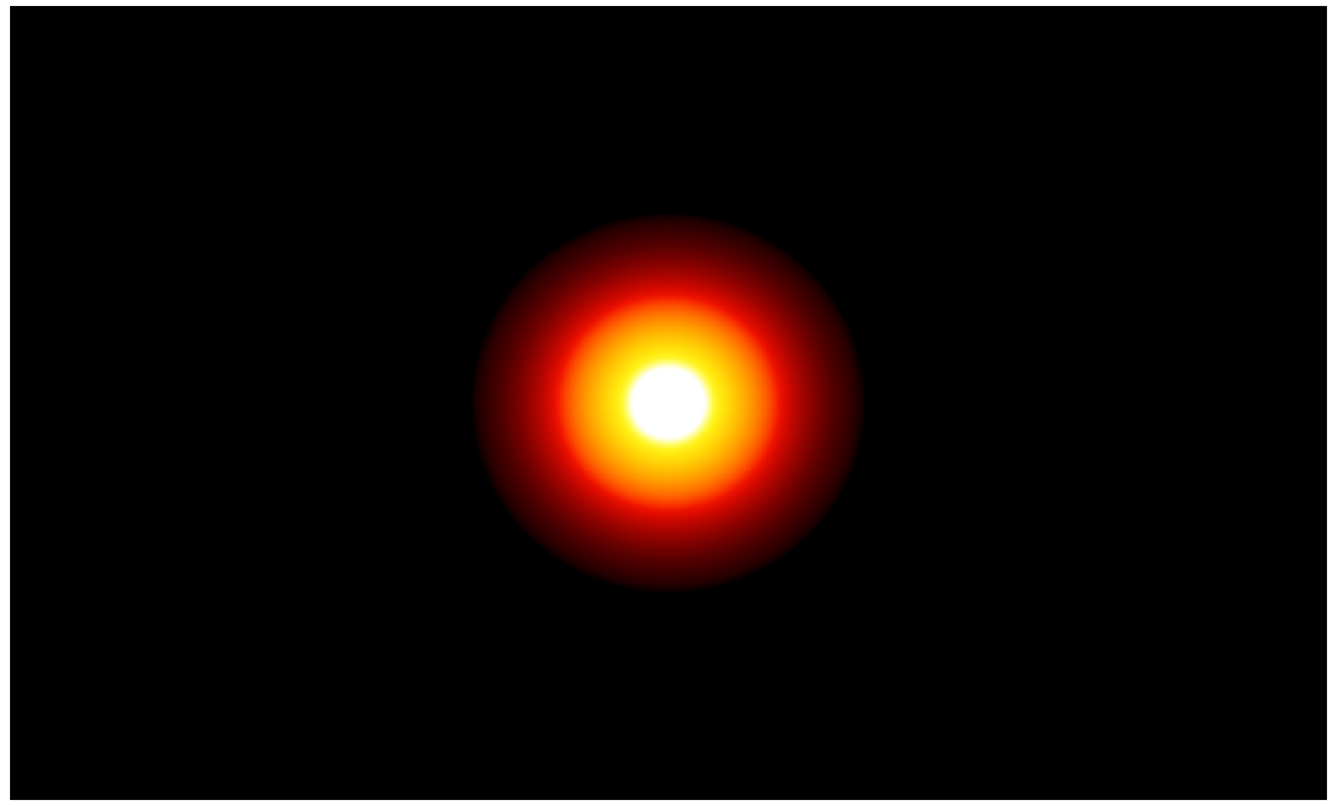
Sampling density for receiver placement

Instead of optimizing for binary mask $\mathbf{M} \in \mathbb{Z}^{m \times n} : \mathbf{M}_{i,j} = \{0,1\}$





optimize for sampling density $\mathbf{w} \in \mathbb{R}^{m \times n}$

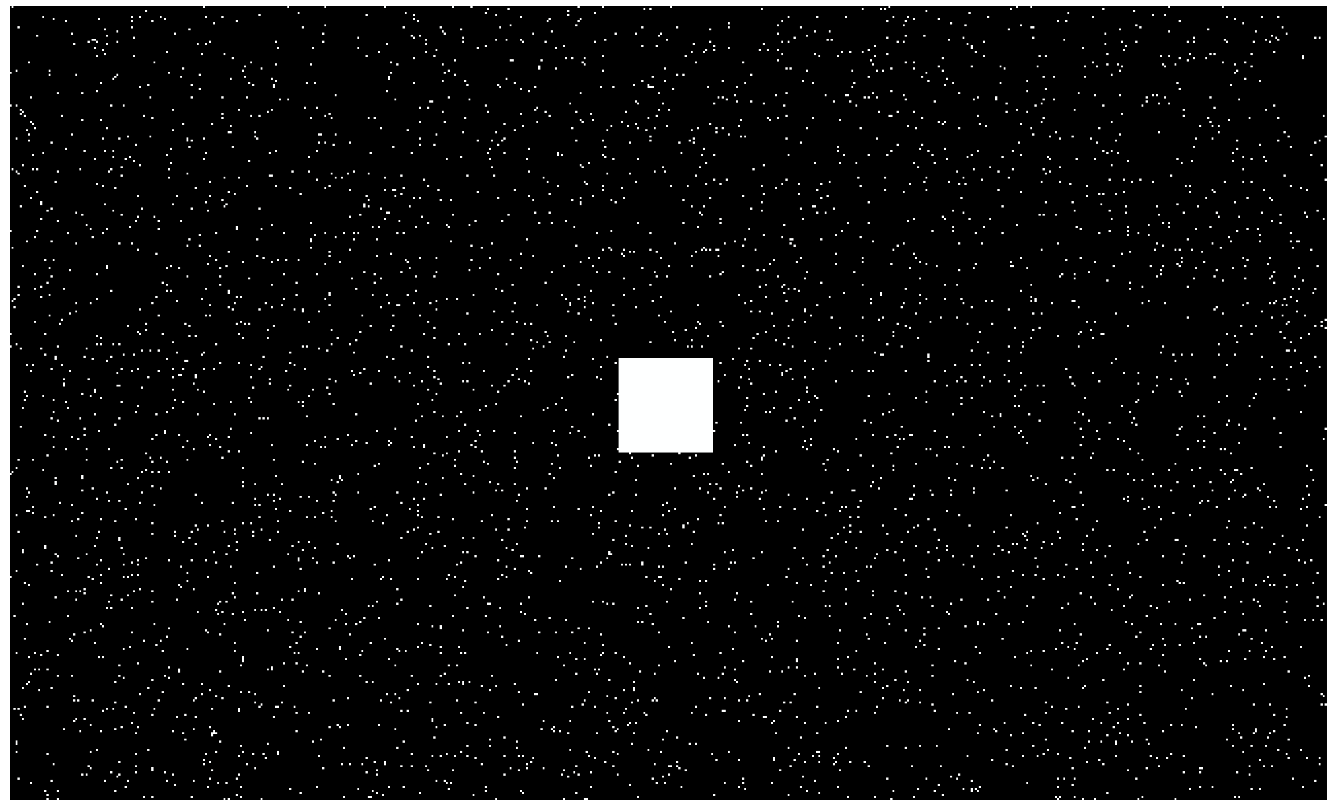


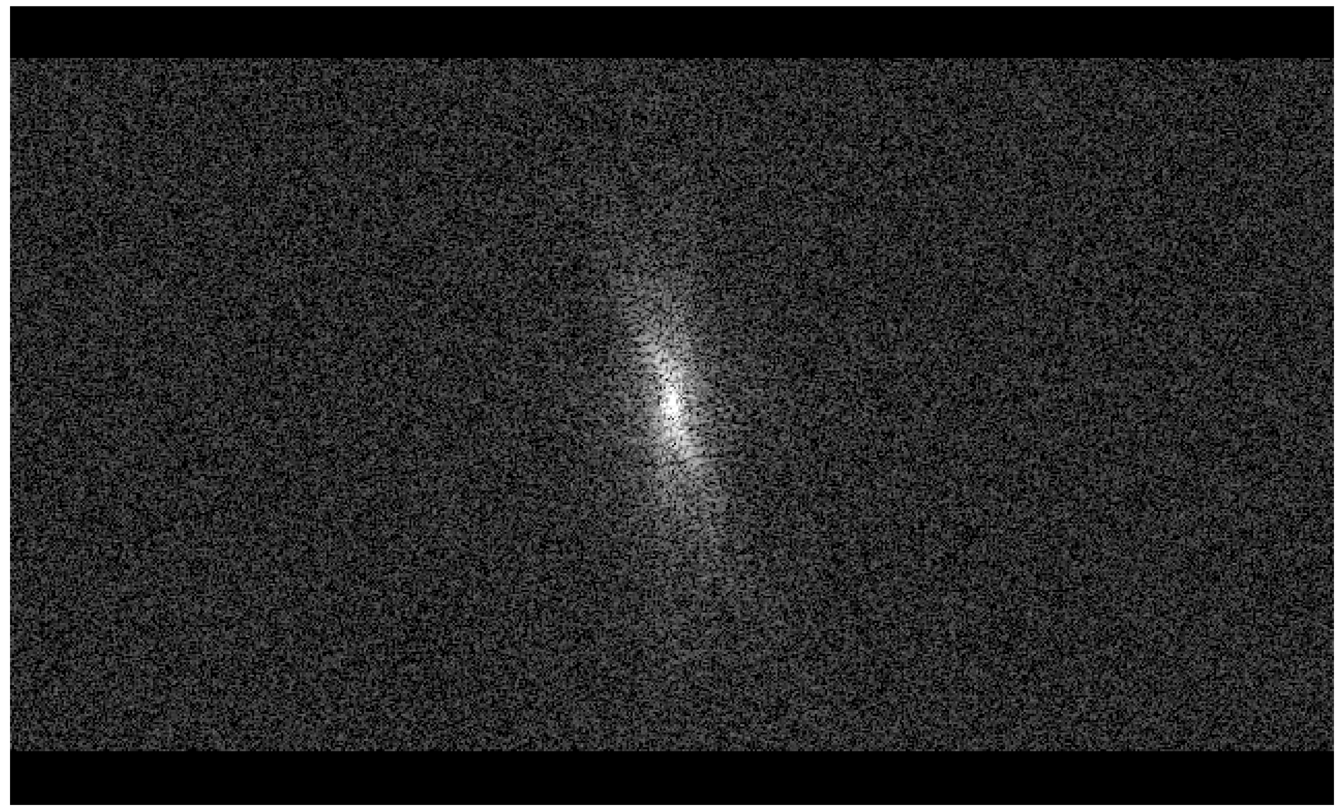
Wu, Sixue, Dirk J. Verschuur, and Gerrit Blacquière. "Automated seismic acquisition geometry design for optimized illumination at the target: A linearized approach." *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021)

Bengio, Yoshua, Nicholas Léonard, and Aaron Courville. "Estimating or propagating gradients through stochastic neurons for conditional computation." *arXiv:1308.3432* (2013).

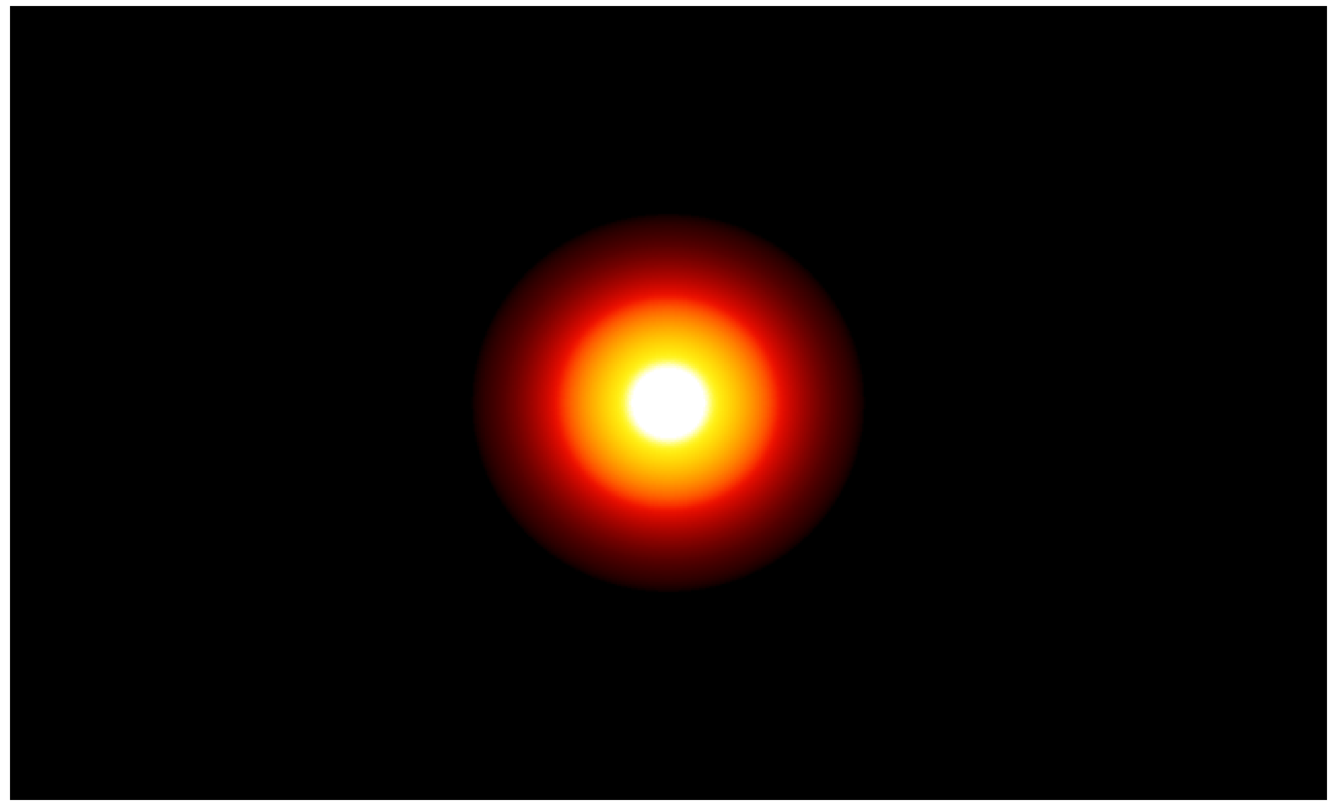
Sampling density for receiver placement

Instead of optimizing for binary mask $\mathbf{M} \in \mathbb{Z}^{m \times n}$: $\mathbf{M}_{i,j} = \{0,1\}$





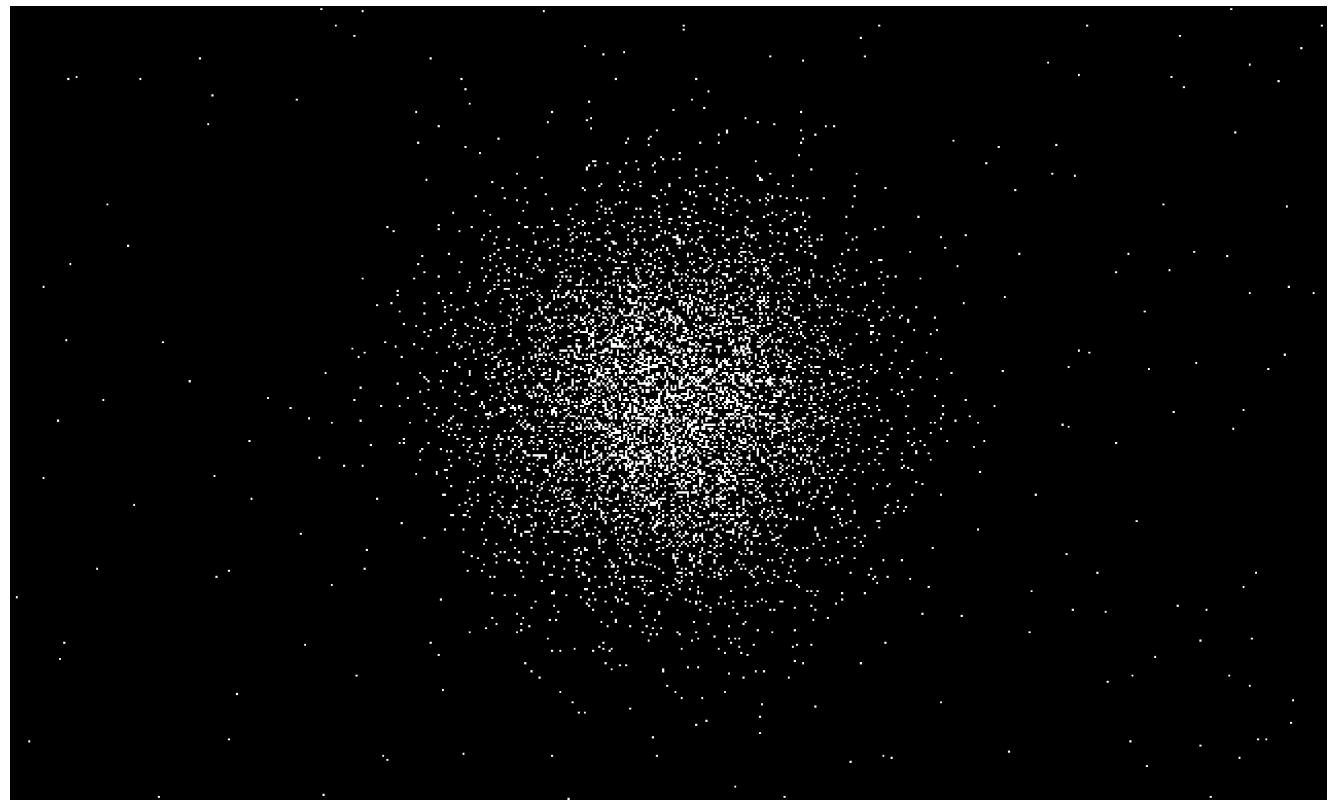
optimize for sampling density $\mathbf{w} \in \mathbb{R}^{m \times n}$



binarize with indicator

$$\mathbf{M}(\mathbf{w}) := \mathbf{1}_{\mathbf{w}<\mathbf{u}}$$

where $\mathbf{u} \sim U(0,1)$.



Wu, Sixue, Dirk J. Verschuur, and Gerrit Blacquière. "Automated seismic acquisition geometry design for optimized illumination at the target: A linearized approach." *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021)

Bengio, Yoshua, Nicholas Léonard, and Aaron Courville. "Estimating or propagating gradients through stochastic neurons for conditional computation." *arXiv:1308.3432* (2013).

MRI experimental design with normalizing flows

FASTMRI pairs of high quality images $\mathbf{x}^{(i)}$ and fully sampled k-space data $\mathbf{y}^{(i)}$:

Jointly train normalizing flow and sampling density:

$$\hat{\theta}, \hat{\mathbf{w}} = \underset{\theta, \mathbf{w}}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^{N} \left(-\frac{1}{2} \| f_{\theta}(\mathbf{x}^{(i)}; \mathbf{A}^{\top}\mathbf{M}(\mathbf{w}) \odot \mathbf{y}^{(i)}) \|_2^2 + \log \left| \det \mathbf{J}_{f_{\theta}} \right| \right).$$

Zbontar, Jure, et al. "fastMRI: An open dataset and benchmarks for accelerated MRI." *arXiv preprint arXiv:1811.08839* (2018).

MRI experimental design with normalizing flows

FASTMRI pairs of high quality images $\mathbf{x}^{(i)}$ and fully sampled k-space data $\mathbf{y}^{(i)}$:

Jointly train normalizing flow and sampling density:

$$\hat{\theta}, \hat{\mathbf{w}} = \underset{\theta, \mathbf{w}}{\operatorname{argmax}} \; \frac{1}{N} \sum_{i=1}^{N} \left(-\frac{1}{2} \| f_\theta(\mathbf{x}^{(i)}; \mathbf{A}^\top \mathbf{M}(\mathbf{w}) \odot \mathbf{y}^{(i)}) \|_2^2 + \log \left| \det \mathbf{J}_{f_\theta} \right| \right) .$$

Binarize during training and enforce budget $s = 0.025$

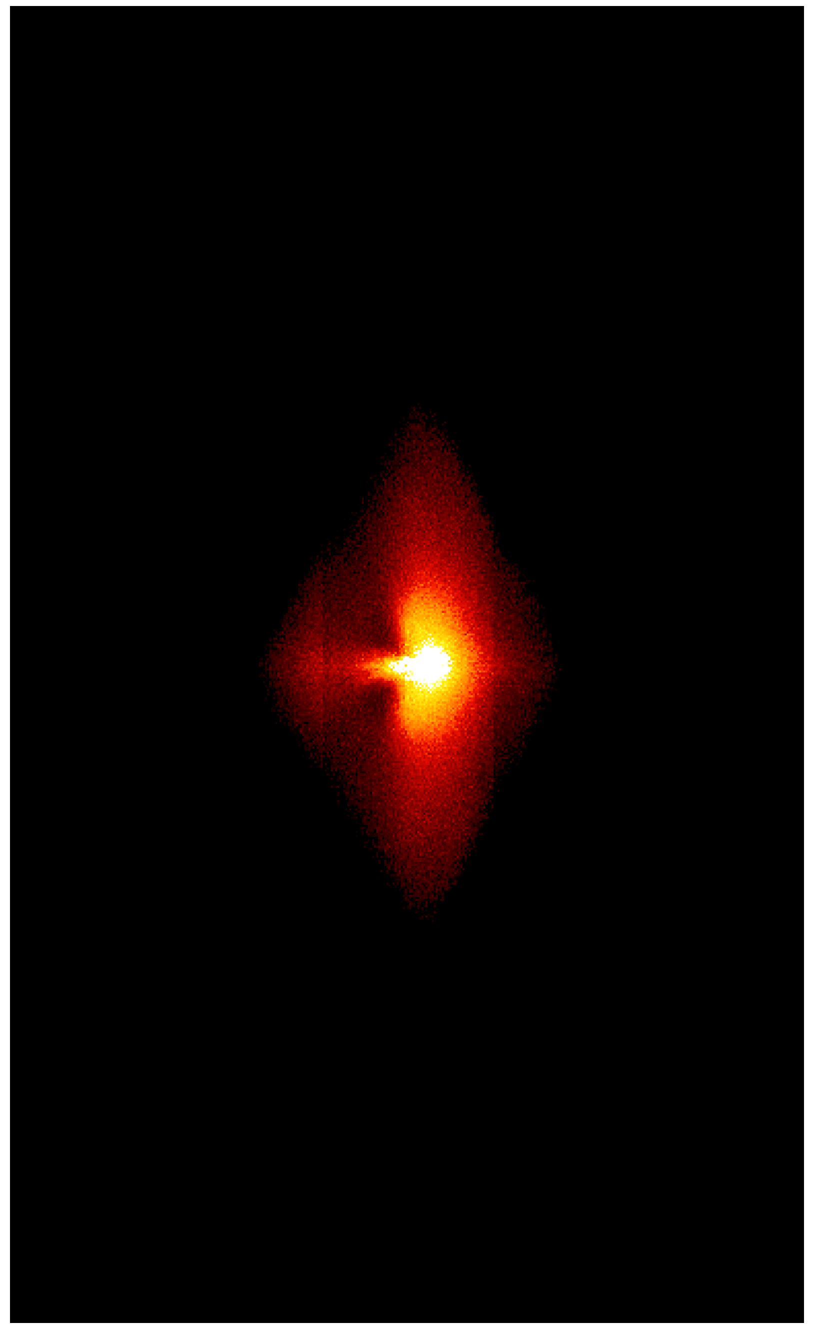
$$\mathbf{M}(\mathbf{w}) := \mathbf{1}_{s \frac{\mathbf{w}}{\overline{\mathbf{w}}} < \mathbf{u}}$$

where $\mathbf{u} \sim U(0,1)$.

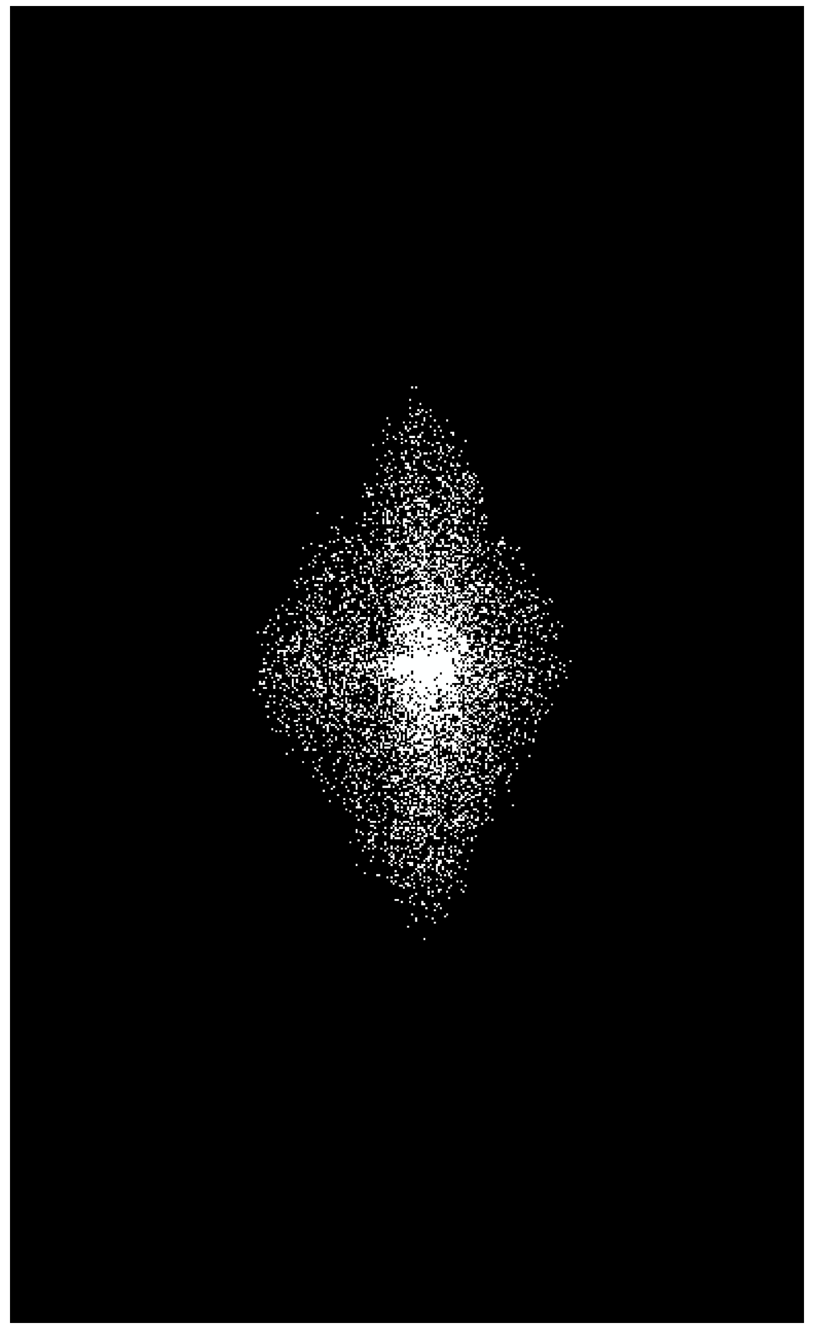
Zbontar, Jure, et al. "fastMRI: An open dataset and benchmarks for accelerated MRI." *arXiv preprint arXiv:1811.08839* (2018).

Optimized experimental design

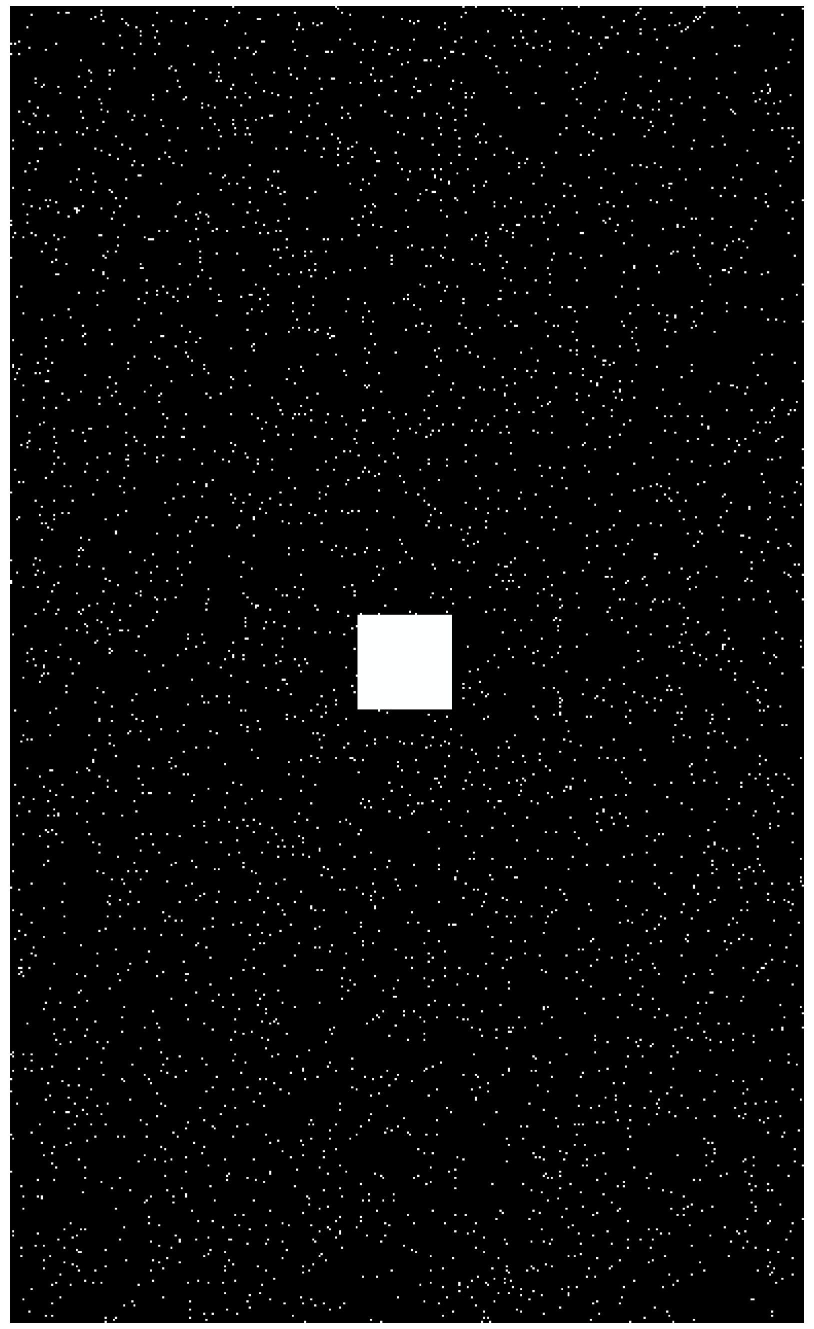
optimal density



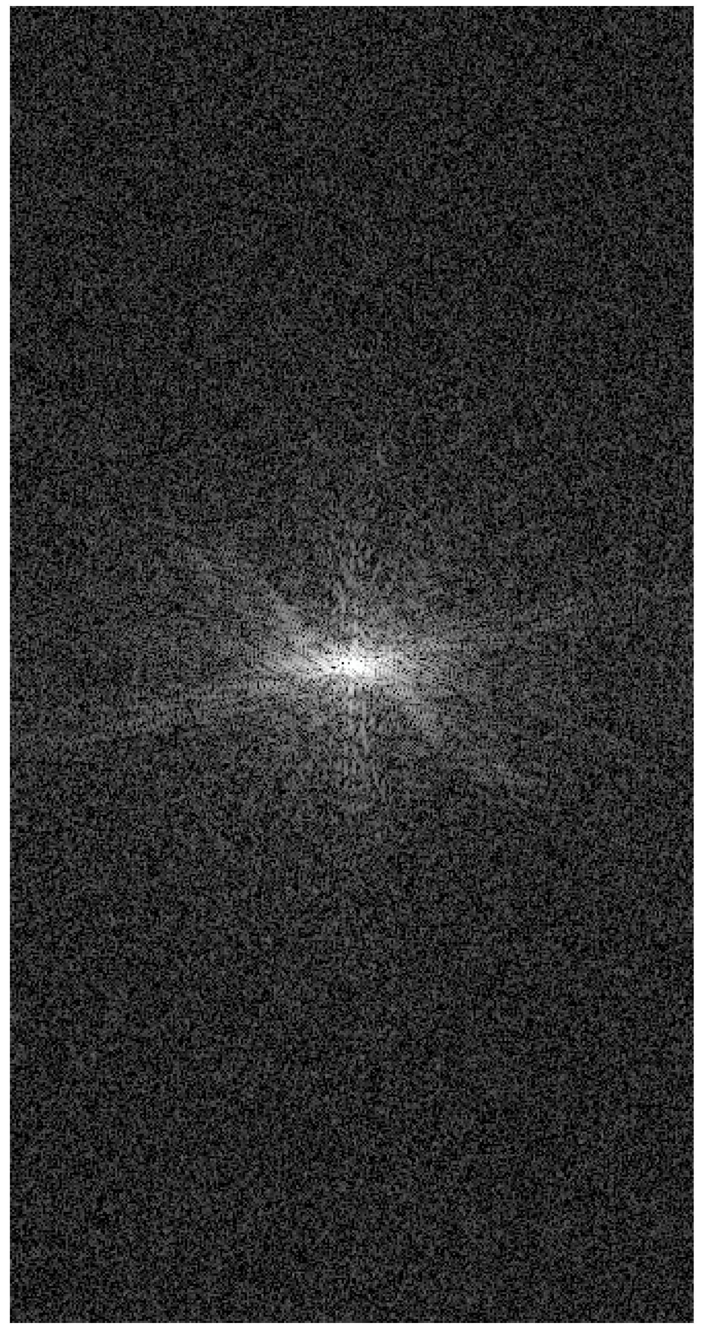
optimal binary



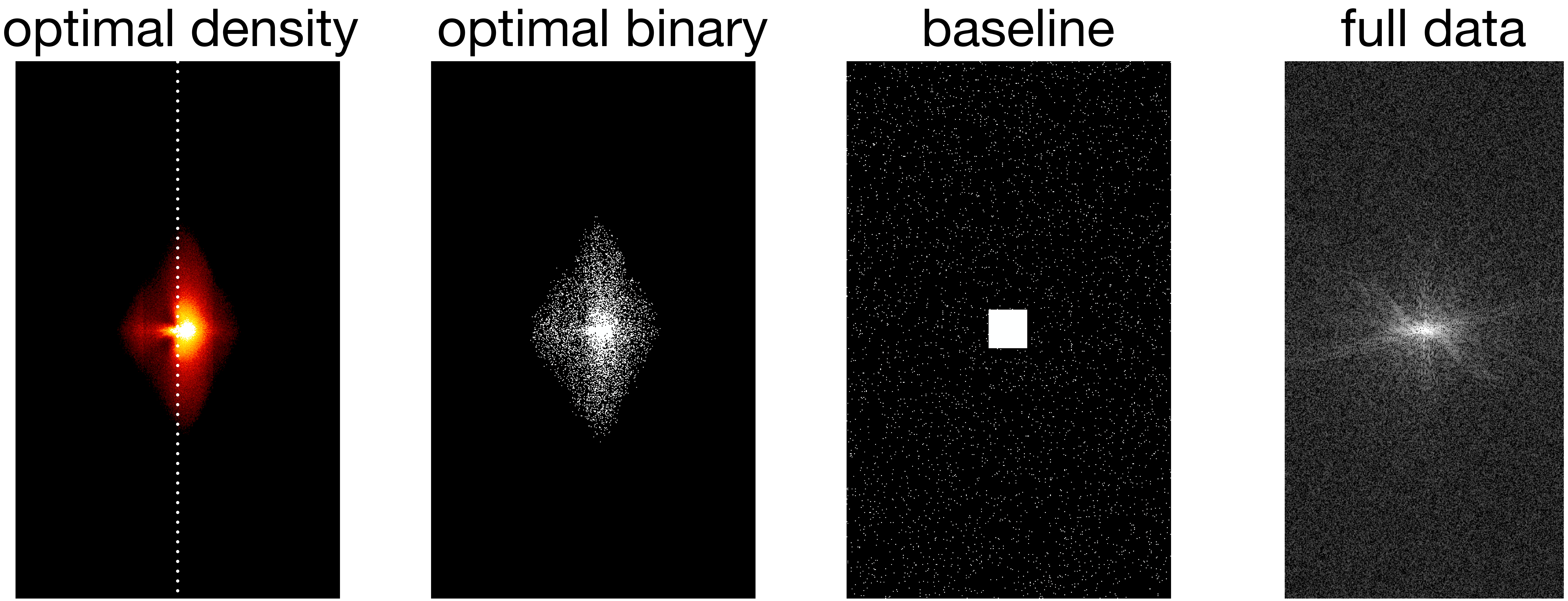
baseline



full data



Optimized experimental design

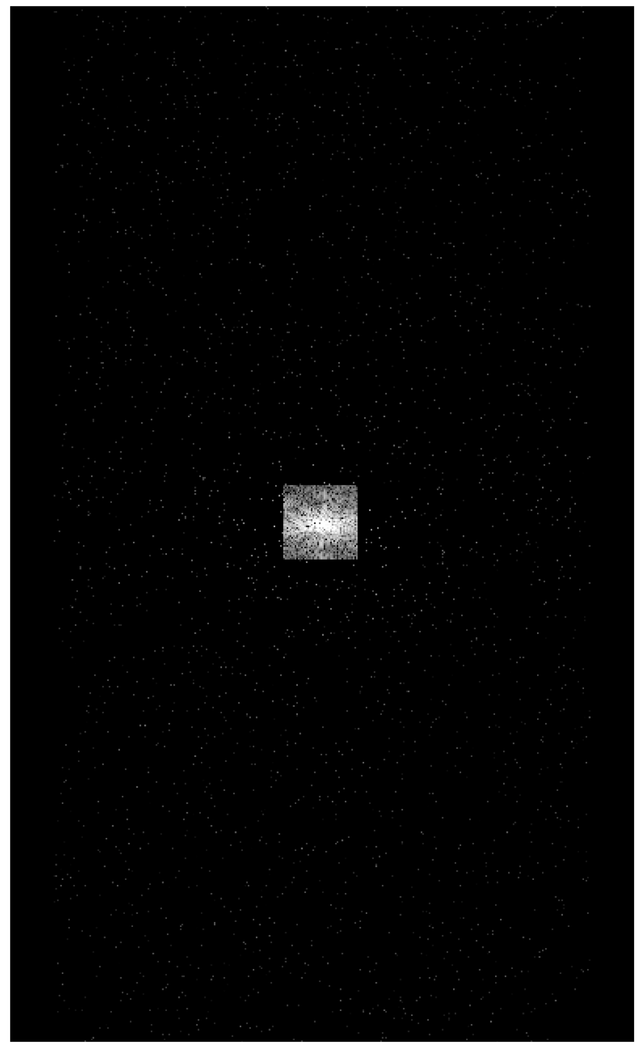


We conclude our optimized density is:

1. centered -> prioritizes low frequencies
2. ellipsoid -> prioritizes vertical elements in k-space
3. asymmetric -> learns to exploit Hermitian symmetry

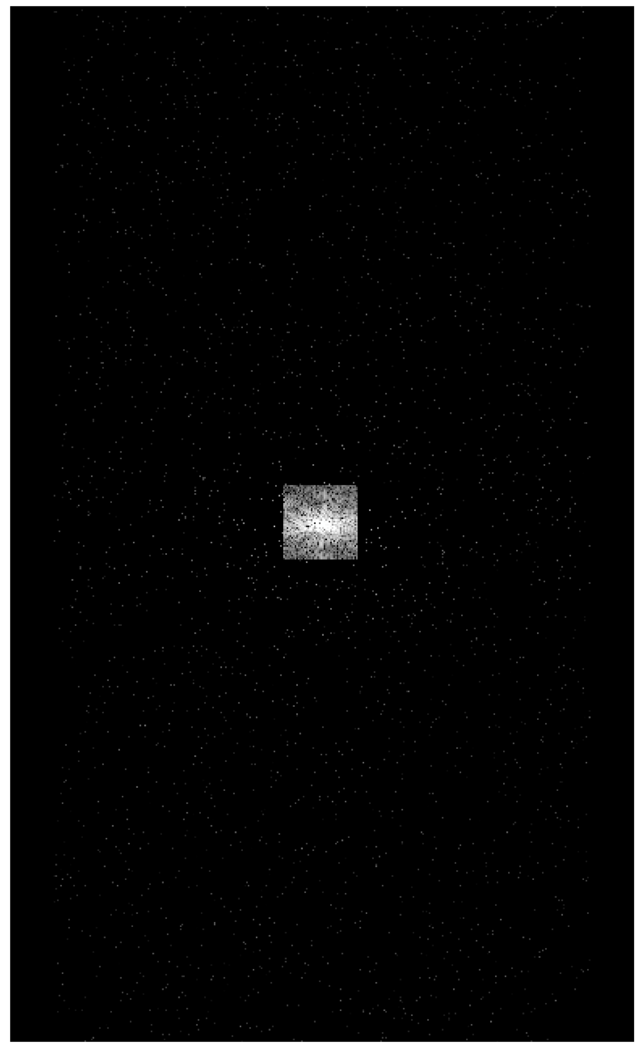
Posterior sampling w/ optimal design

Baseline posterior samples:

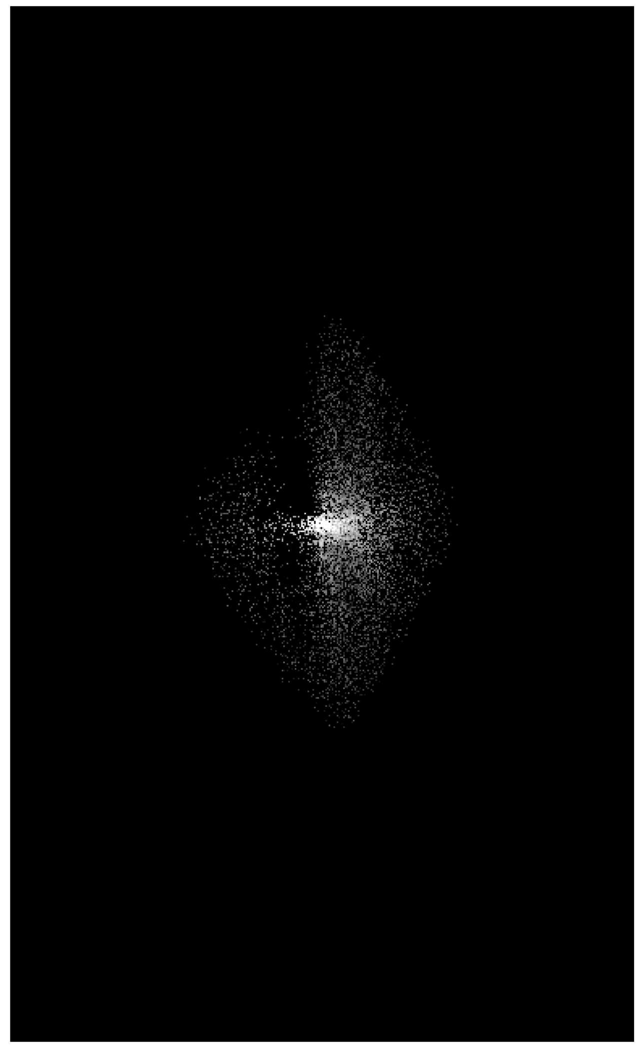


Posterior sampling w/ optimal design

Baseline posterior samples:

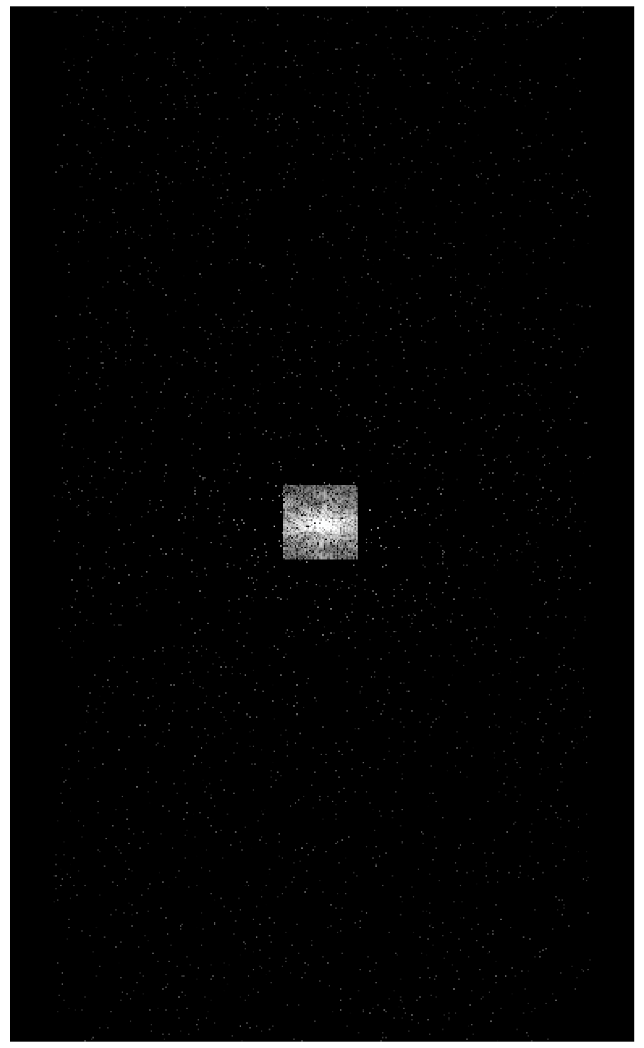


Our posterior samples
w/ optimal design:



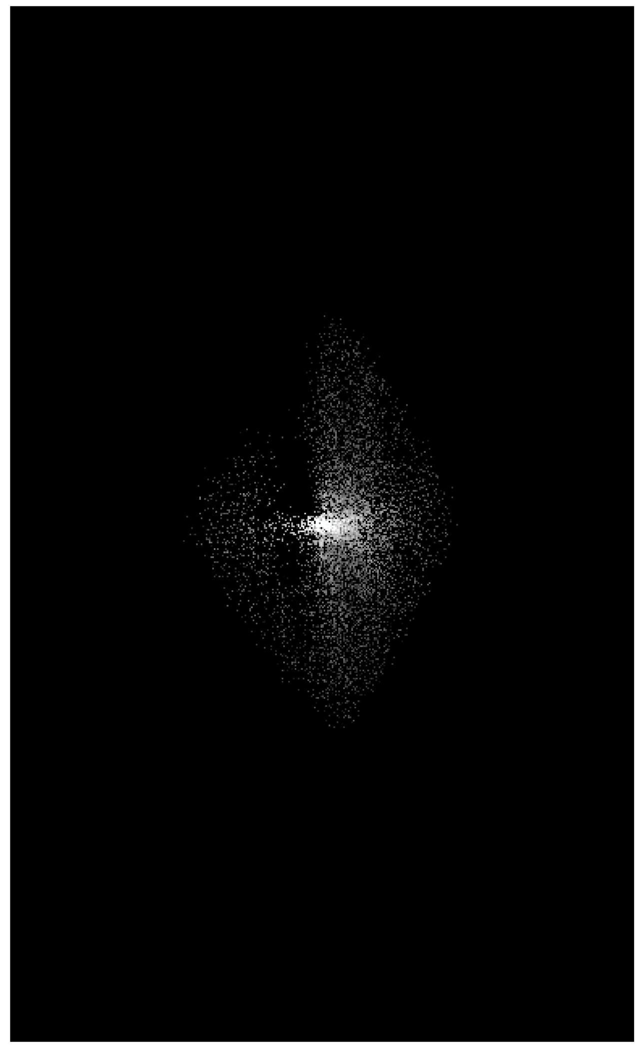
Posterior sampling w/ optimal design

Baseline posterior samples:



Reference image

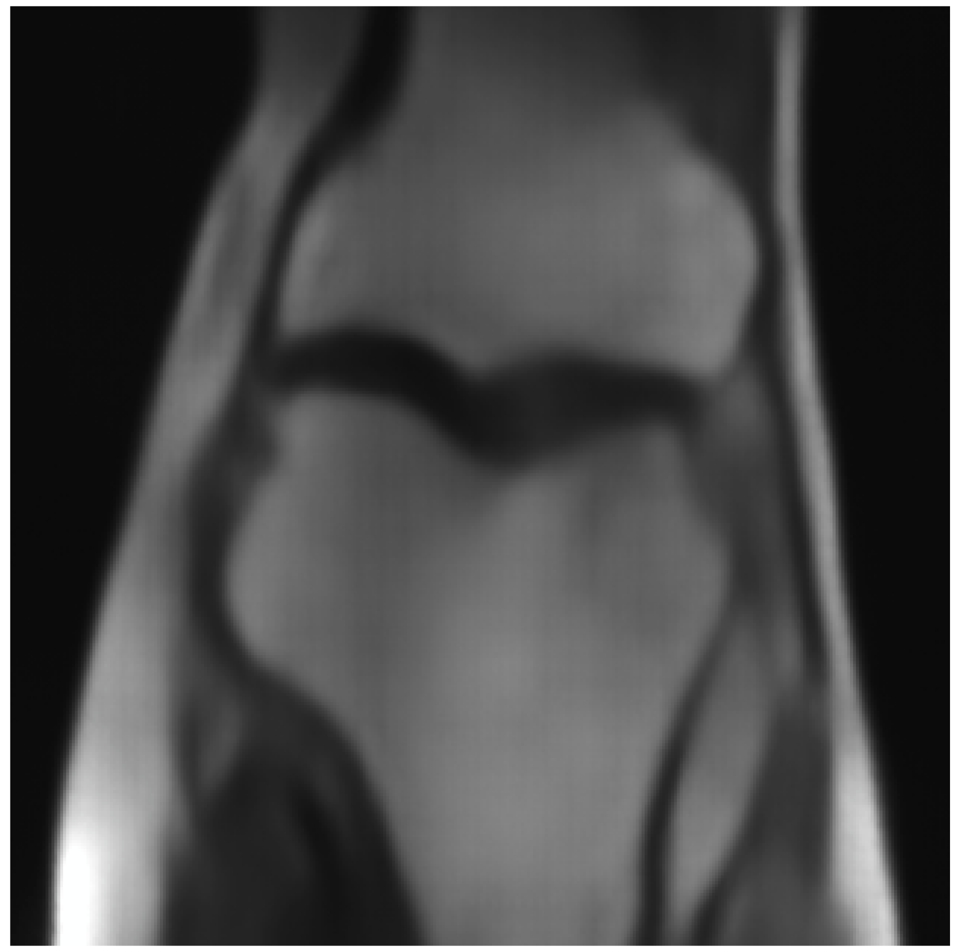
Our posterior samples w/ optimal design:



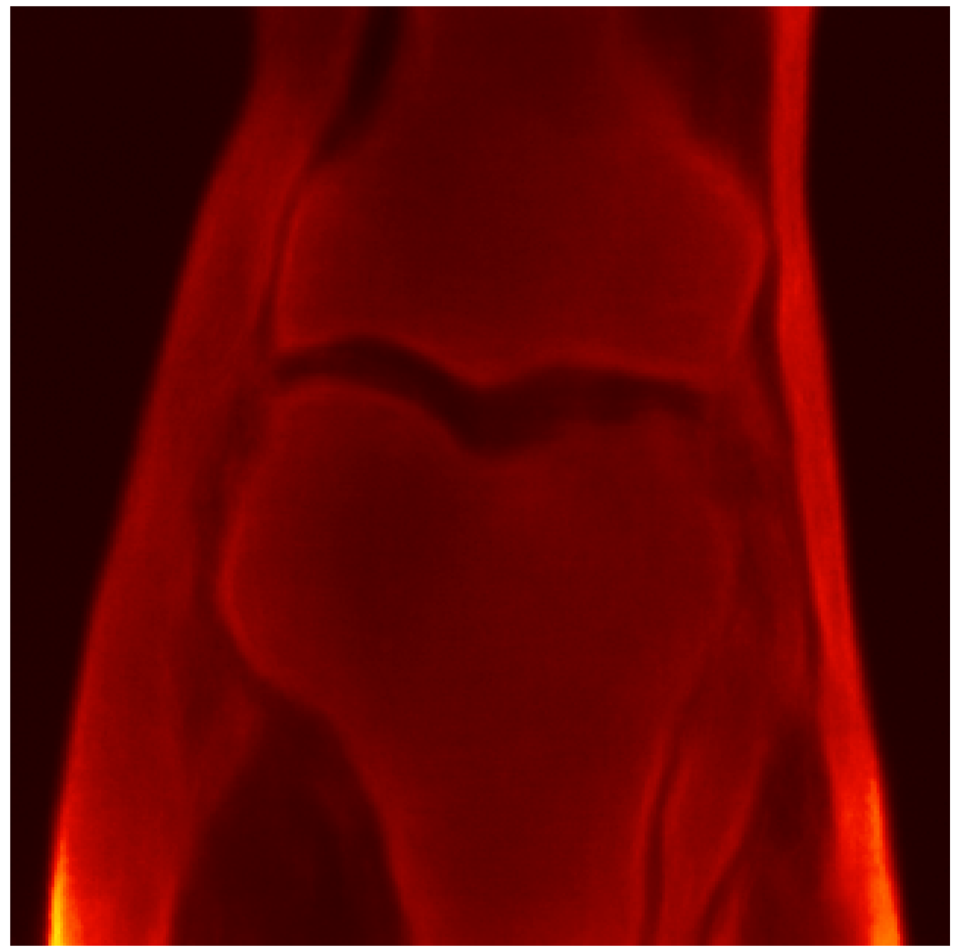
Posterior statistics

Fast sampling w/ normalizing flow to efficiently estimate statistical moments i.e. mean, standard deviation:

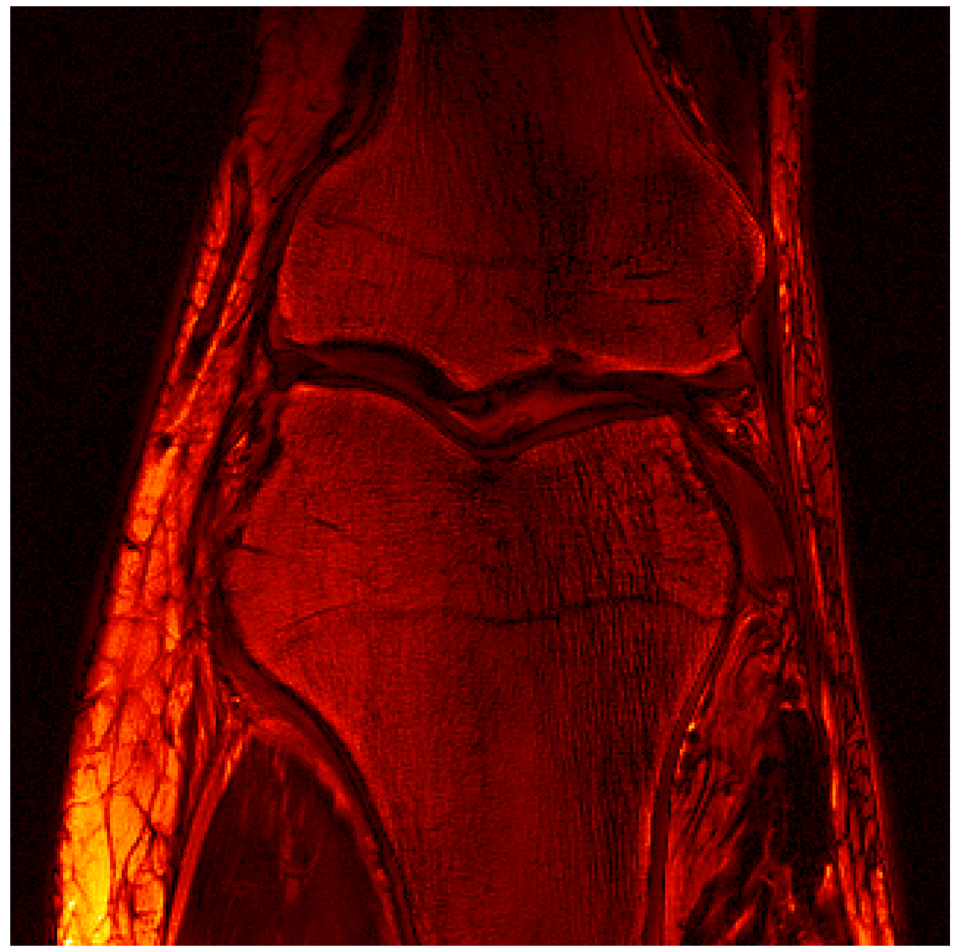
Mean SSIM=0.57



Standard deviation



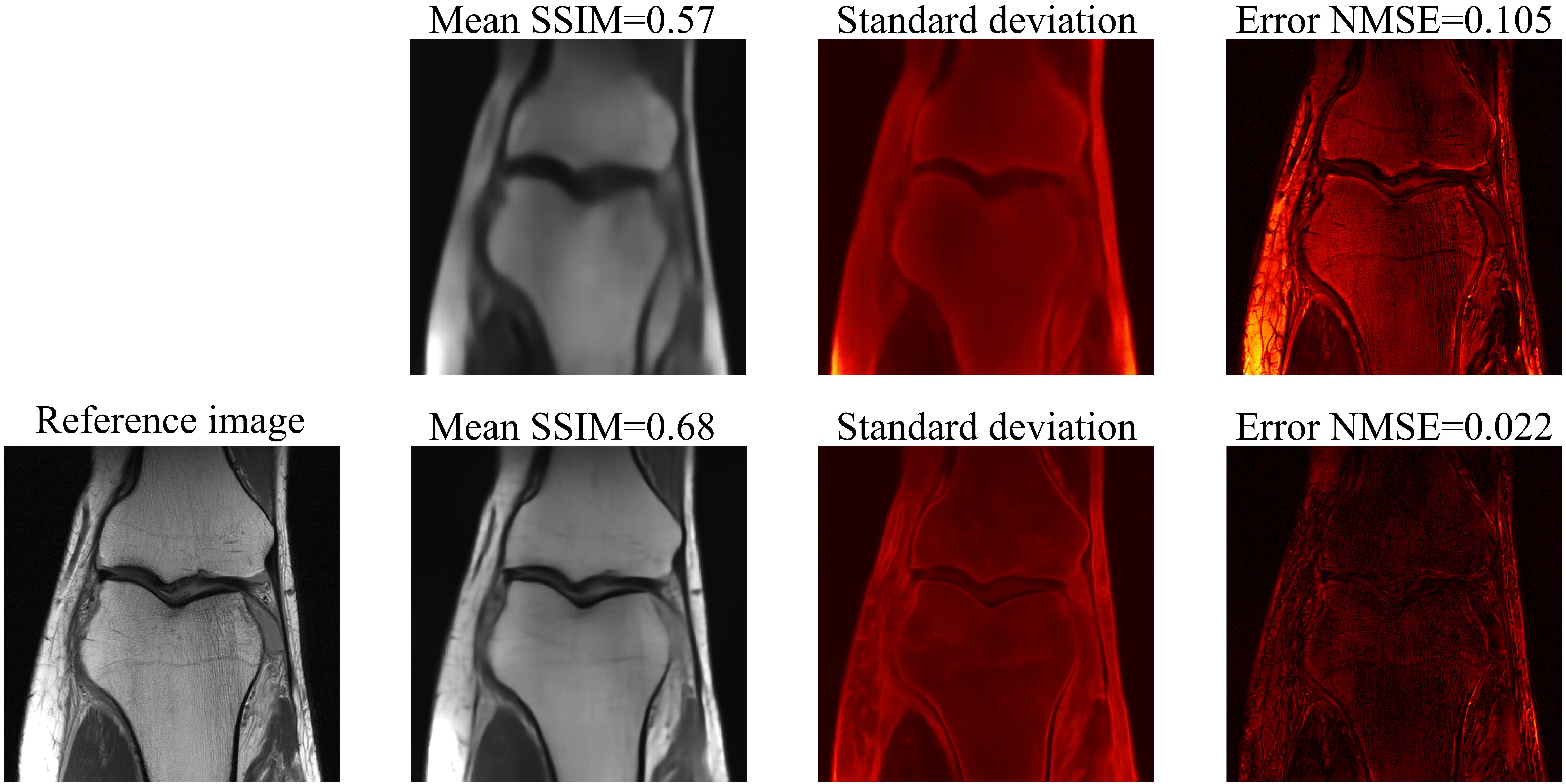
Error NMSE=0.105



Reference image

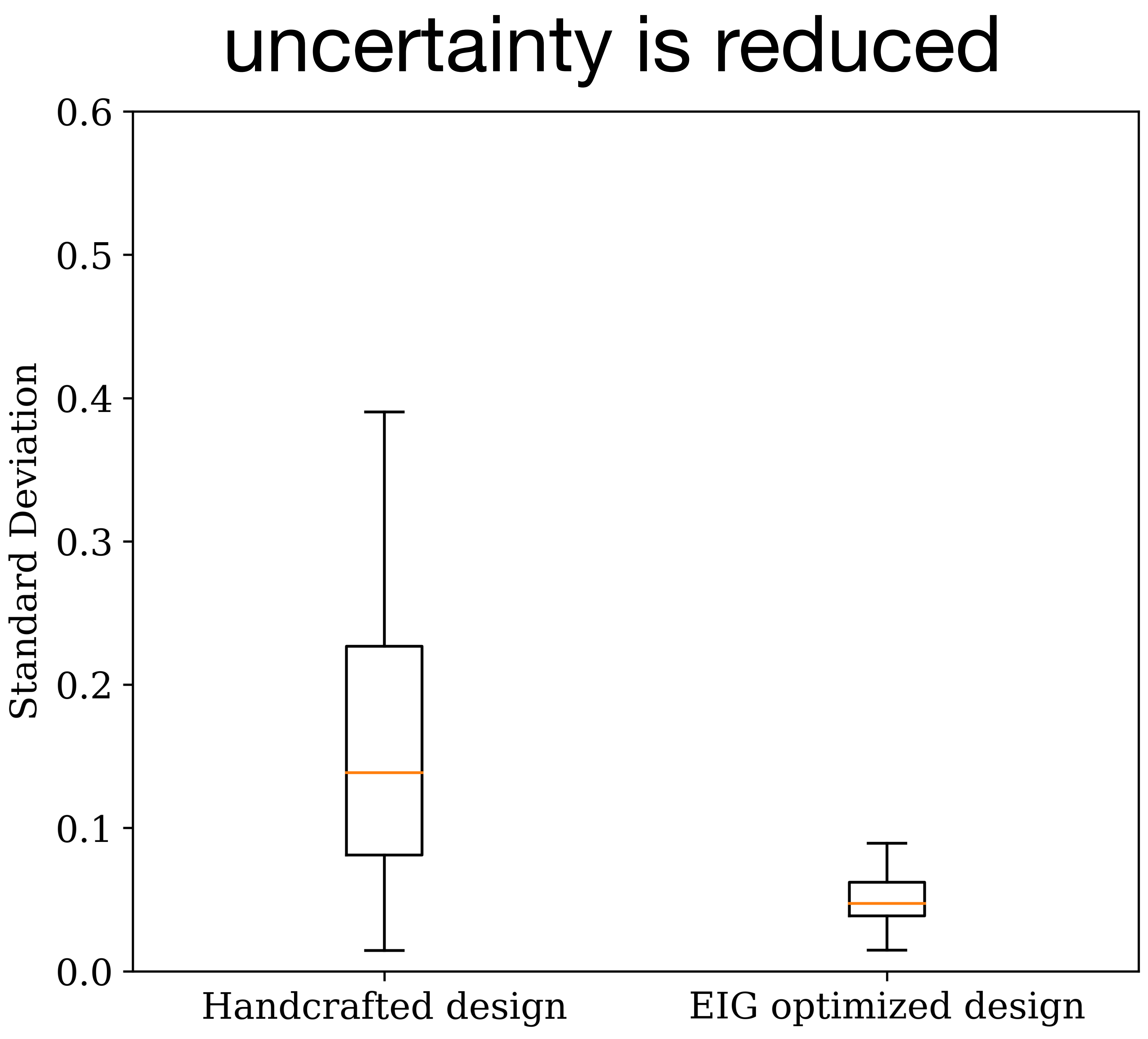
Posterior statistics

Fast sampling w/ normalizing flow to efficiently estimate statistical moments i.e. mean, standard deviation:



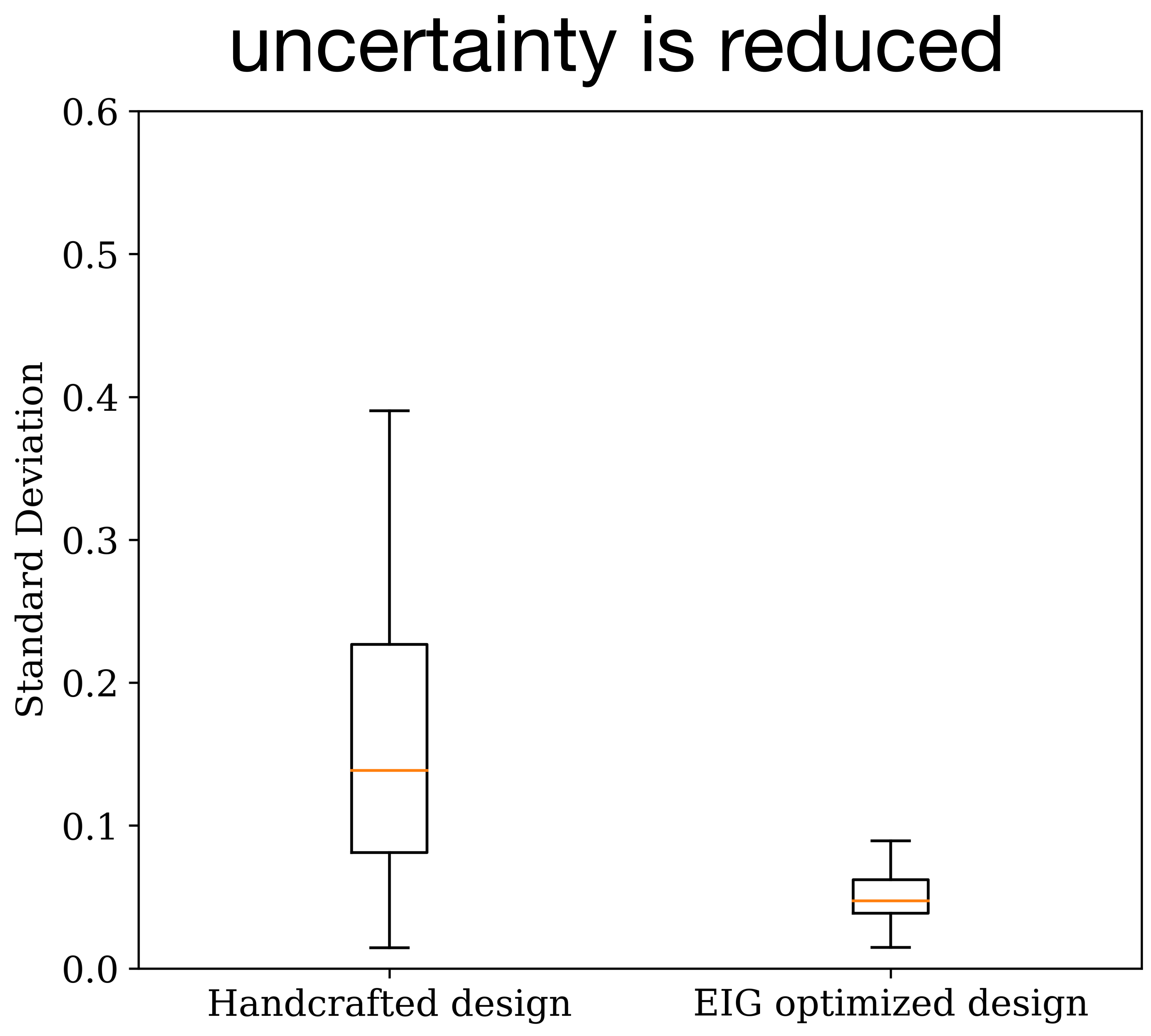
Evaluation on leave-out test set

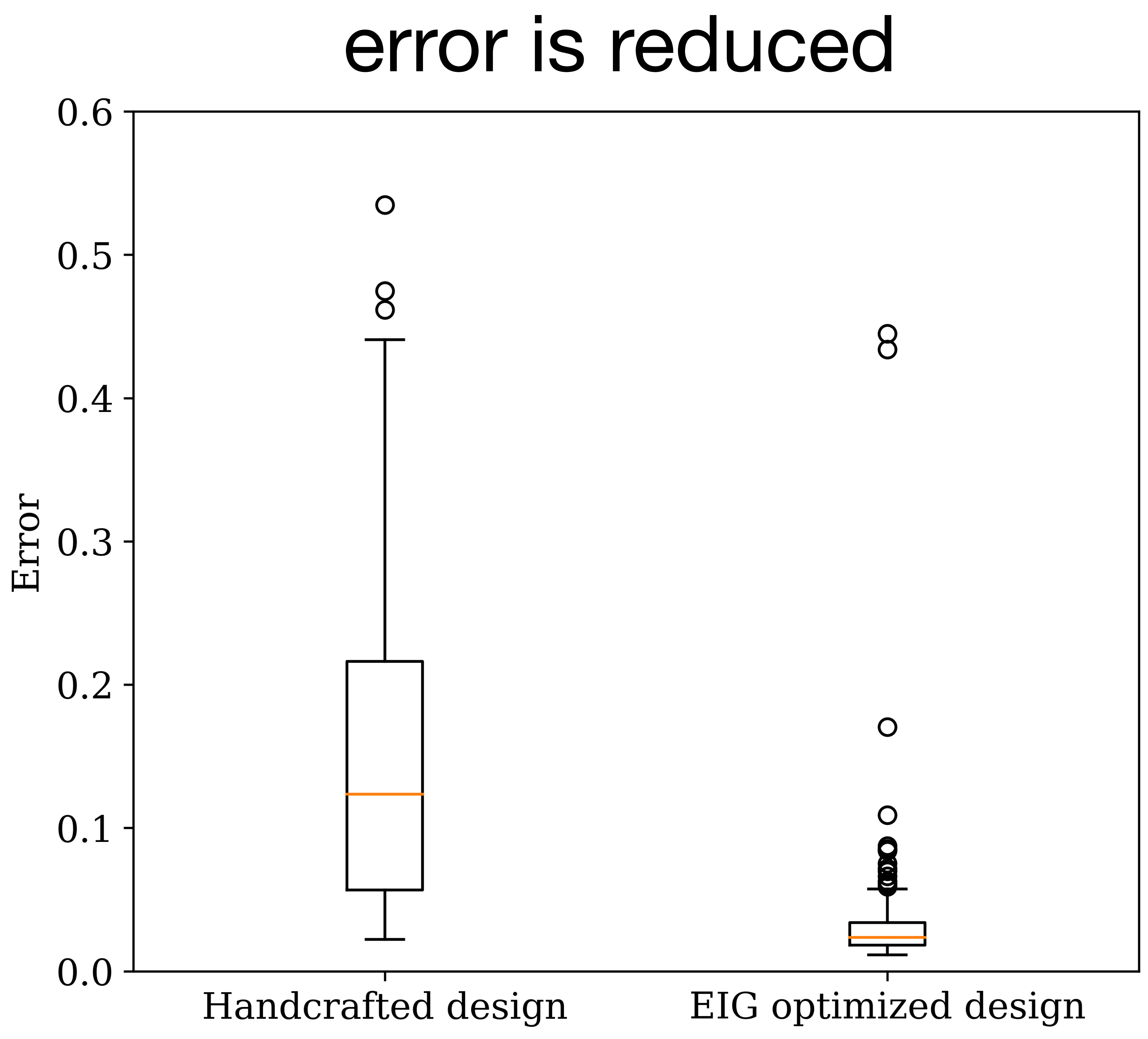
Posterior sampler generalizes to many observations thus can evaluate on many (100) test examples.



Evaluation on leave-out test set

Posterior sampler generalizes to many observations thus can evaluate on many (100) test examples.





Note on scalability

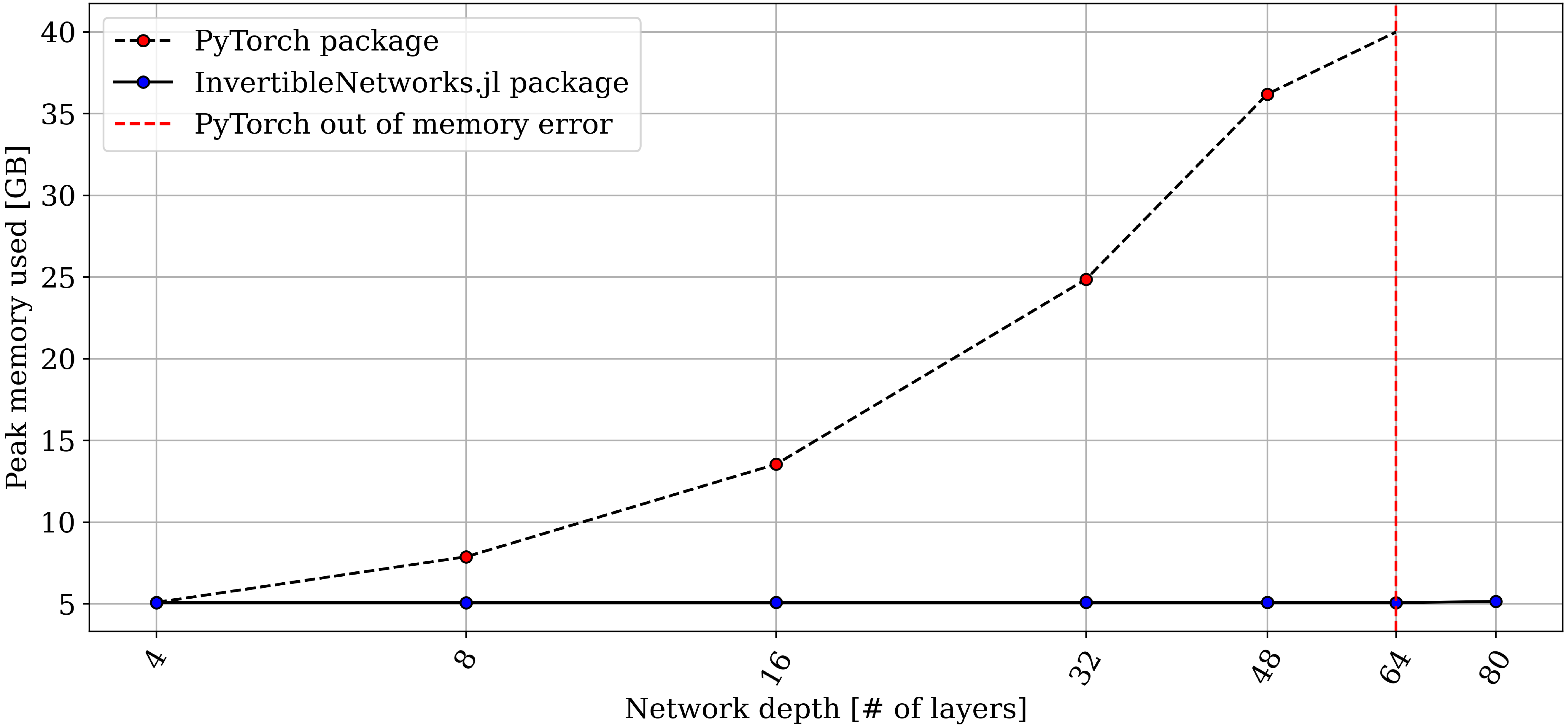
Normalizing flows give you crucial memory efficiency for free…

Orozco, R., Witte, P., Louboutin, M., Siahkoohi, A., Rizzuti, G., Peters, B., & Herrmann, F. J. (2023). InvertibleNetworks. jl: A Julia package for scalable normalizing flows. *arXiv:2312.13480*.

Note on scalability

Normalizing flows give you crucial memory efficiency for free…

if you actually take advantage of it.

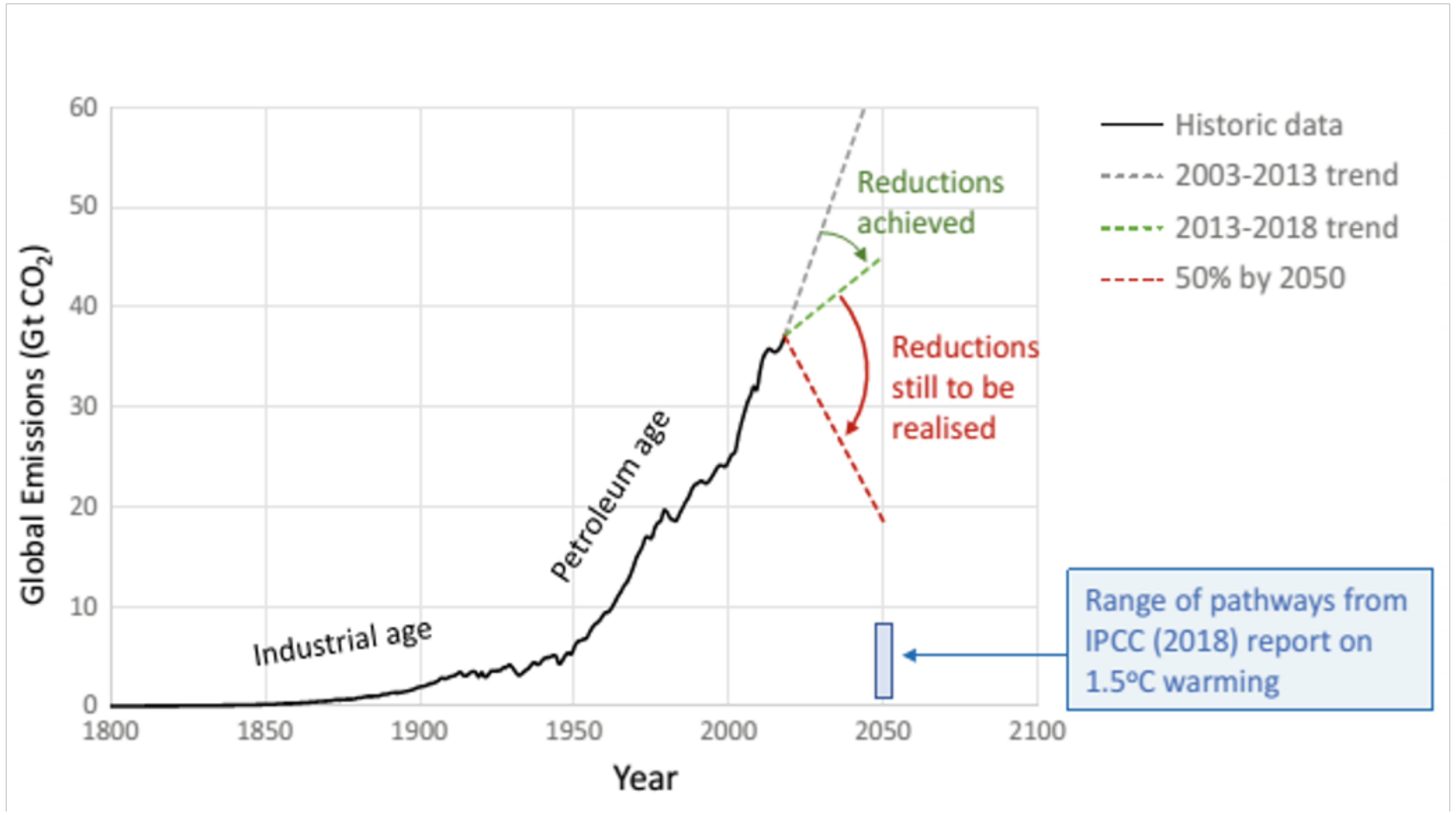


Orozco, R., Witte, P., Louboutin, M., Siahkoohi, A., Rizzuti, G., Peters, B., & Herrmann, F. J. (2023). InvertibleNetworks. jl: A Julia package for scalable normalizing flows. *arXiv:2312.13480*.

Application: monitoring carbon dioxide for mitigating climate change

State and future of climate change

Forecasts say it is not enough to reduce CO_2 emissions



we need to have negative CO_2 emissions i.e. take out CO_2 already in atmosphere…

but where do we store it?

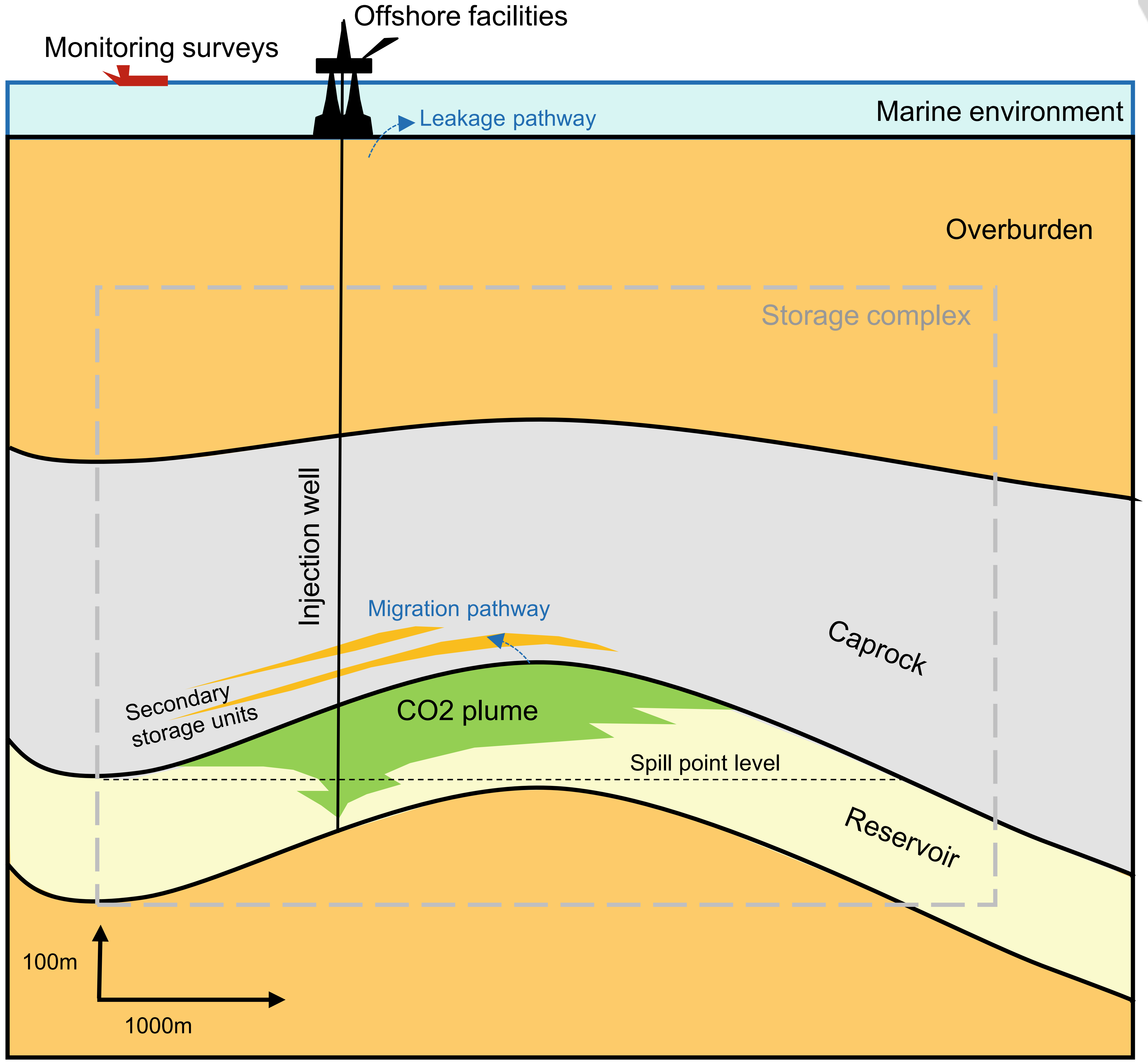
Ringrose, P. 2020. How to store CO_2 underground: Insights from early-mover CCS Projects, volume 129. Springer.

Underground carbon dioxide storage

Demonstrated solution for large scale storage

- subsurface structures create natural barriers
- long term solution - CO_2 chemically seals into rock at geological time scales

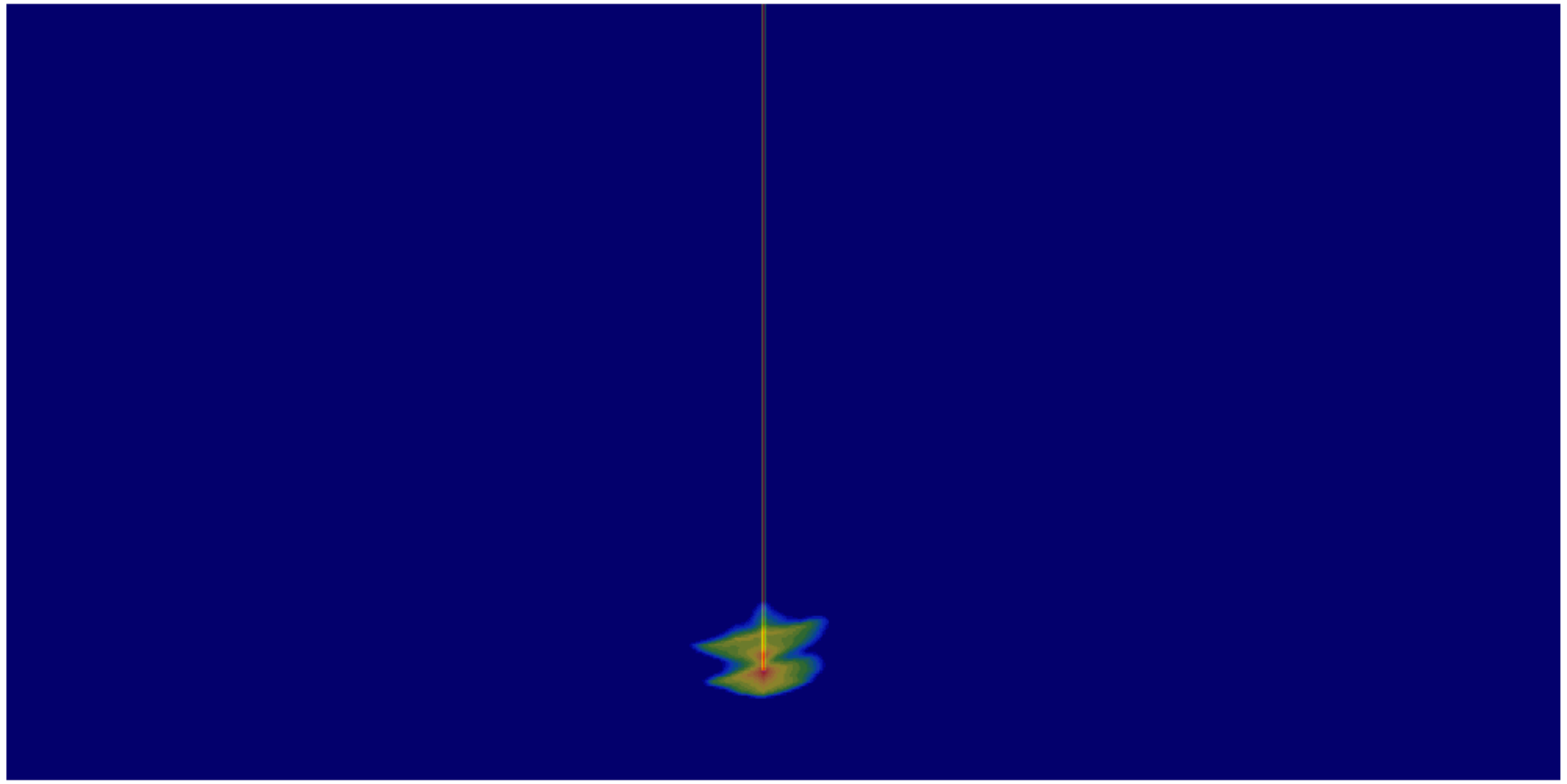
but the plume is not stationary…



Ringrose, P. 2020. How to store CO_2 underground: Insights from early-mover CCS Projects, volume 129. Springer.
Jun, Y.-S.; Zhang, L.; Min, Y.; and Li, Q. 2017. Nanoscale Chemical Processes Affecting Storage Capacities and Seals during Geologic CO_2 Sequestration. Accounts of Chemical Research, 50(7): 1521–1529. PMID: 28686035.

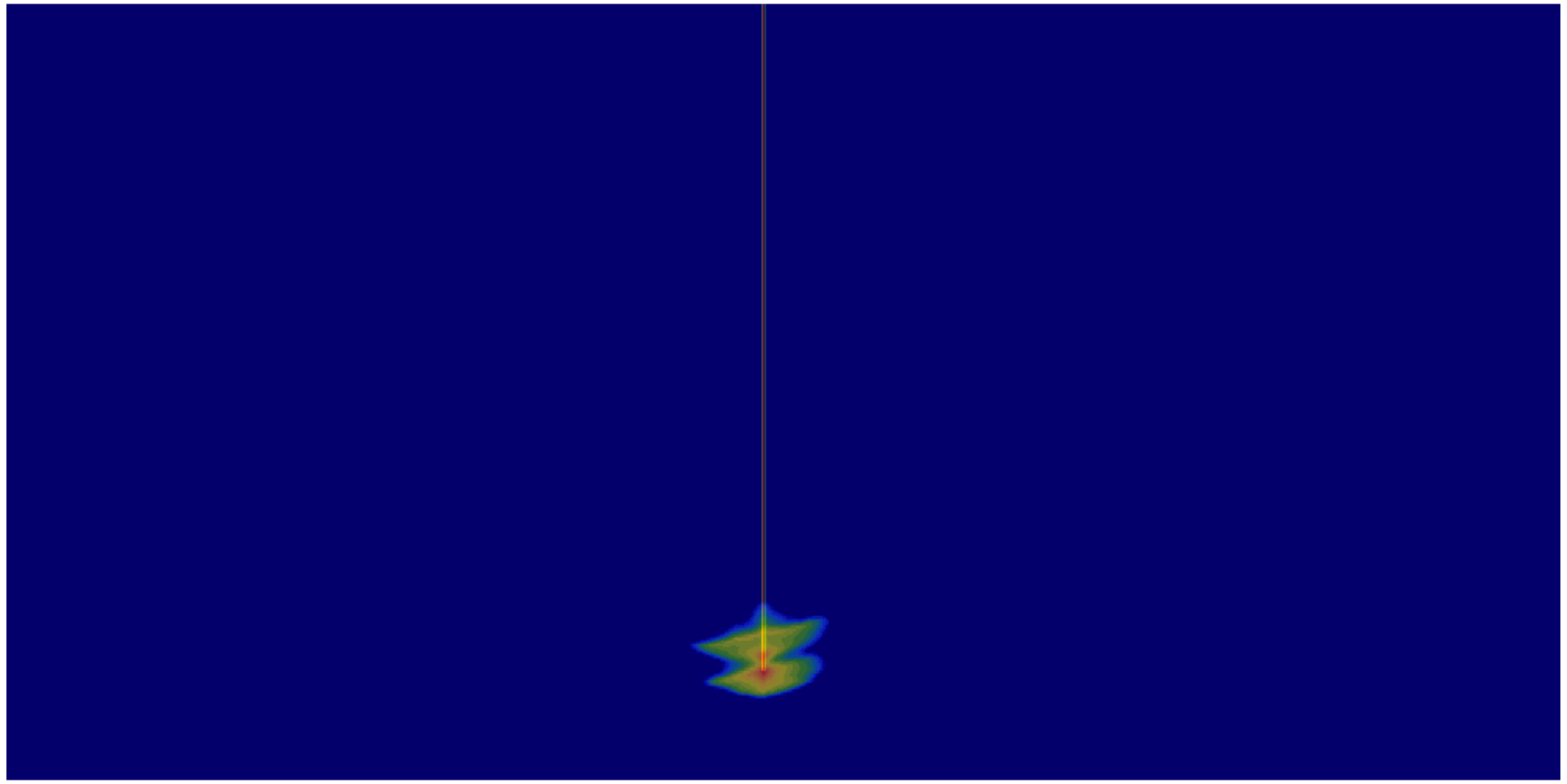
Carbon dioxide monitoring

CO_2 plume evolves over time due to injection and permeability effects



Carbon dioxide monitoring

CO_2 plume evolves over time due to injection and permeability effects



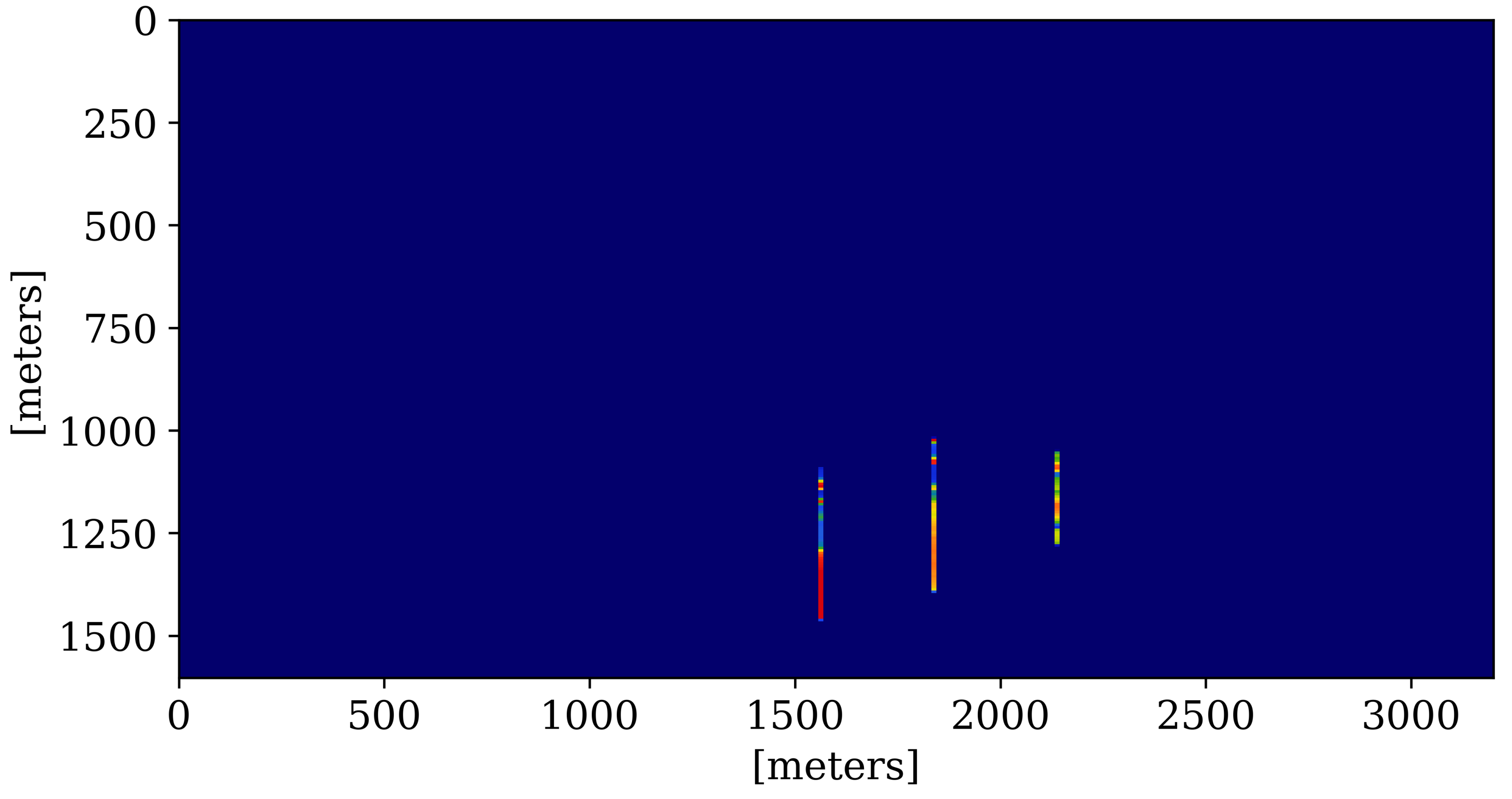
thus monitoring plume is important to:

- prevent leakage
- avoid “seismic events”
- stay in licensed area.

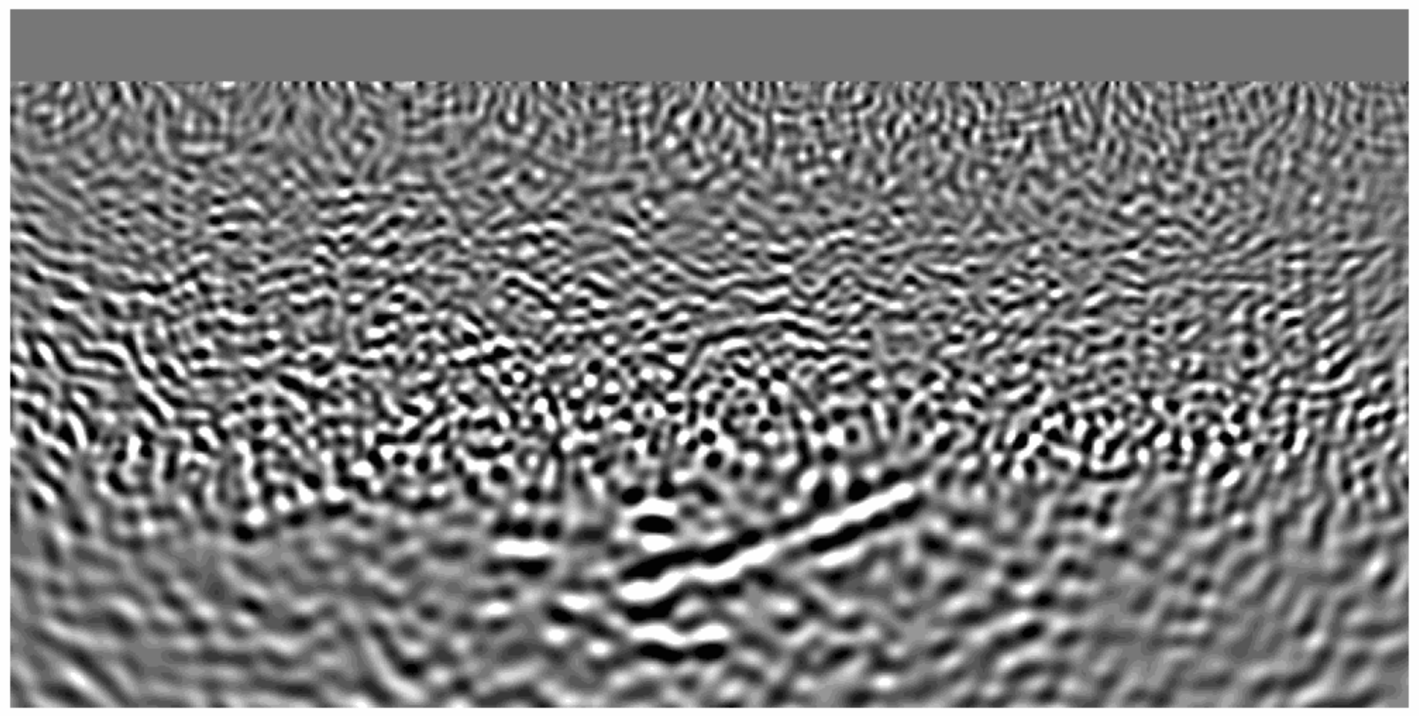
Carbon dioxide monitoring

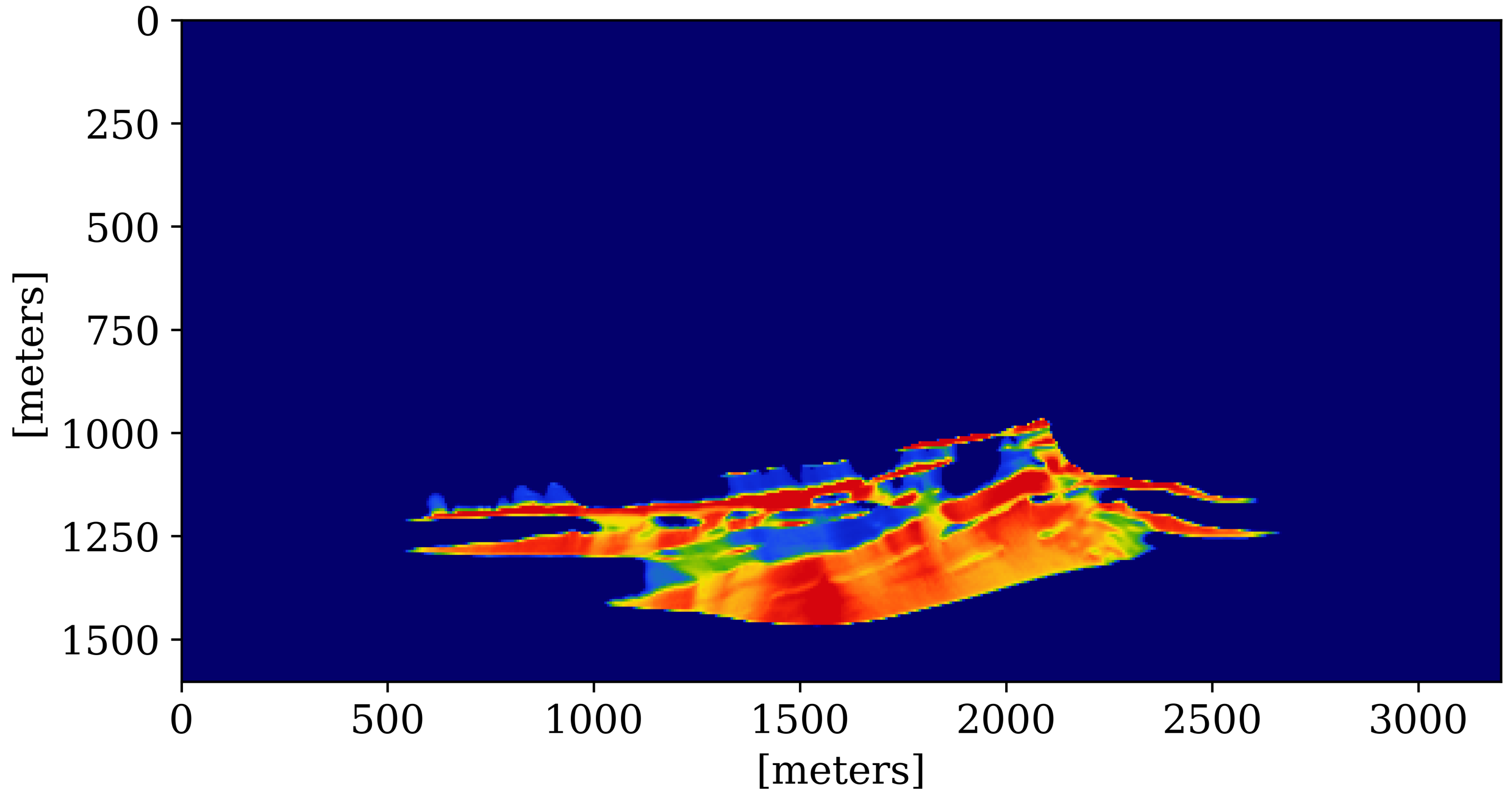
Two types of time-lapse CO_2 plume observations

- ► direct but local – borehole wells



- ► indirect but global – seismic





Optimal well locations

CO_2 project lasts years thus can drill more wells but:

- many location options
- expensive (1 million dollars - 100 million dollars)

Optimal well locations

CO_2 project lasts years thus can drill more wells but:

- ► many location options
- ► expensive (1 million dollars - 100 million dollars)

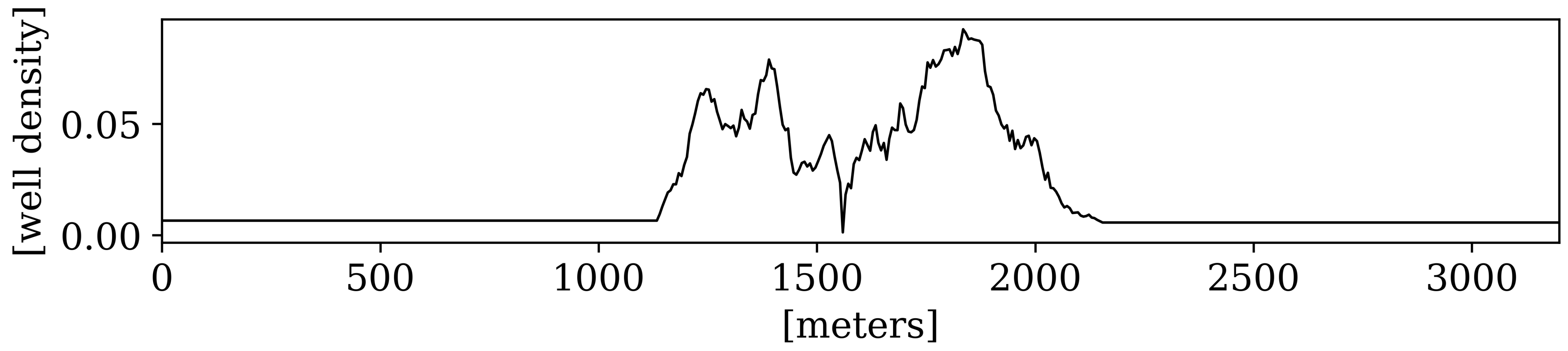
Operators deciding well locations should be informed by

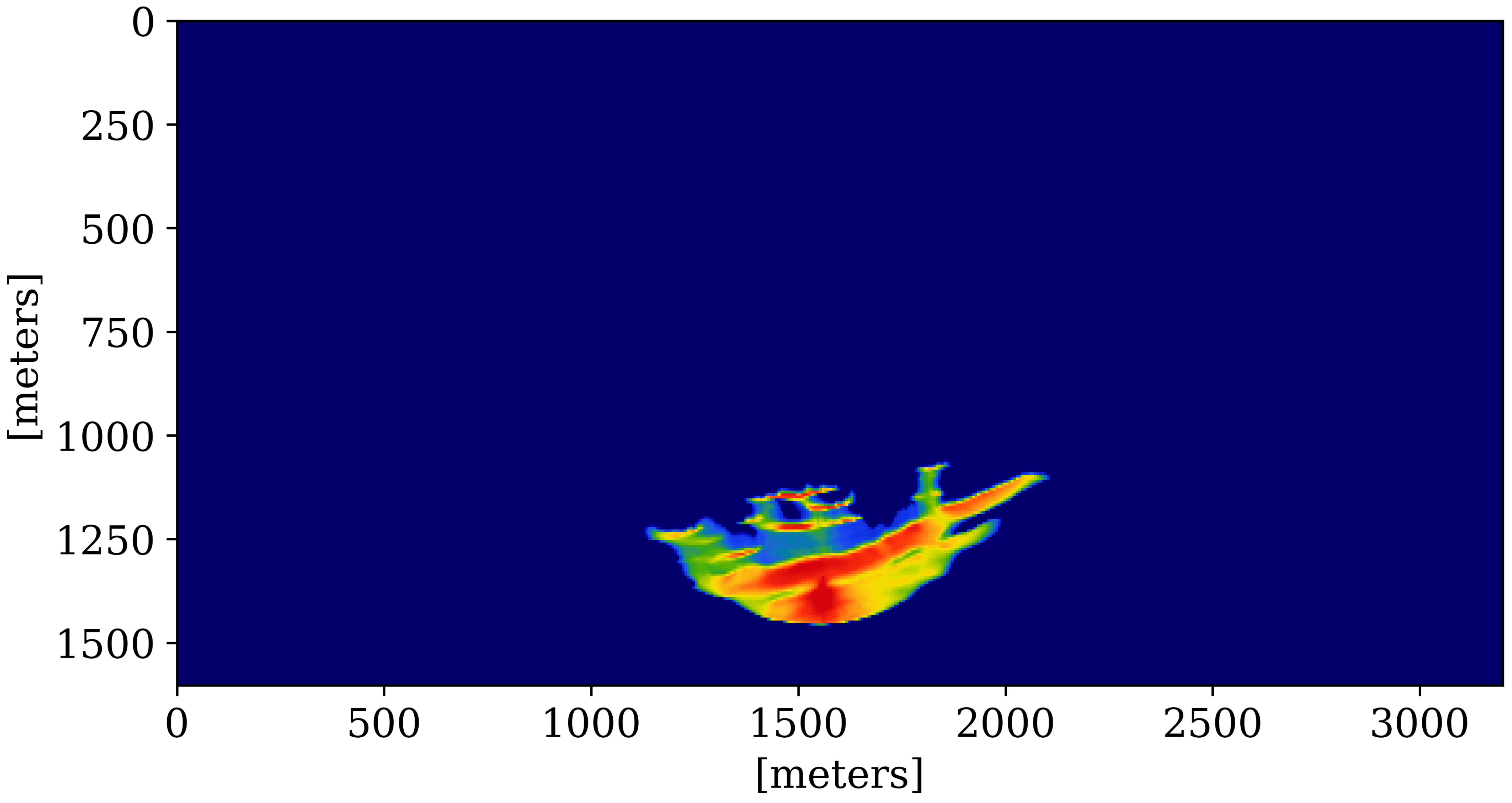
- ► current knowledge of the CO_2 plumes (prior)
- ► physics simulations of plume forecasts (likelihood)

Optimal well locations

Optimize for probability *density* of well placement

- well budget agnostic
  - decide number of wells post-hoc
- easier optimization
  - stochastic sampling during training avoids local minima





Small module in full-stack digital twin

MS189
Uncertainty Quantification for Digital Twins - Part III of III

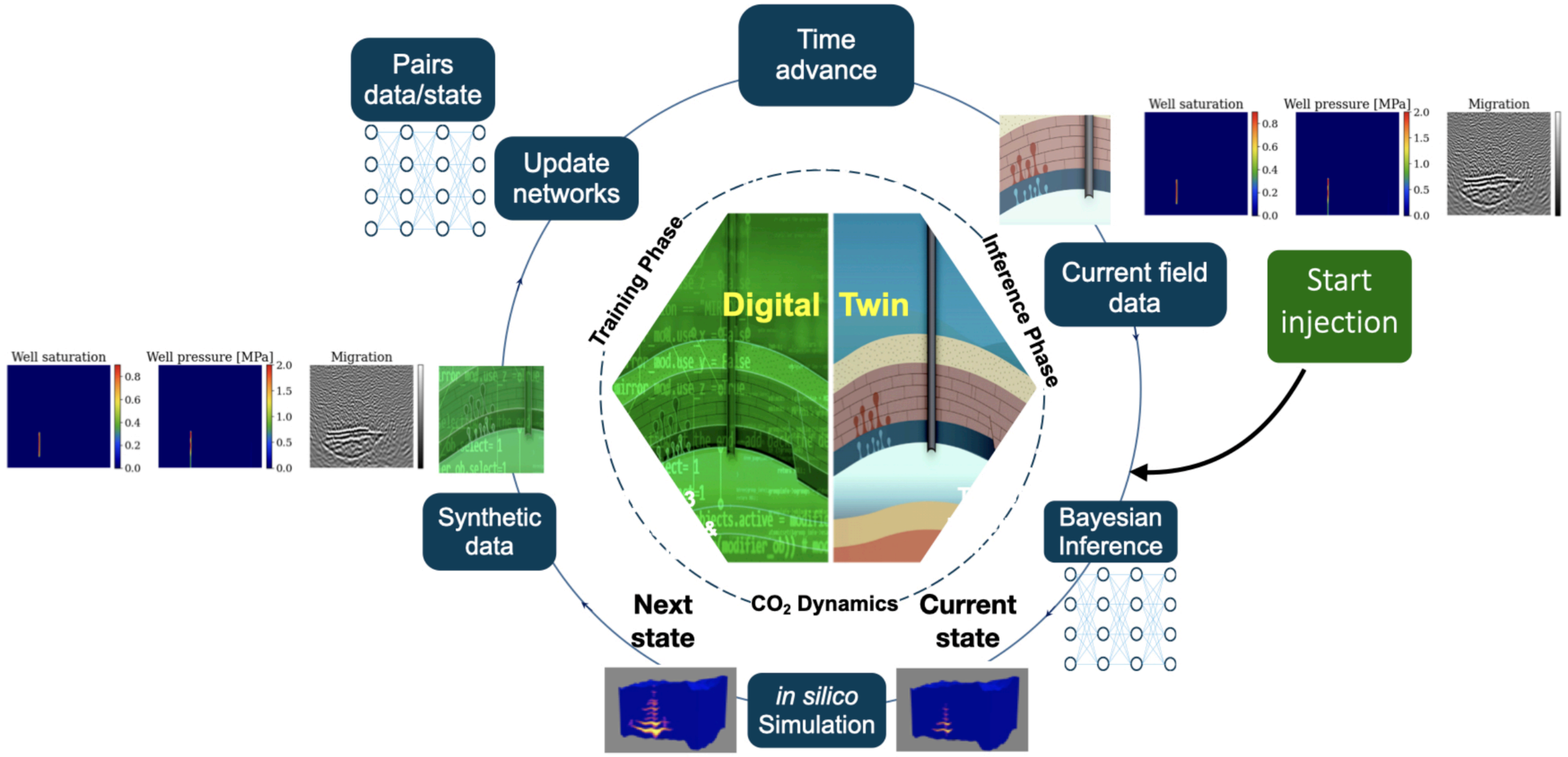
8:30 AM - 10:30 AM
Room: San Giusto - Hotel Savoia Excelsior Palace

For Part II, see MS169

A digital twin (DT) is a computational system that continuously and repeatedly assimilates obser
otherwise guides decisions, using predictions from the updated model. Often DTs are employed f
to models to decisions. The resulting data assimilation and optimal control/decision problems mu
tractable for large-scale complex systems. This minisymposium addresses mathematical, statistic
control, and optimal experimental design subproblems, as well as the reduced order models and s

Organizer: Nicole Aretz
University of Texas at Austin, U.S.
Omar Ghattas
University of Texas at Austin, U.S.
Youssef M. Marzouk
Massachusetts Institute of Technology, U.S.

8:30-8:55 An Uncertainty-Aware Digital Twin for Geological Carbon Storage abstract
Felix Herrmann and Abhinav Gahlot, Georgia Institute of Technology, U.S.



CO2 storage project life cycle

Prior samples $p(\mathbf{x}_t)$

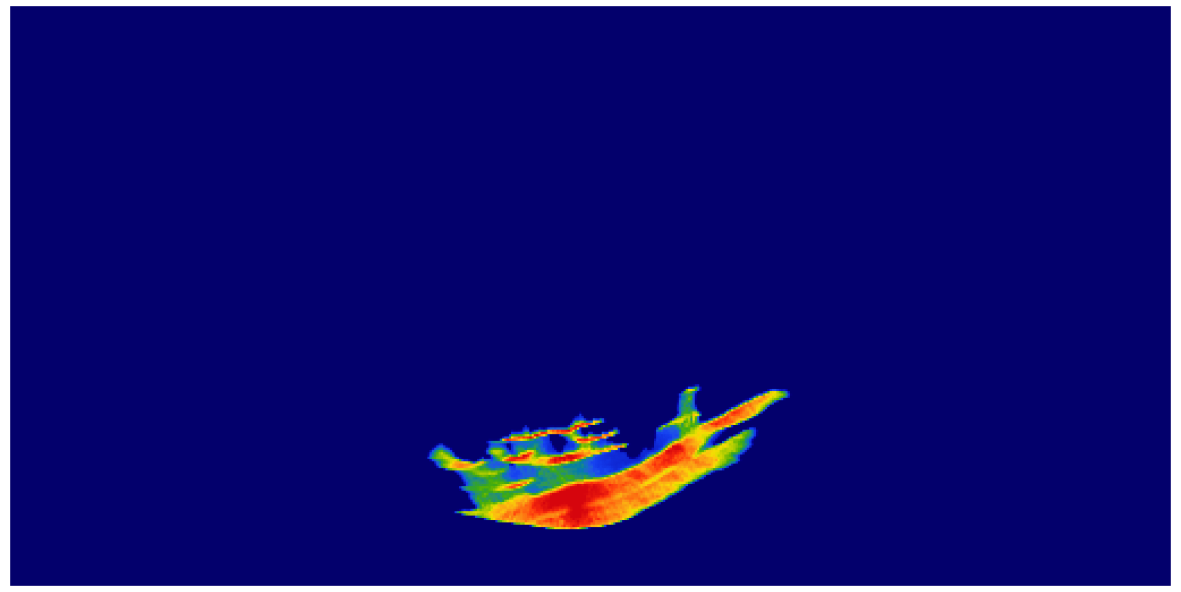
Fluid flow simulations

CO2 storage project life cycle

Prior samples $p(\mathbf{x}_t)$

Fluid flow simulations

Forecasted plumes $p(\mathbf{x}_{t+1} \mid \mathbf{x}_t)$

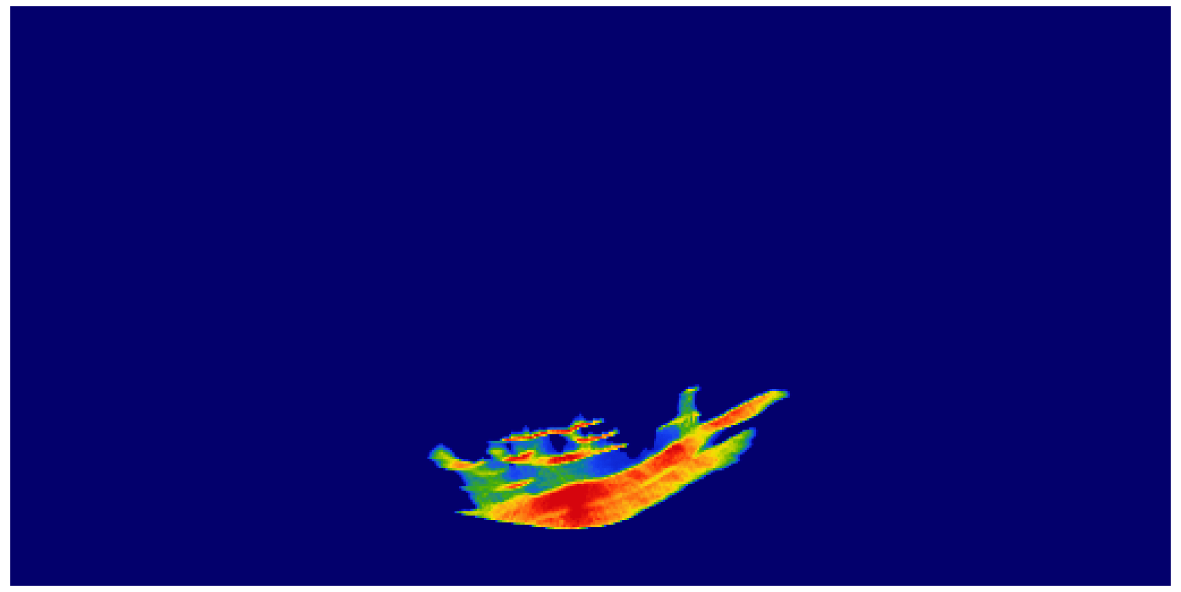


CO_2 storage project life cycle

Prior samples $p(\mathbf{x}_t)$

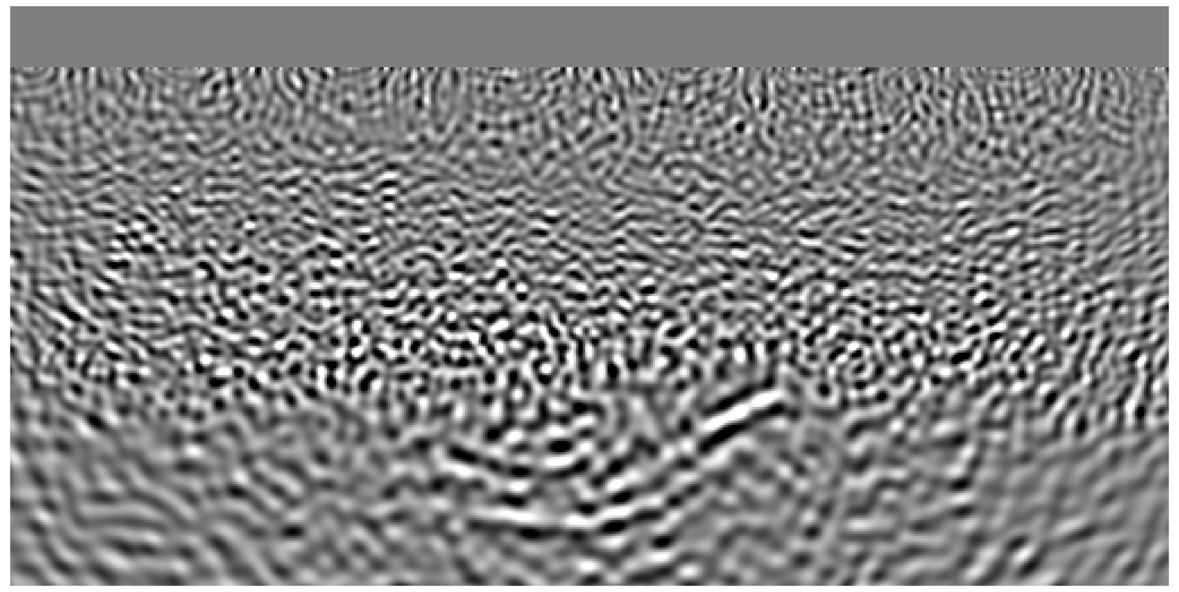
Fluid flow simulations →

Forecasted plumes $p(\mathbf{x}_{t+1} \mid \mathbf{x}_t)$



Synthetic observations →

Train inference network and well design using pairs $p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})$

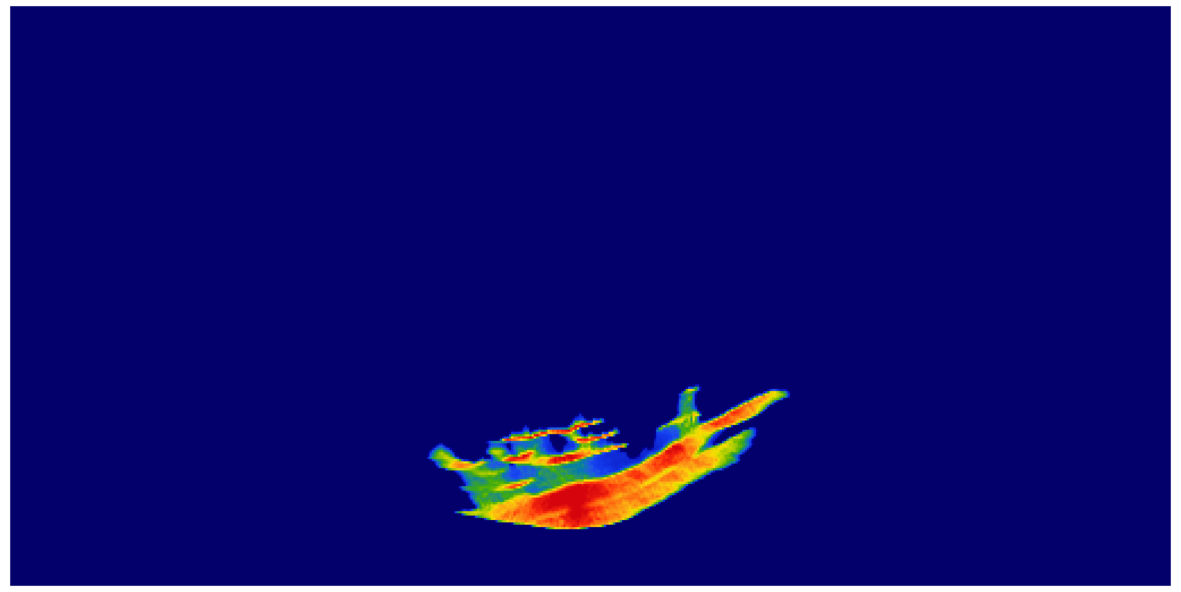


CO2 storage project life cycle

Prior samples $p(\mathbf{x}_t)$

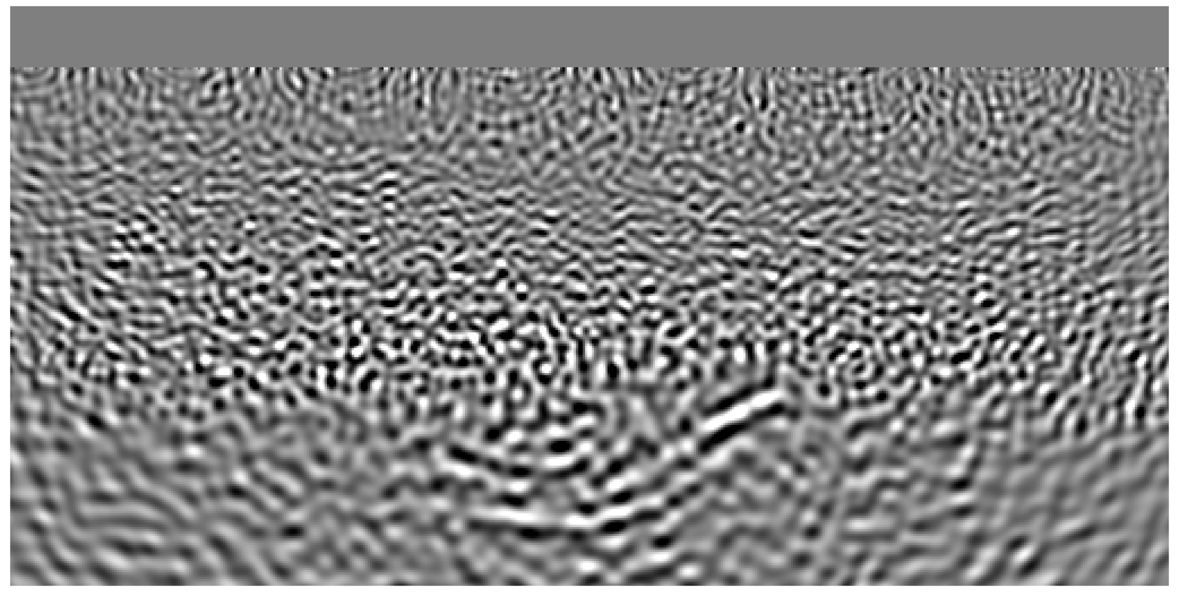
Fluid flow simulations →

Forecasted plumes $p(\mathbf{x}_{t+1} \mid \mathbf{x}_t)$



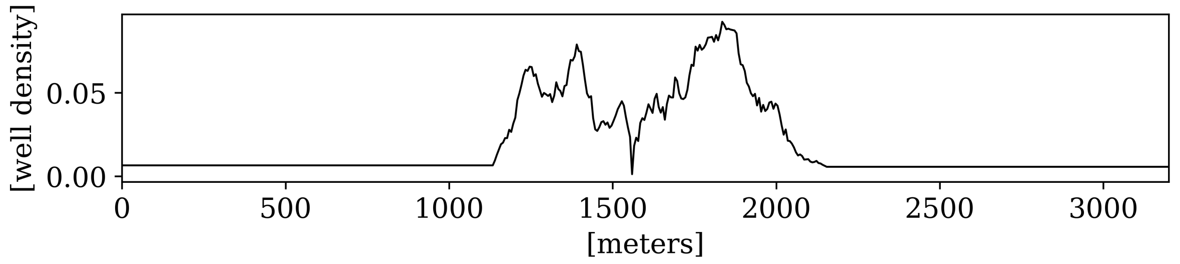
Synthetic observations →

Train inference network and well design using pairs $p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})$



↓

Outputs: posterior sampler $p_{\hat{\theta}}(\mathbf{x}_{t+1} \mid \mathbf{y}_{t+1})$ and optimal well density

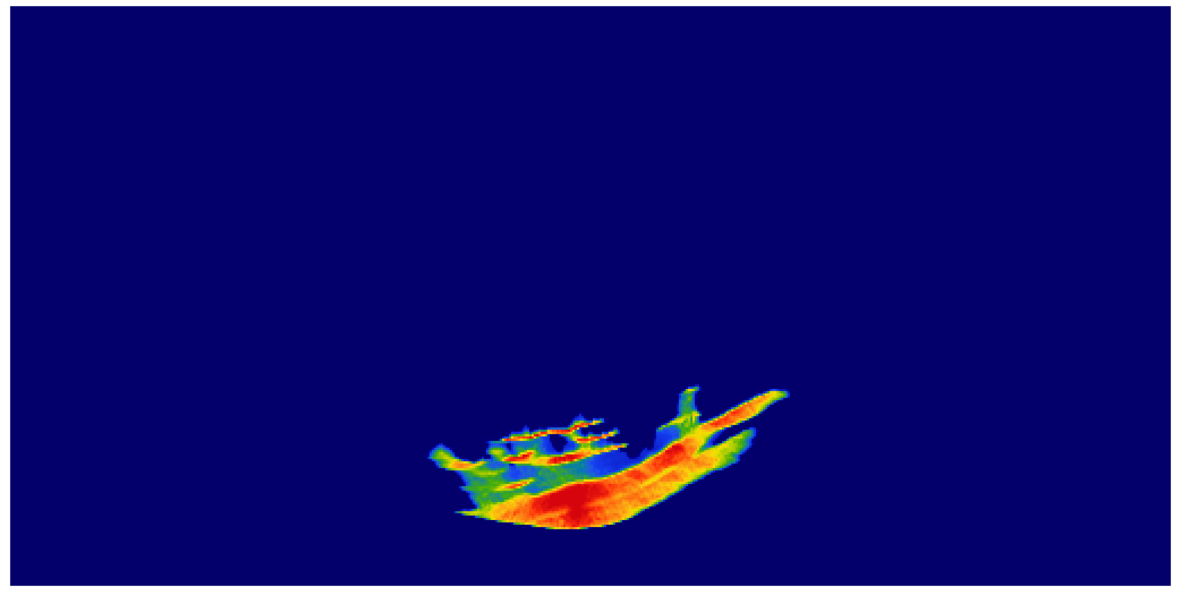


CO2 storage project life cycle

Prior samples $p(\mathbf{x}_t)$

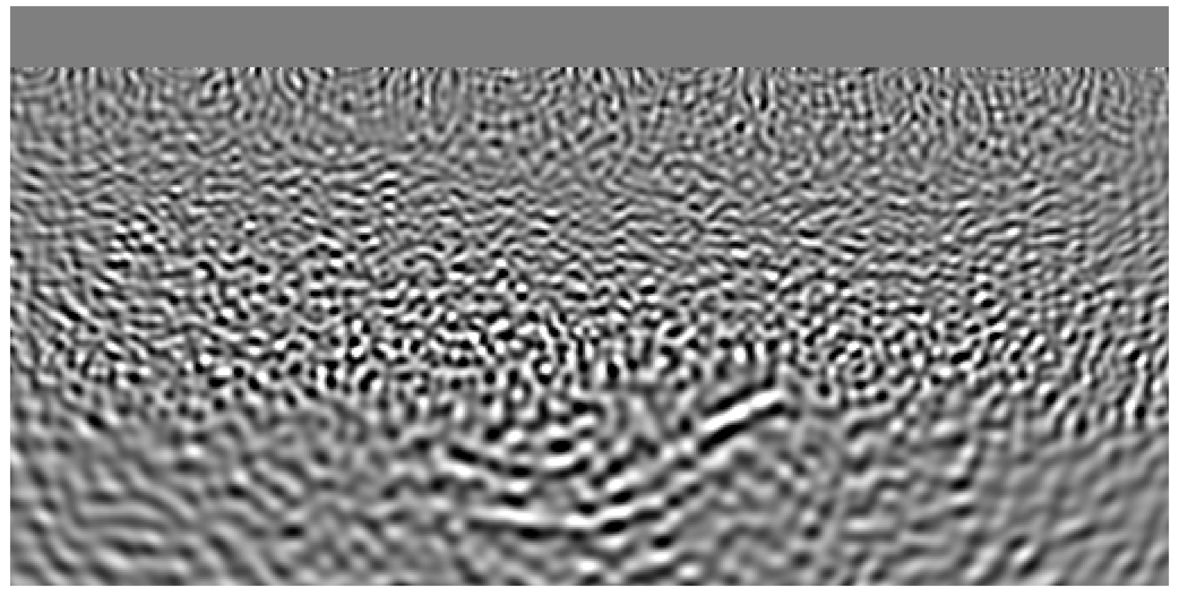
Fluid flow simulations

Forecasted plumes $p(\mathbf{x}_{t+1} \mid \mathbf{x}_t)$

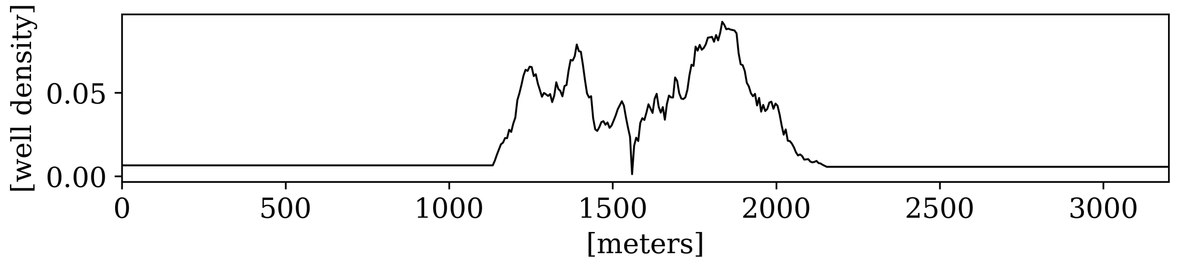


Synthetic observations

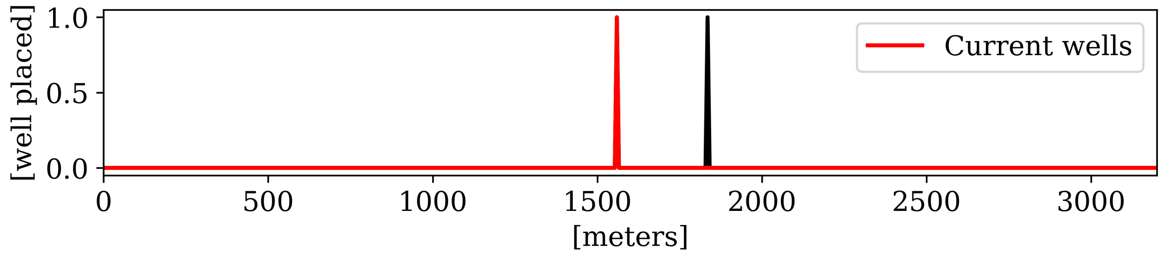
Train inference network and well design using pairs $p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})$



Outputs: posterior sampler $p_{\hat{\theta}}(\mathbf{x}_{t+1} \mid \mathbf{y}_{t+1})$ and optimal well density



Drill well using optimal well density

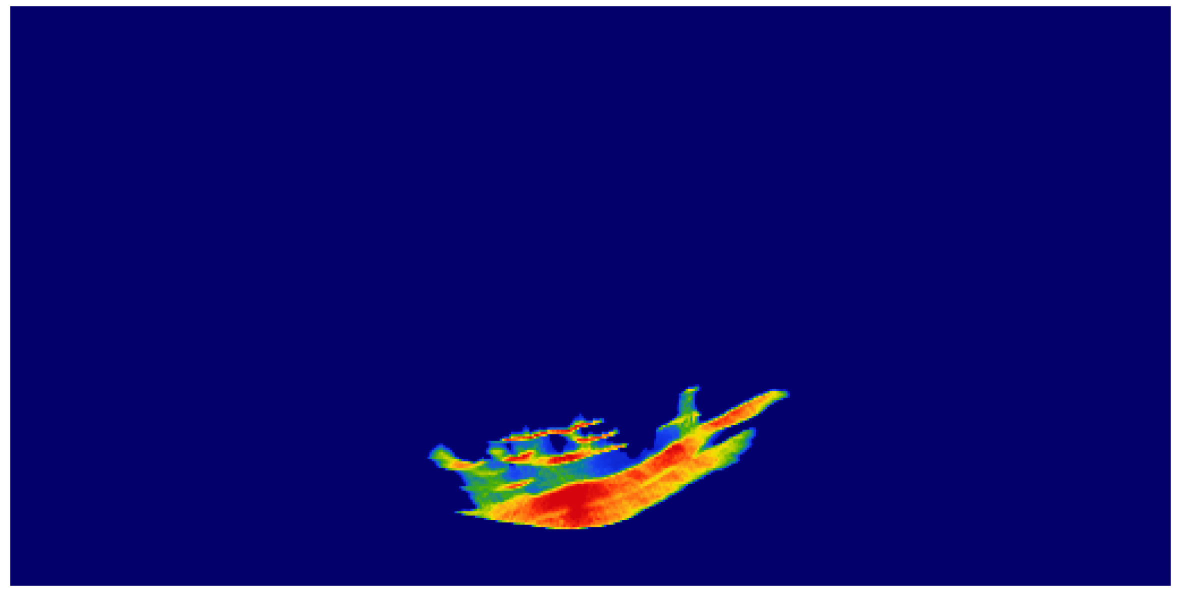


CO2 storage project life cycle

Prior samples $p(\mathbf{x}_t)$

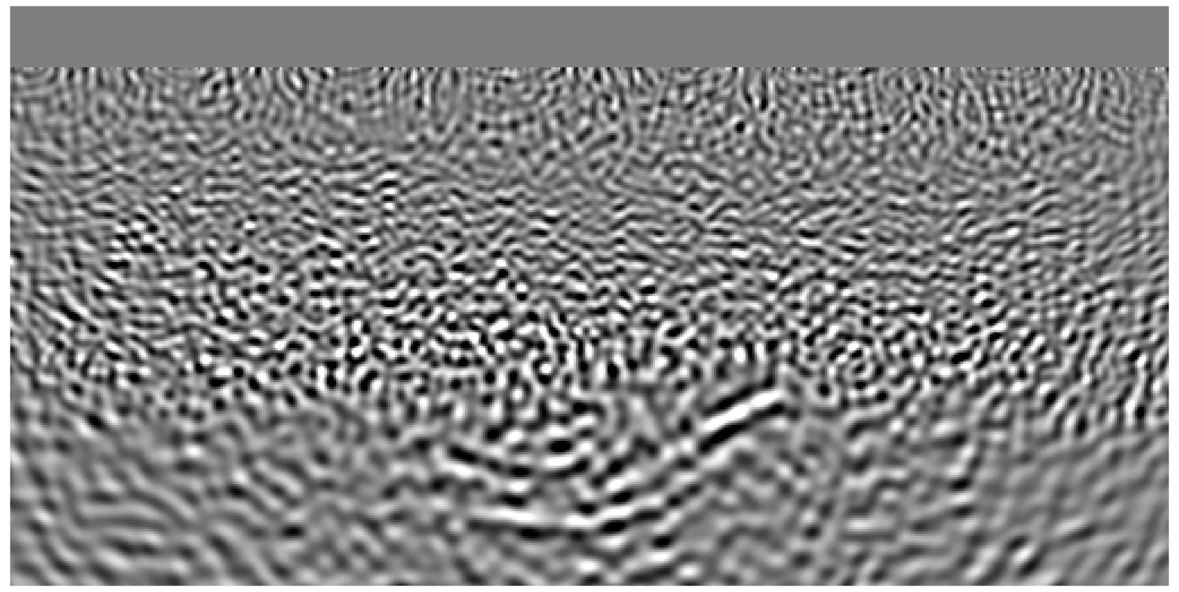
Fluid flow simulations

Forecasted plumes $p(\mathbf{x}_{t+1} \mid \mathbf{x}_t)$

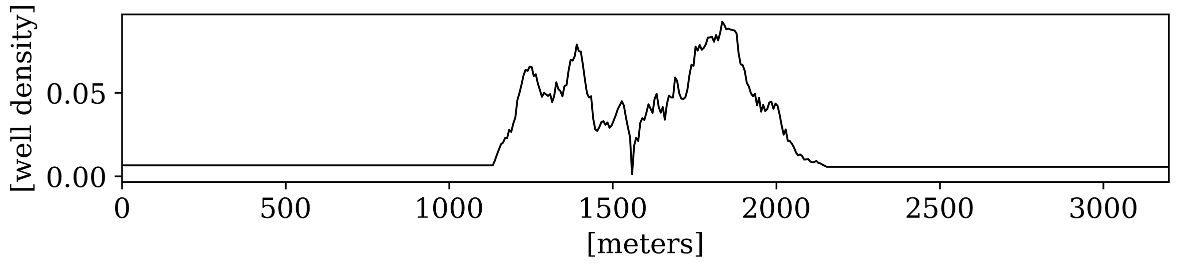


Synthetic observations

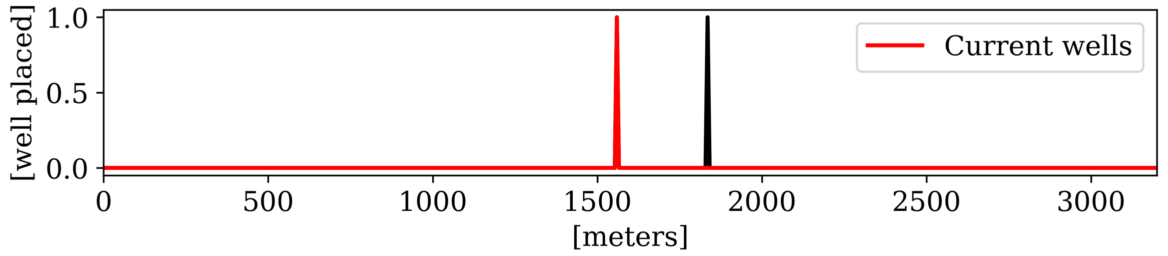
Train inference network and well design using pairs $p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})$



Outputs: posterior sampler $p_{\hat{\theta}}(\mathbf{x}_{t+1} \mid \mathbf{y}_{t+1})$ and optimal well density

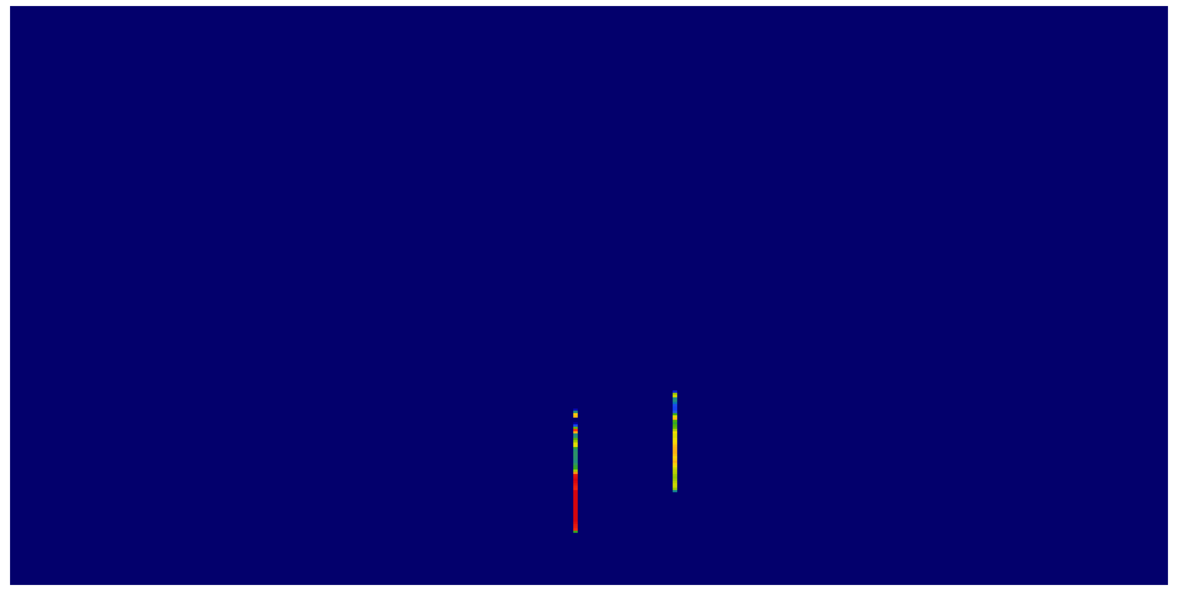


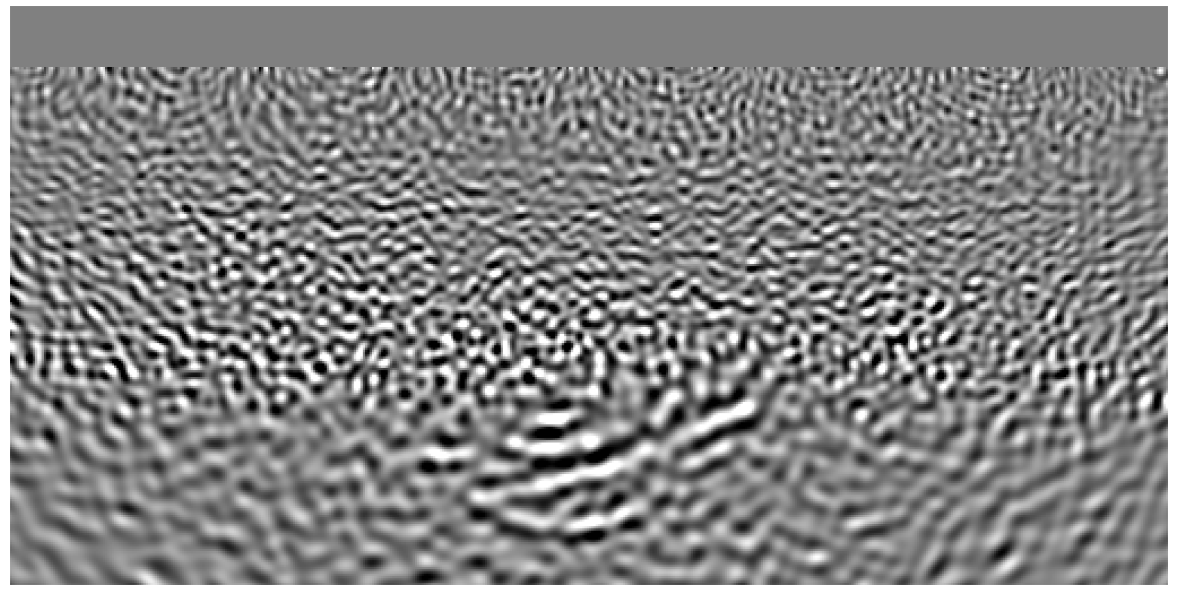
Drill well using optimal well density



Field observation

Collect field data $\mathbf{y}_{t+1}^{obs}$ w/ optimal well



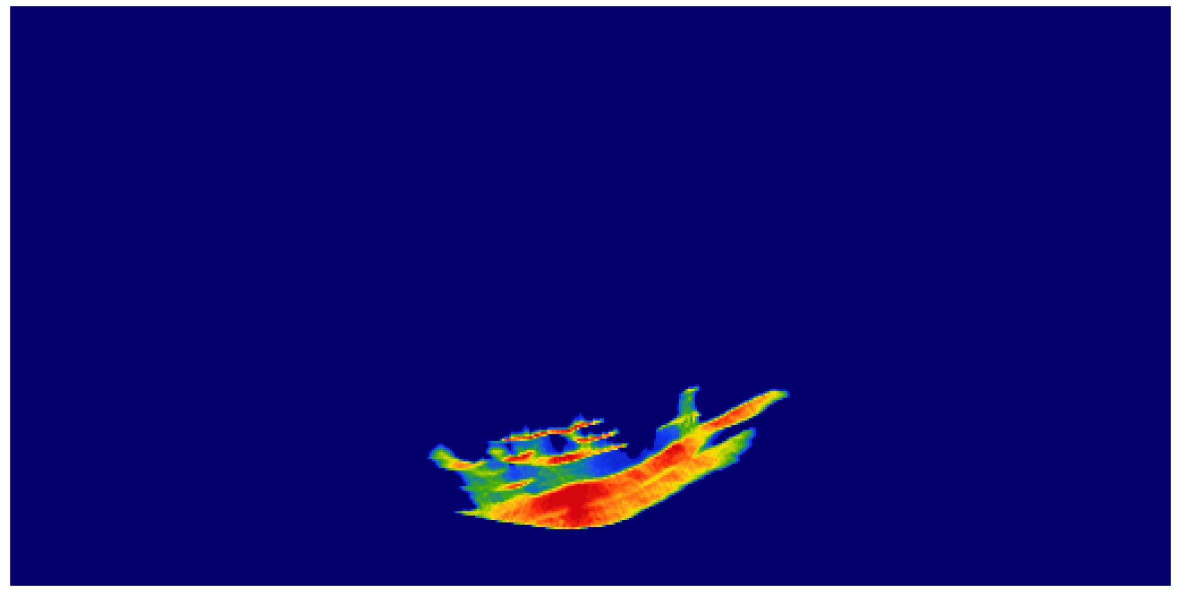


CO2 storage project life cycle

Prior samples $p(\mathbf{x}_t)$

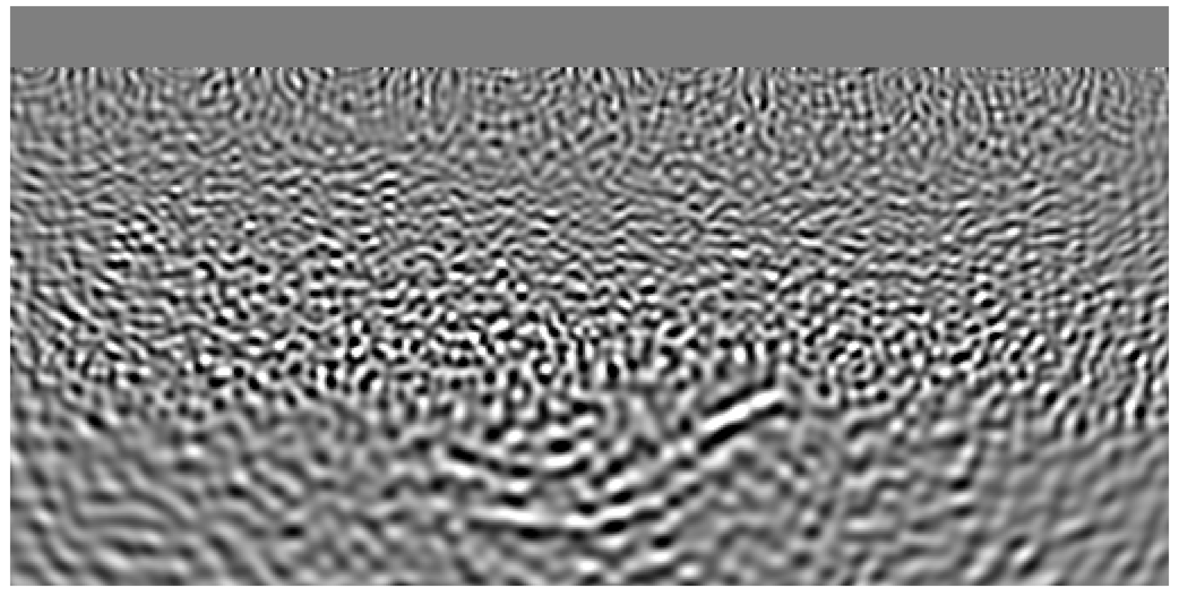
Fluid flow simulations

Forecasted plumes $p(\mathbf{x}_{t+1} \mid \mathbf{x}_t)$

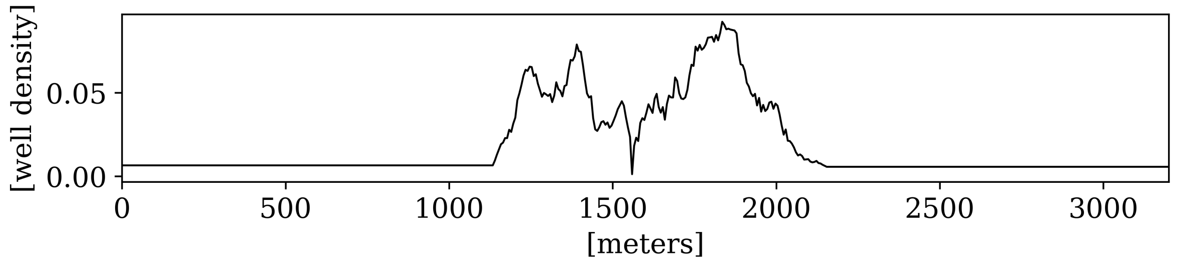


Synthetic observations

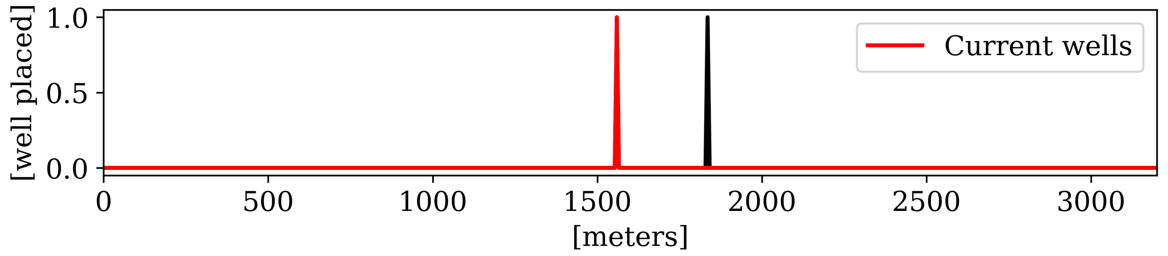
Train inference network and well design using pairs $p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})$



Outputs: posterior sampler $p_{\hat{\theta}}(\mathbf{x}_{t+1} \mid \mathbf{y}_{t+1})$ and optimal well density

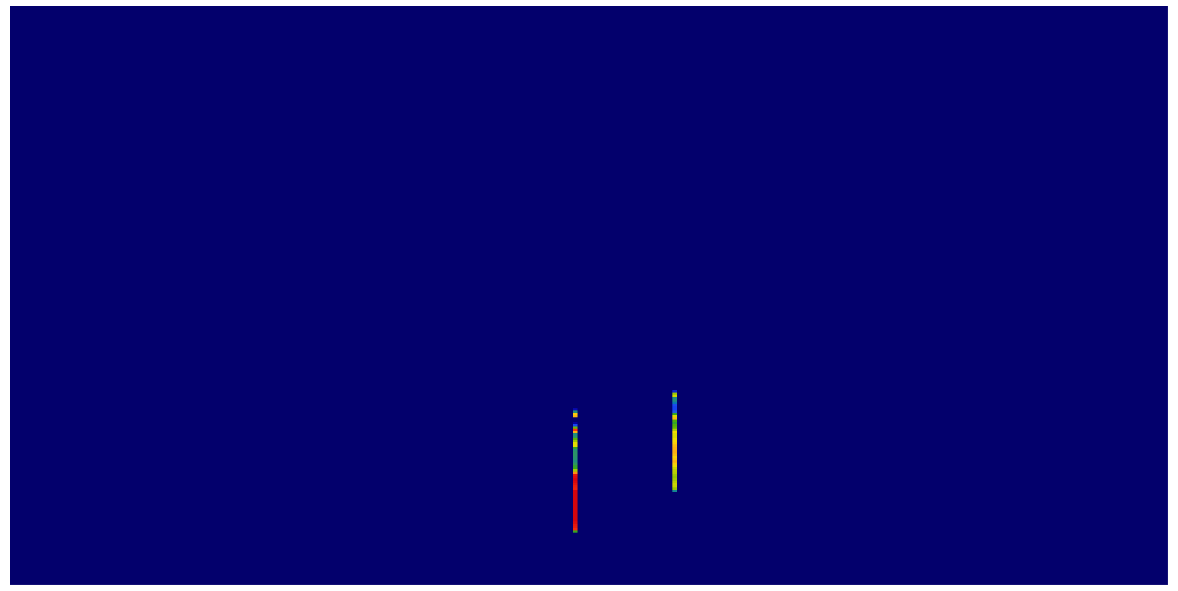


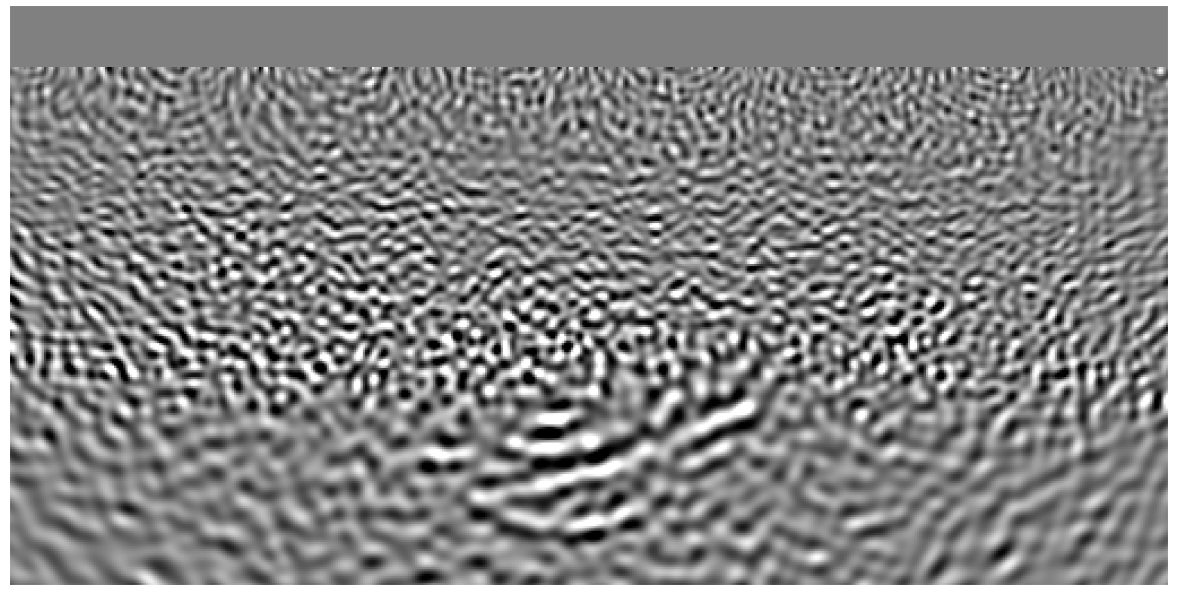
Drill well using optimal well density



Field observation

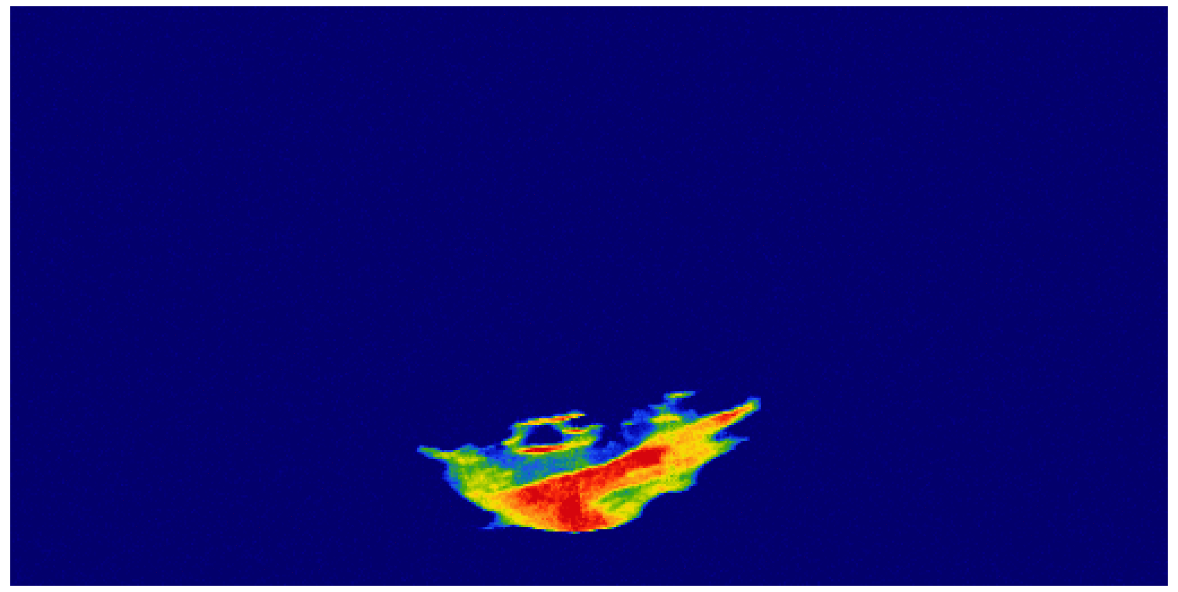
Collect field data $\mathbf{y}_{t+1}^{obs}$ w/ optimal well





Posterior inference

Inference from field data $p_{\hat{\theta}}(\mathbf{x}_{t+1} \mid \mathbf{y}_{t+1}^{obs})$

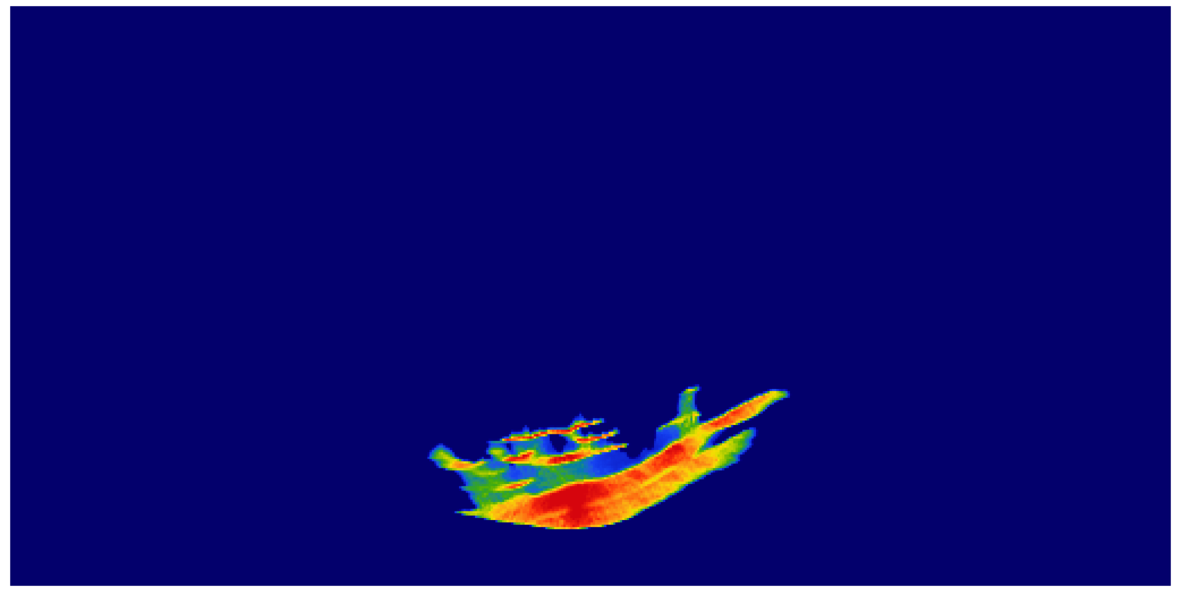


CO2 storage project life cycle

Prior samples $p(\mathbf{x}_t)$

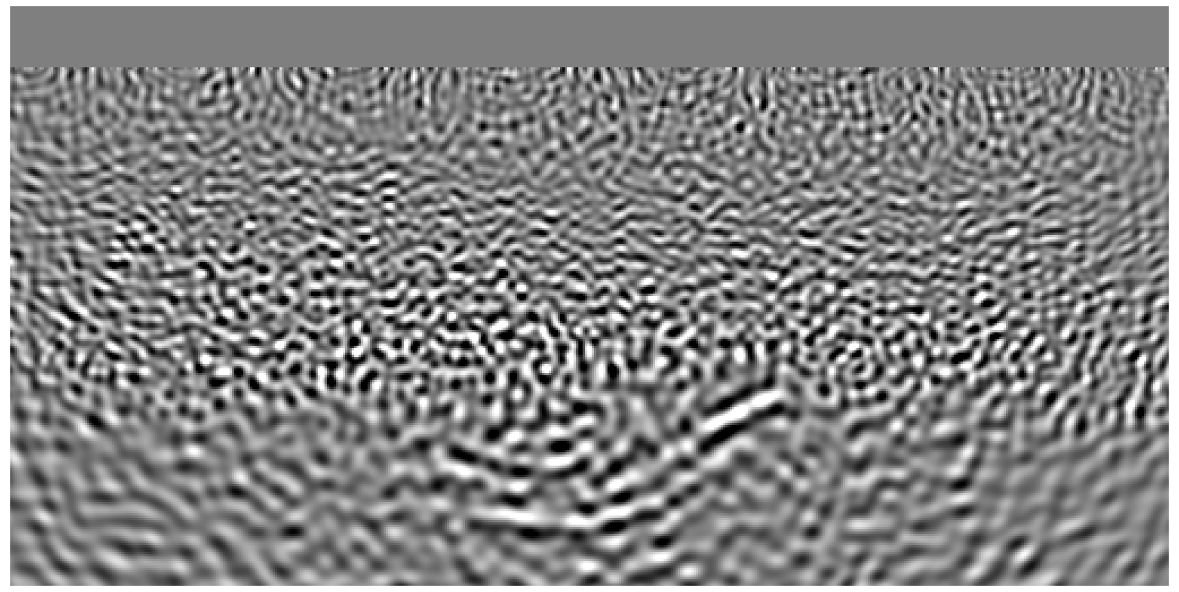
Fluid flow simulations

Forecasted plumes $p(\mathbf{x}_{t+1} \mid \mathbf{x}_t)$

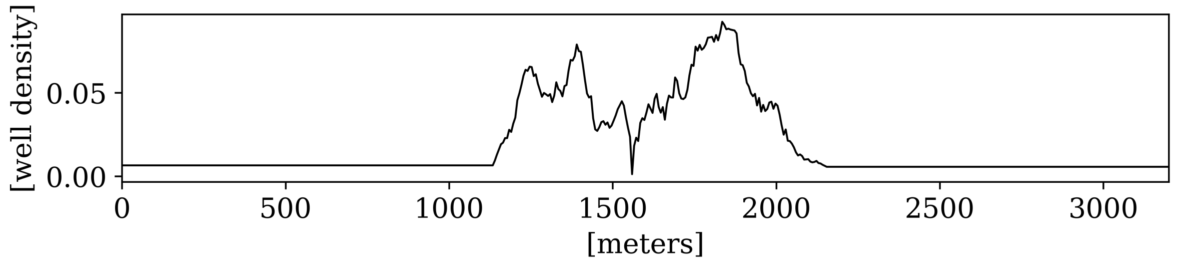


Synthetic observations

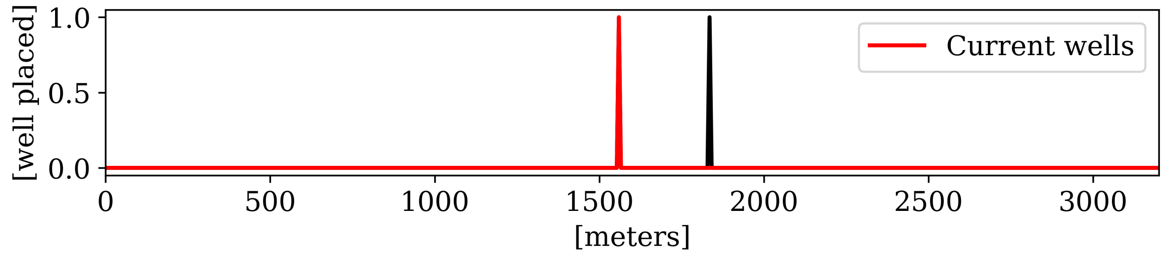
Train inference network and well design using pairs $p(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})$



Outputs: posterior sampler $p_{\hat{\theta}}(\mathbf{x}_{t+1} \mid \mathbf{y}_{t+1})$ and optimal well density

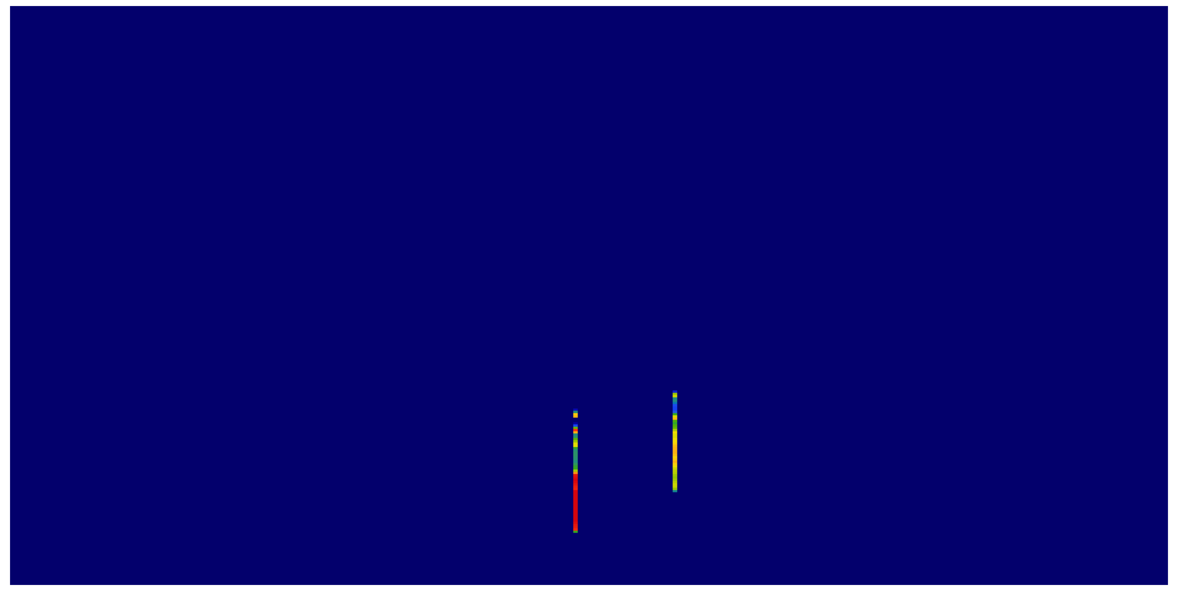


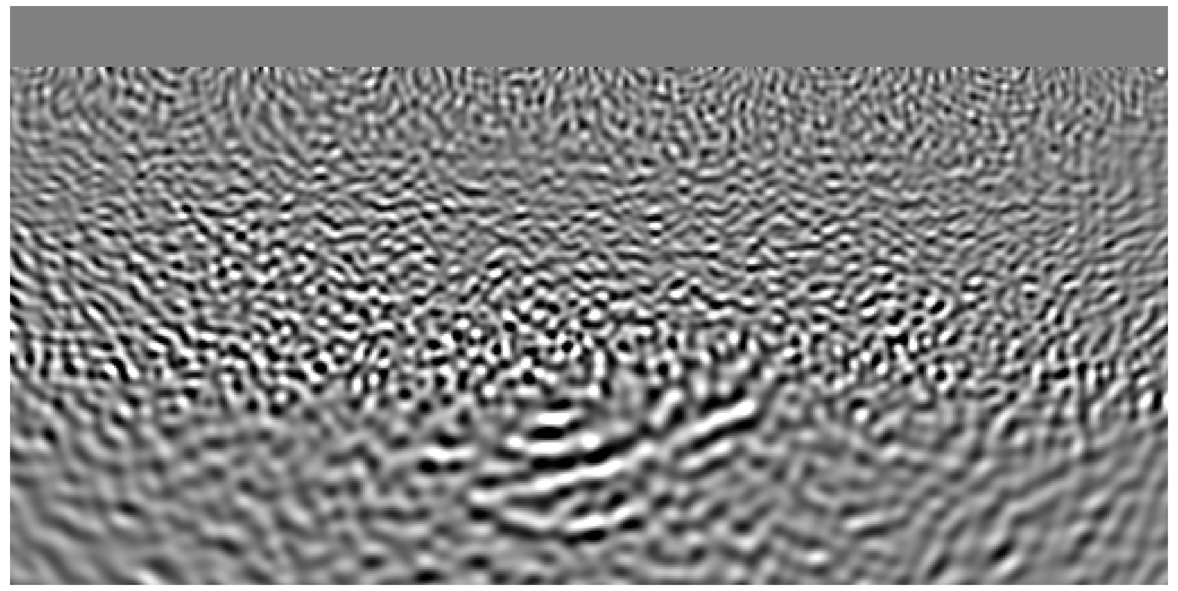
Drill well using optimal well density



Field observation

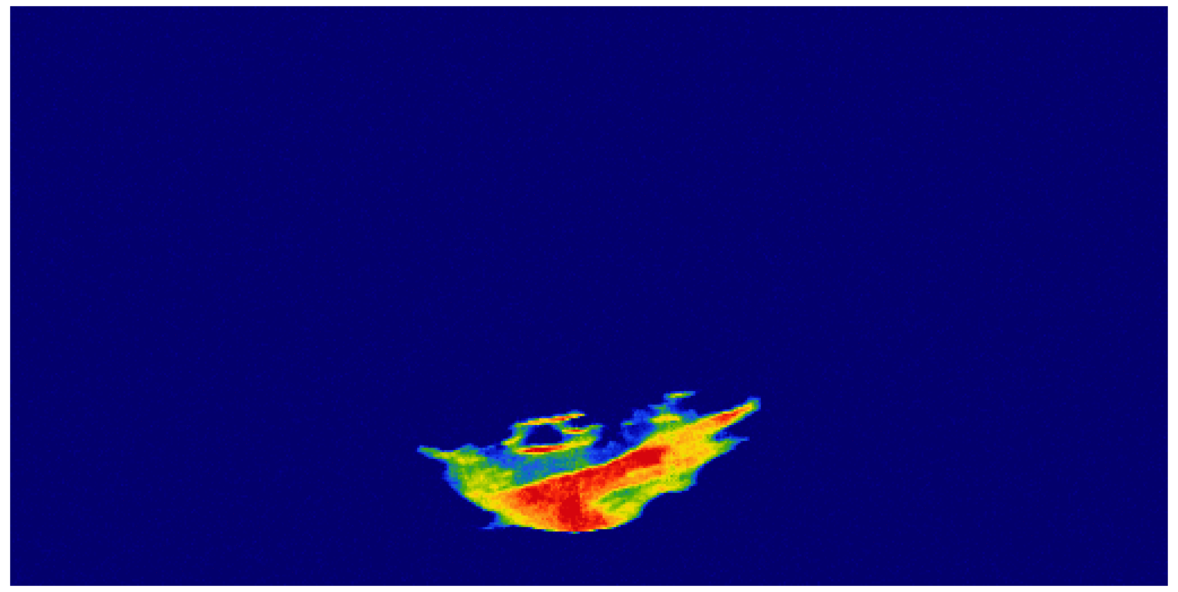
Collect field data $\mathbf{y}^{obs}_{t+1}$ w/ optimal well





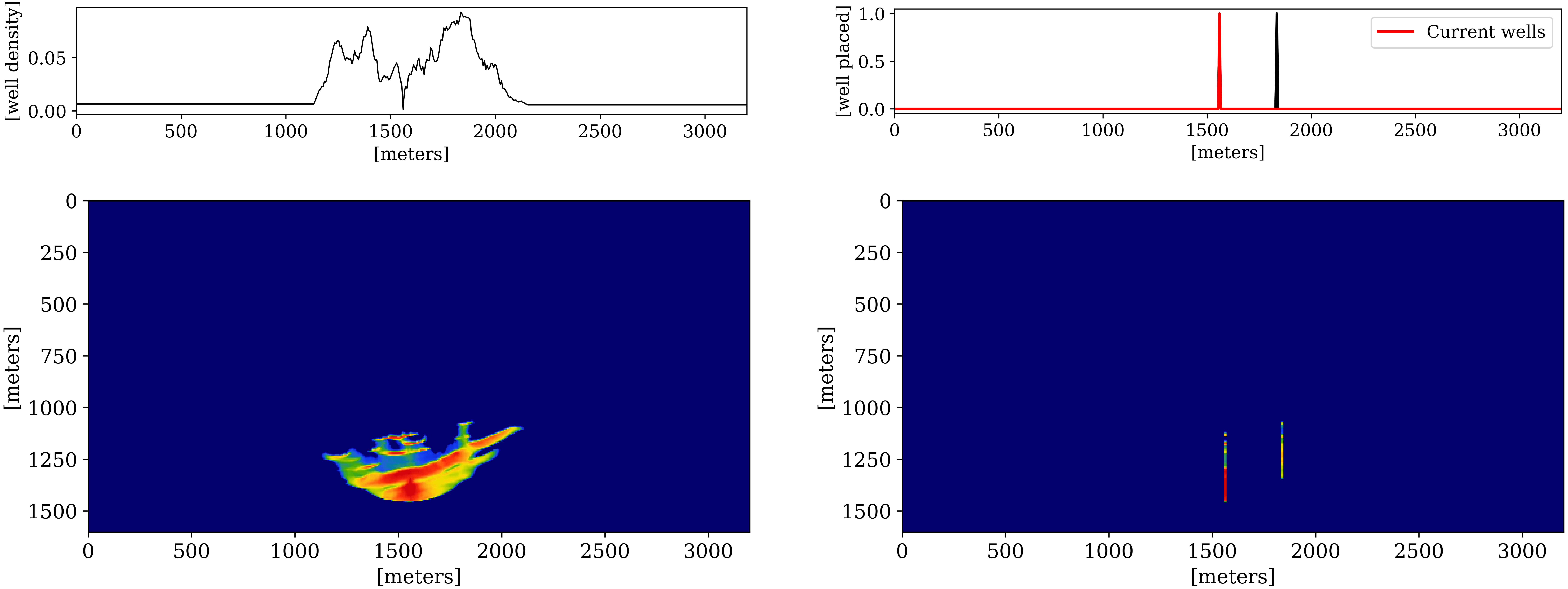
Posterior inference

Inference from field data $p_{\hat{\theta}}(\mathbf{x}_{t+1} \mid \mathbf{y}^{obs}_{t+1})$



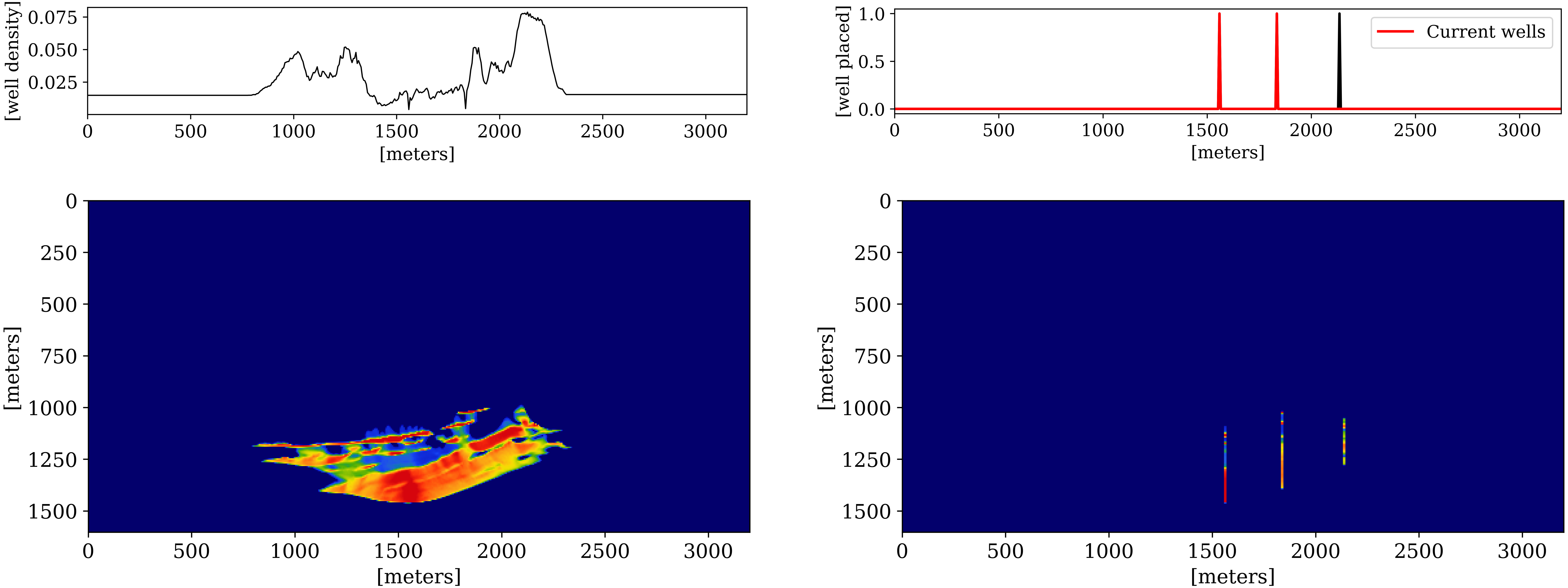
Posterior becomes prior and recurse

Monitor 1



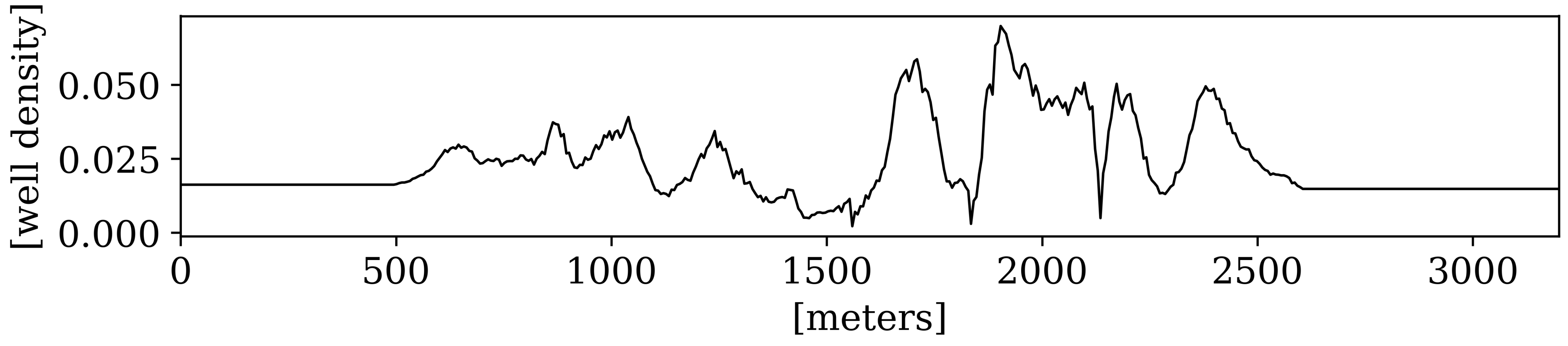
[well density]
0.05
0.00
0
500
1000
1500
2000
2500
3000
[meters]
[well placed]
1.0
0.5
0.0
Current wells
[meters]
0
250
500
750
1000
1250
1500

Monitor 2

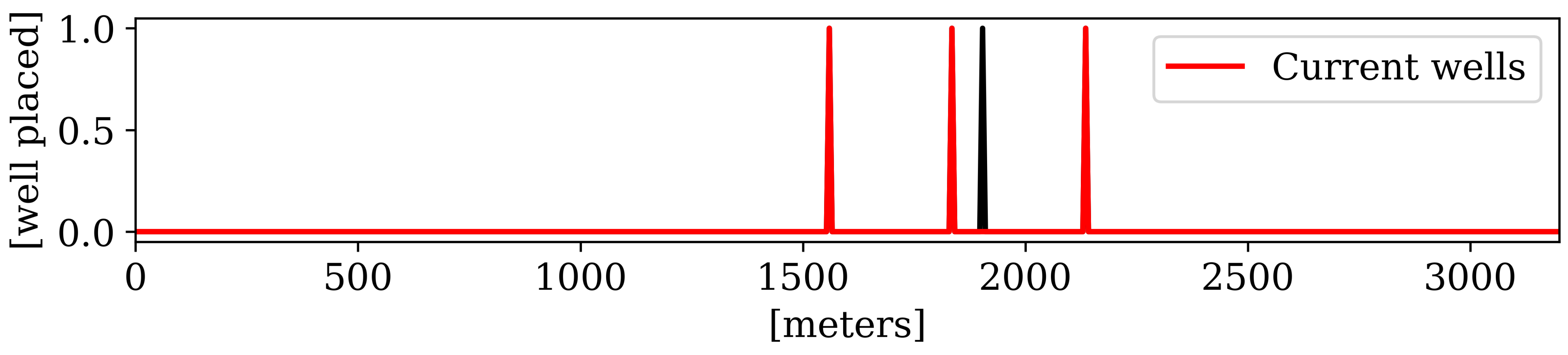


[well density]
0.075
0.050
0.025
0
500
1000
1500
2000
2500
3000
[meters]
[well placed]
1.0
0.5
0.0
Current wells
[meters]
0
250
500
750
1000
1250
1500
[meters]
[meters]

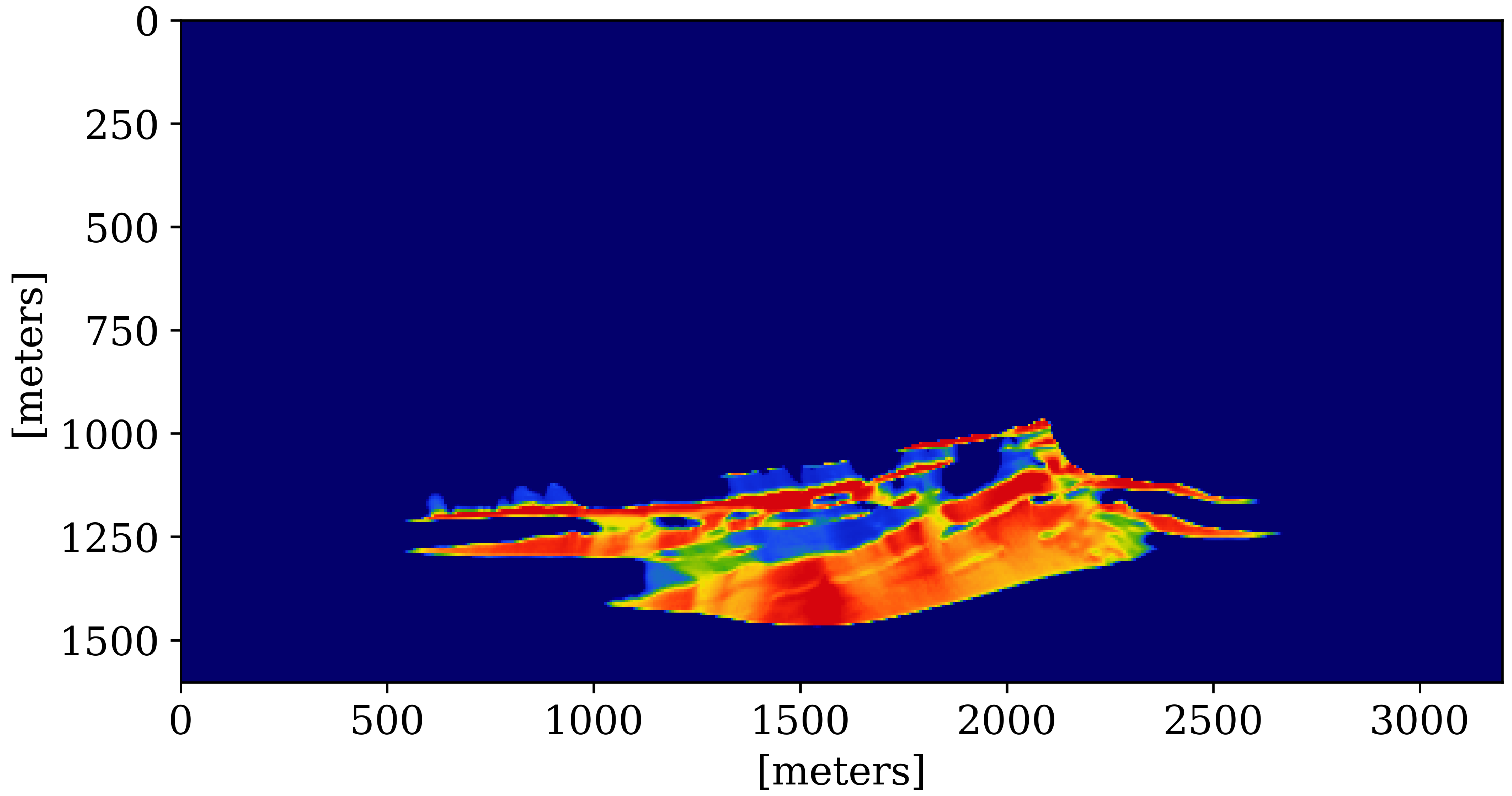
Monitor 3



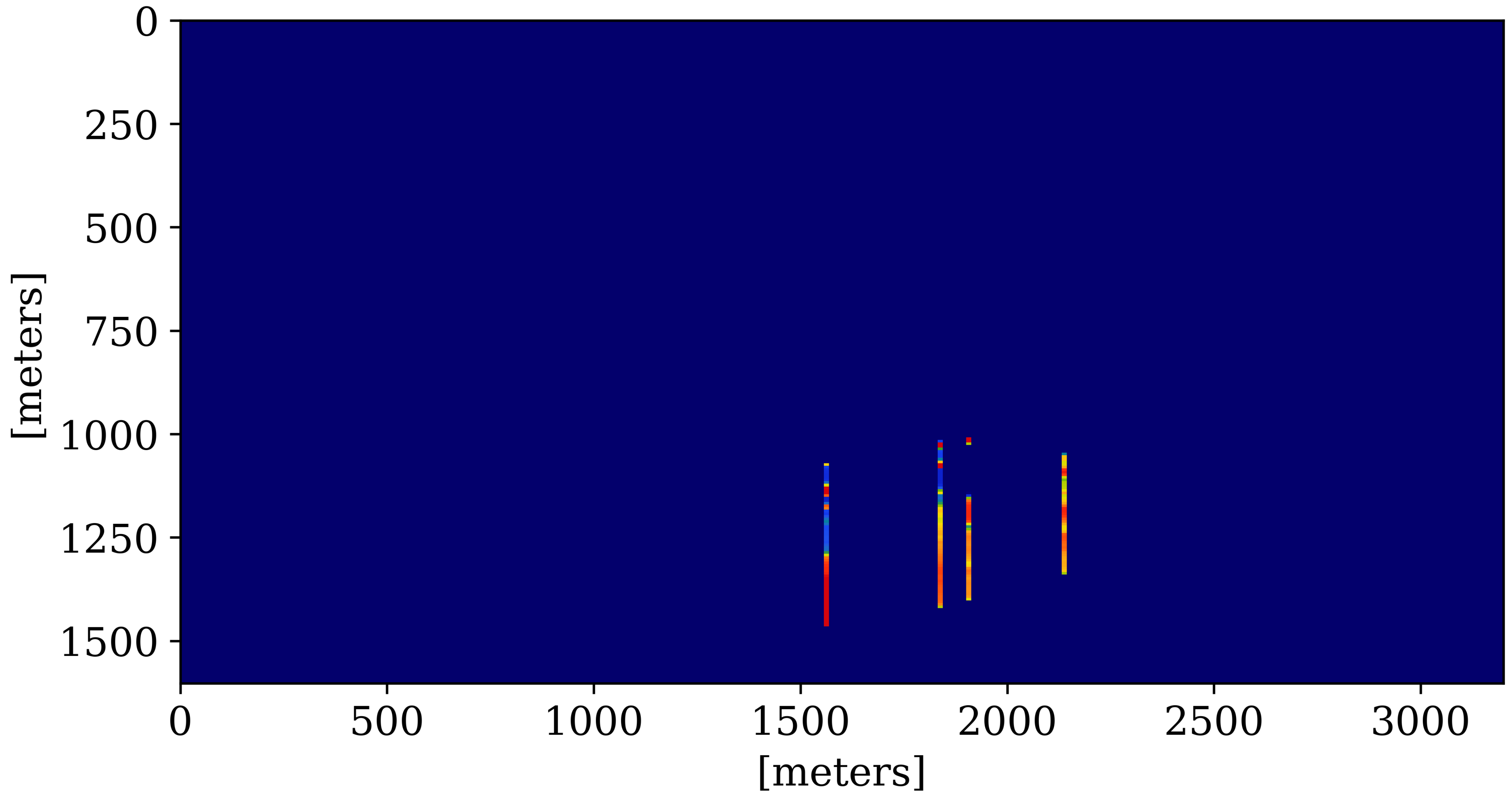
[well density]
0.050
0.025
0.000
0
500
1000
1500
2000
2500
3000
[meters]



[well placed]
1.0
0.5
0.0
Current wells
0
500
1000
1500
2000
2500
3000
[meters]

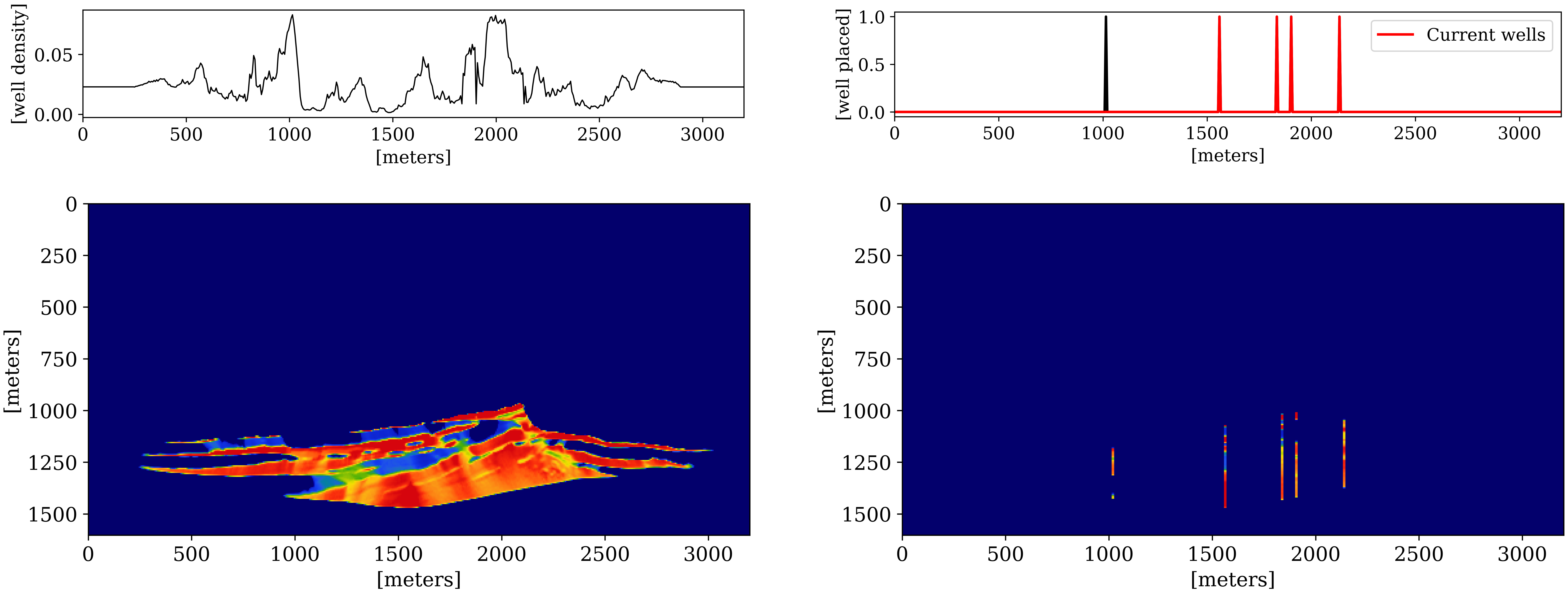


[meters]
0
250
500
750
1000
1250
1500
0
500
1000
1500
2000
2500
3000
[meters]

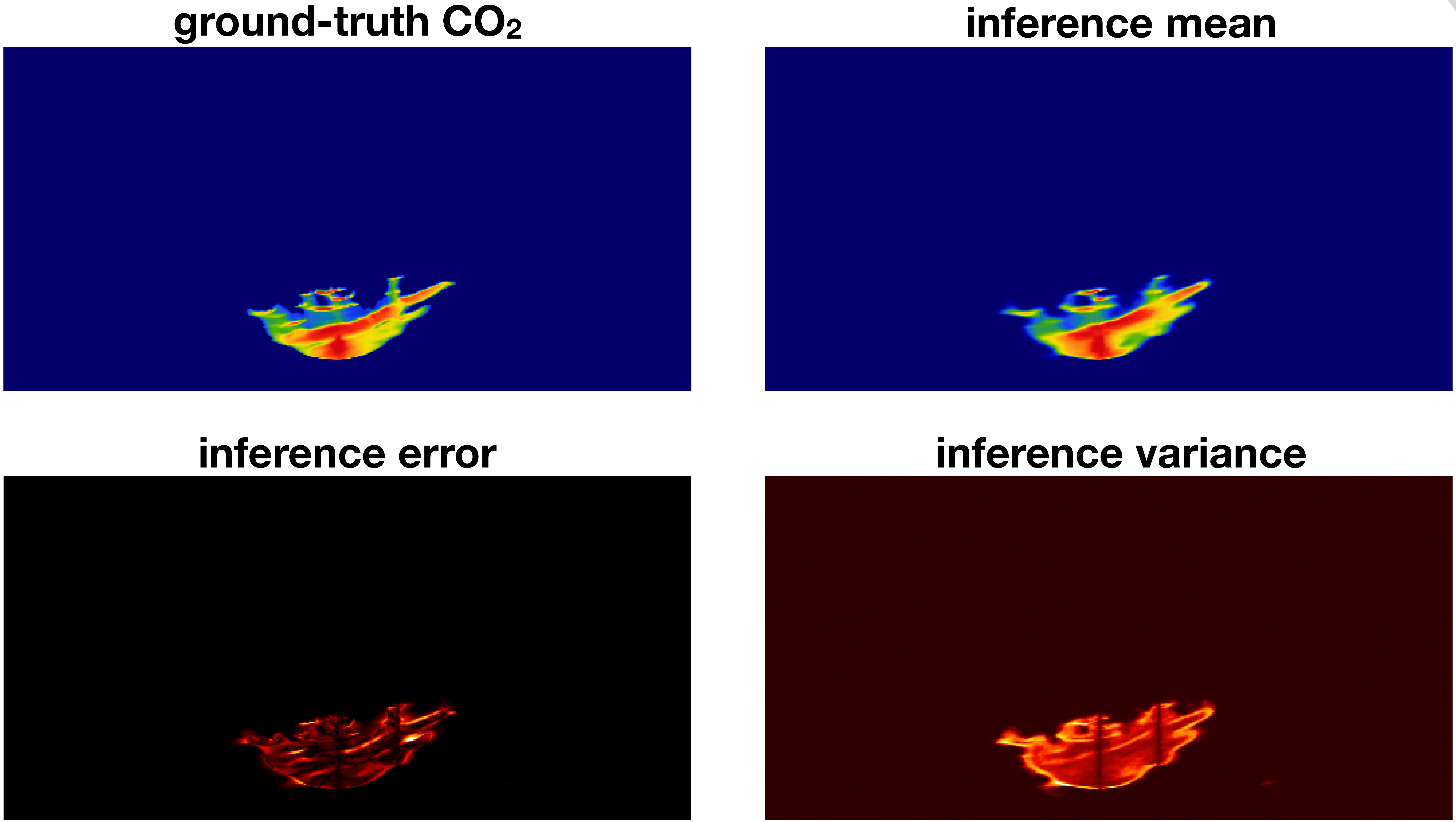


[meters]
0
250
500
750
1000
1250
1500
0
500
1000
1500
2000
2500
3000
[meters]

Monitor 4

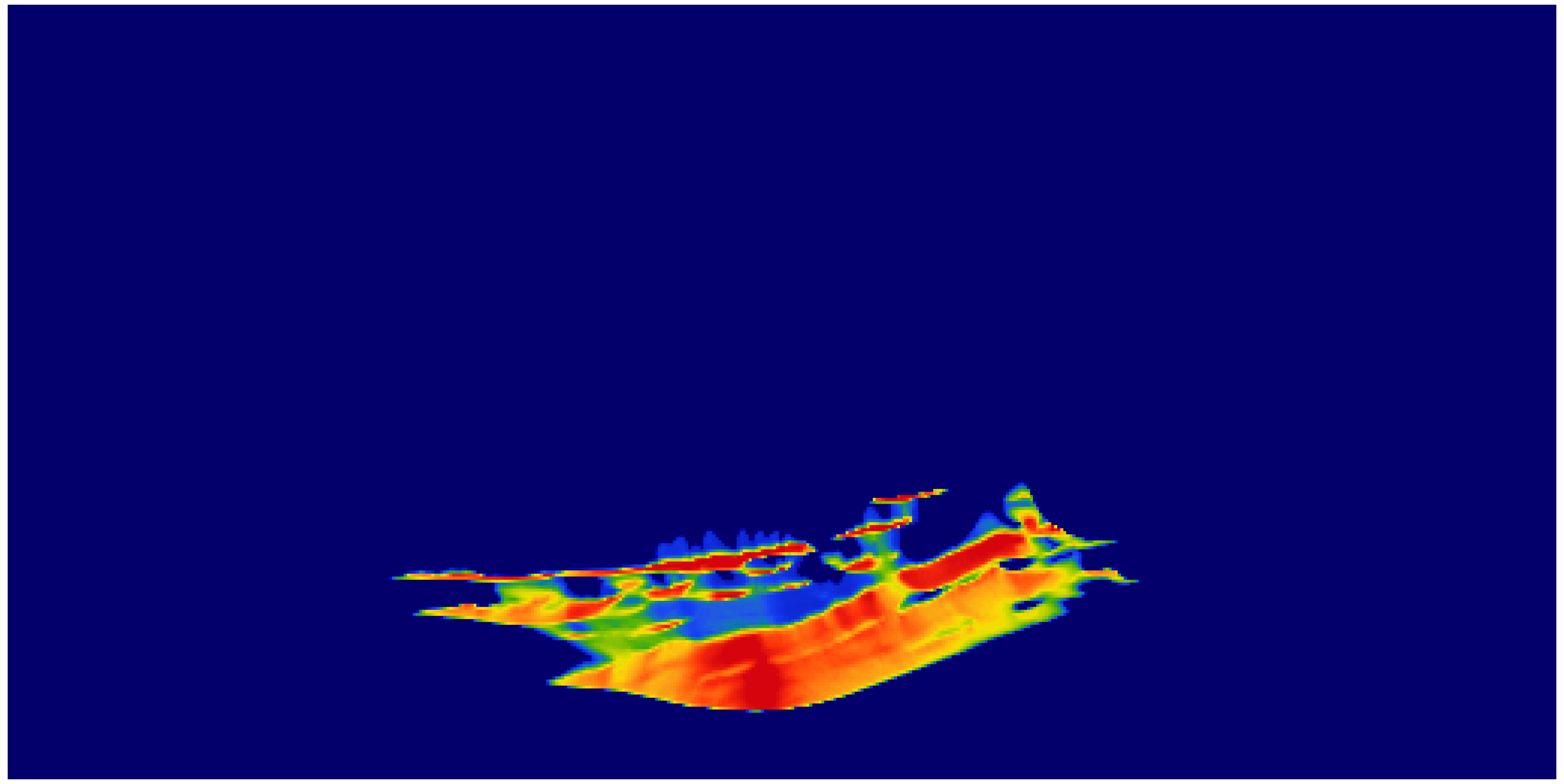

[well density]
0.05
0.00
0
500
1000
1500
2000
2500
3000
[meters]
[well placed]
1.0
0.5
0.0
Current wells
[meters]
0
250
500
750
1000
1250
1500
[meters]

Monitor 1

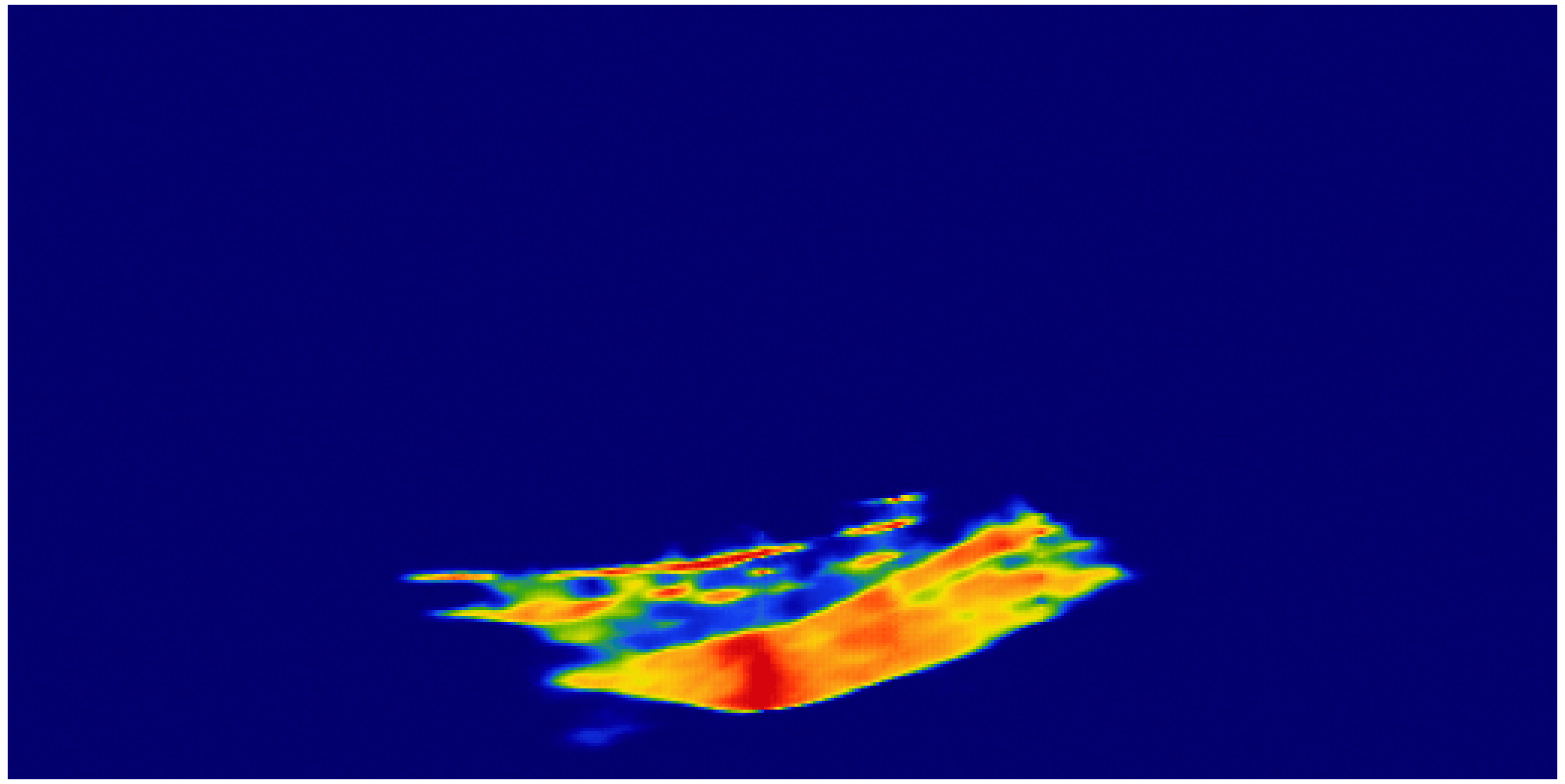

ground-truth CO_2
inference mean
inference error
inference variance

Monitor 2

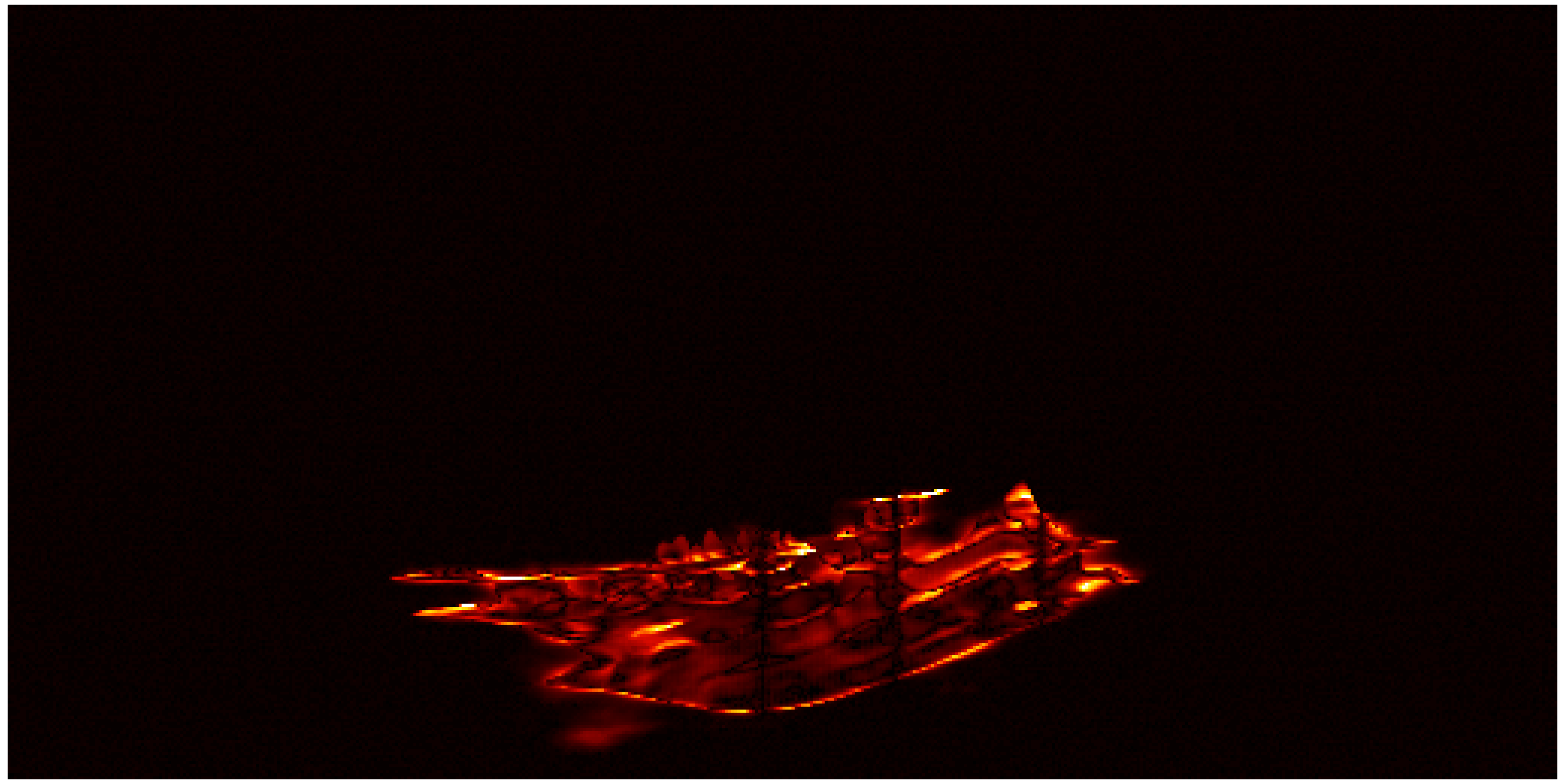
ground-truth CO_2



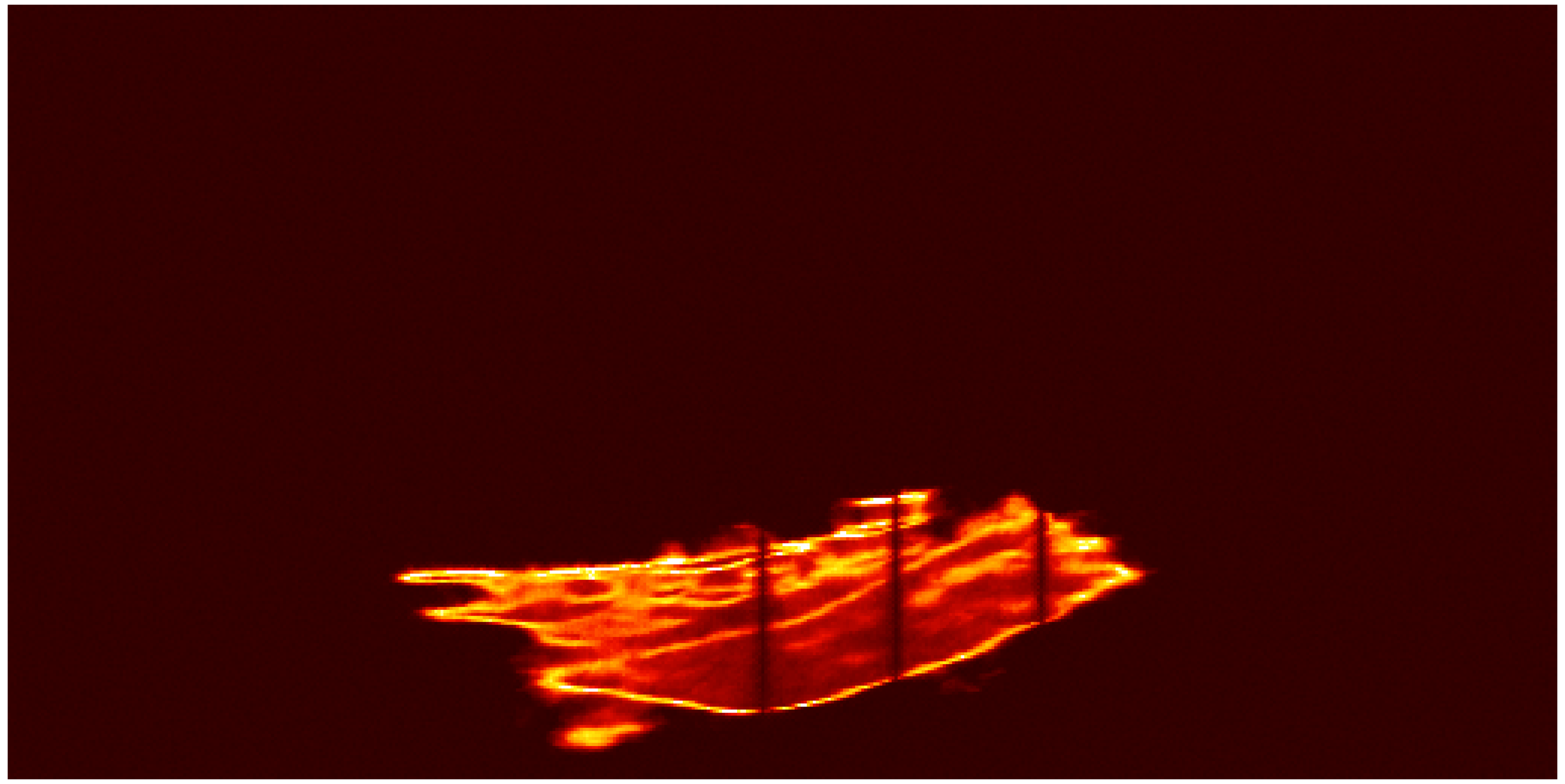
inference mean



inference error

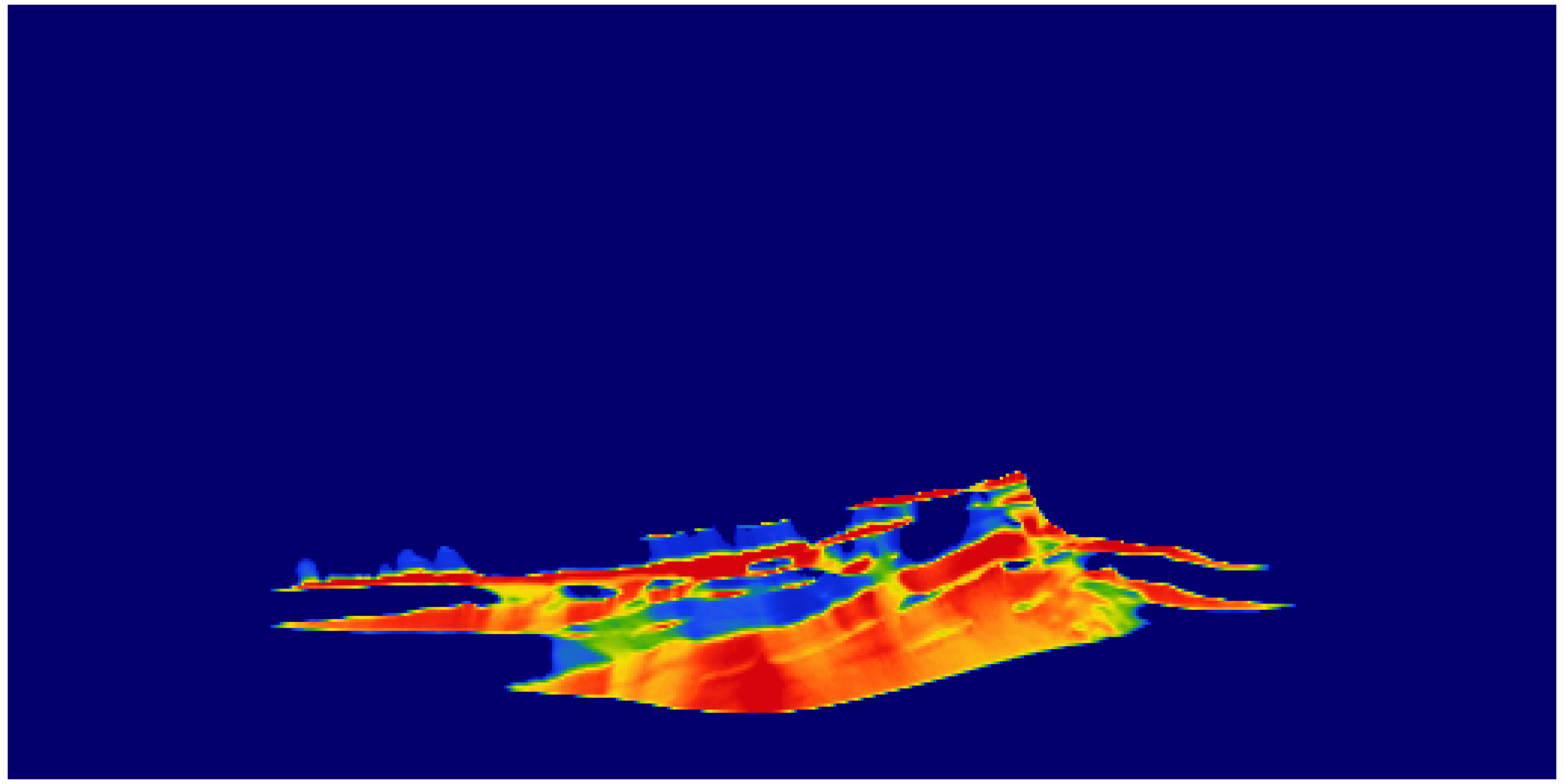


inference variance

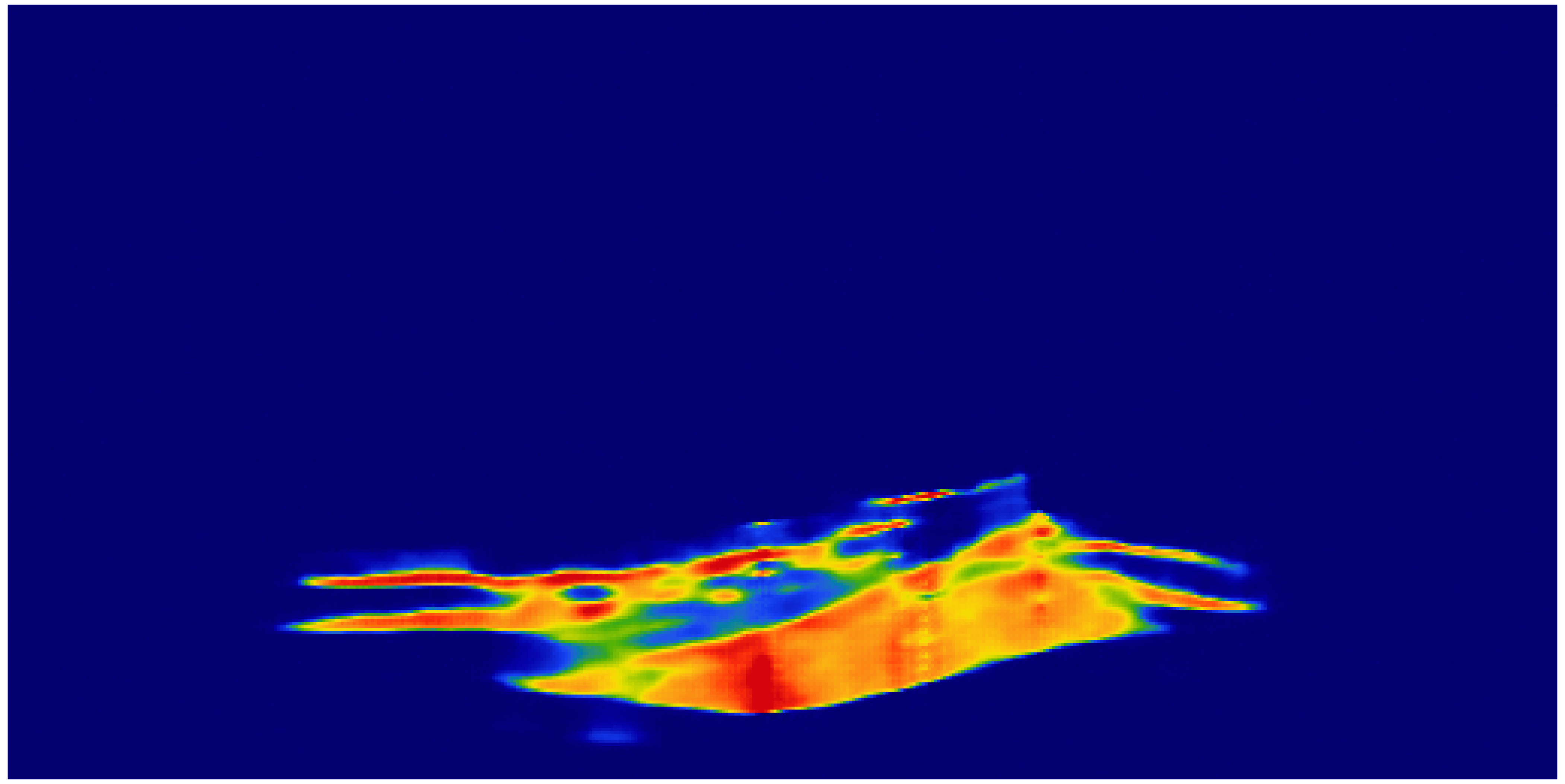


Monitor 3

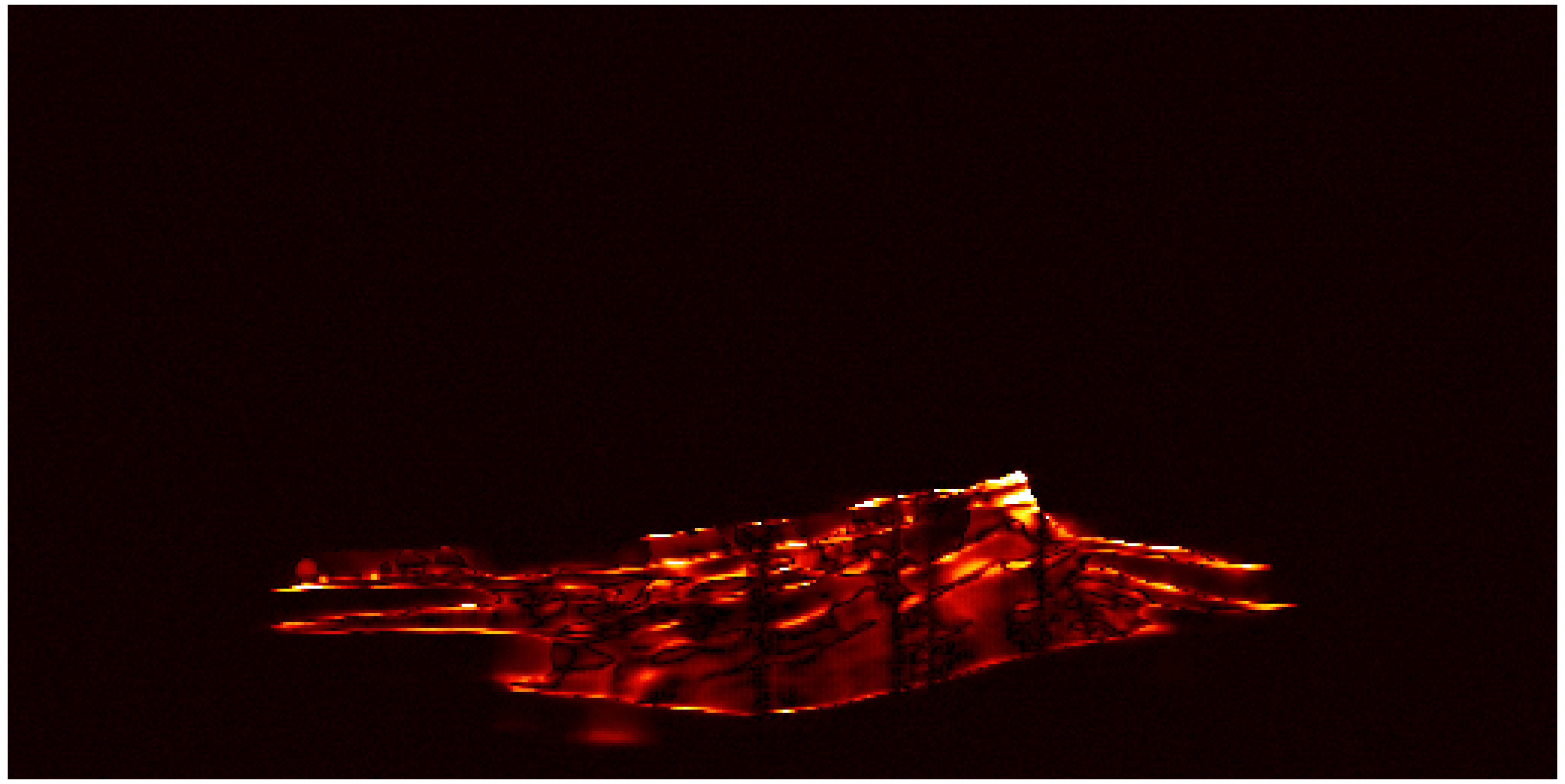
ground-truth CO_2



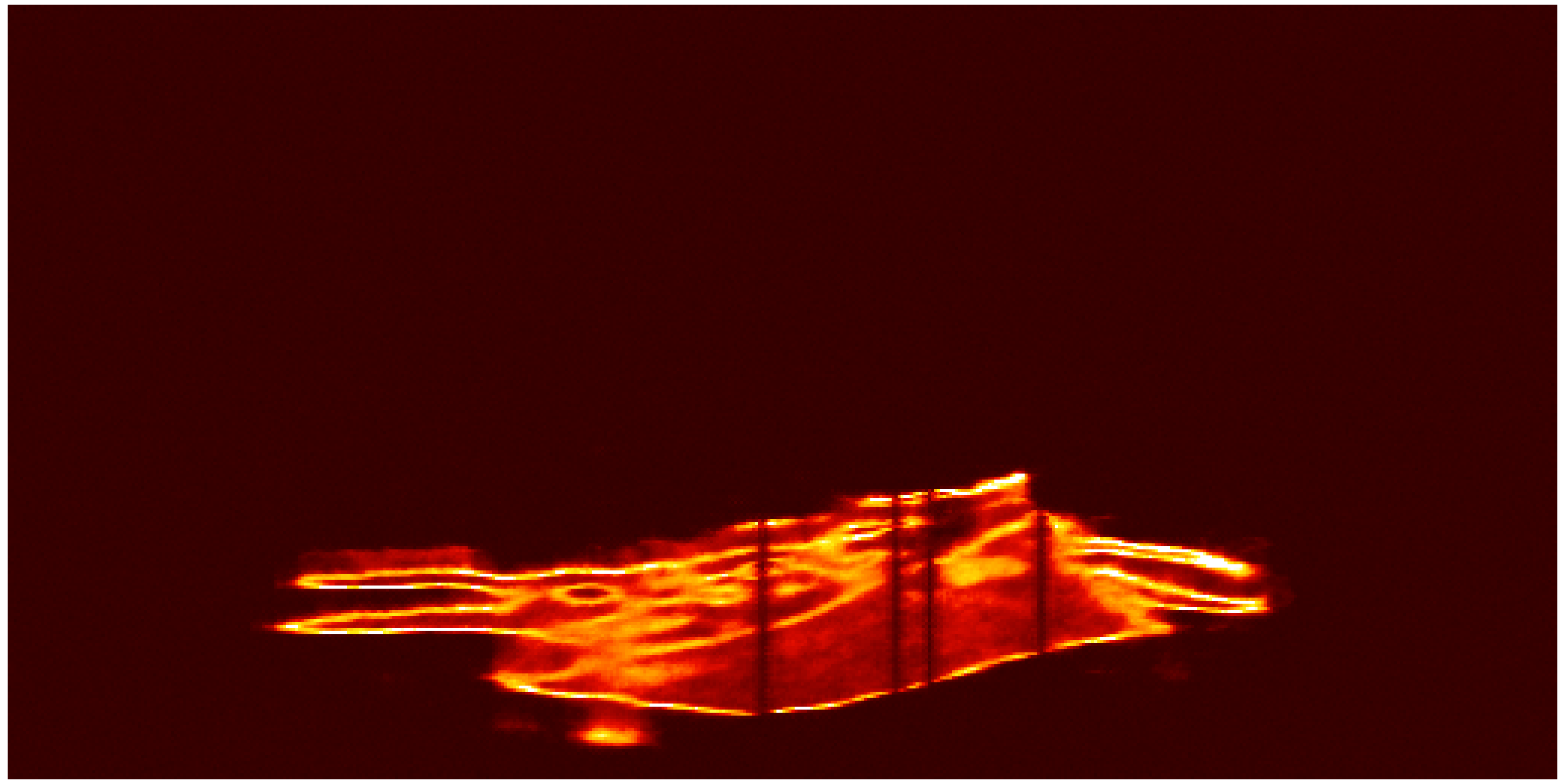
inference mean



inference error

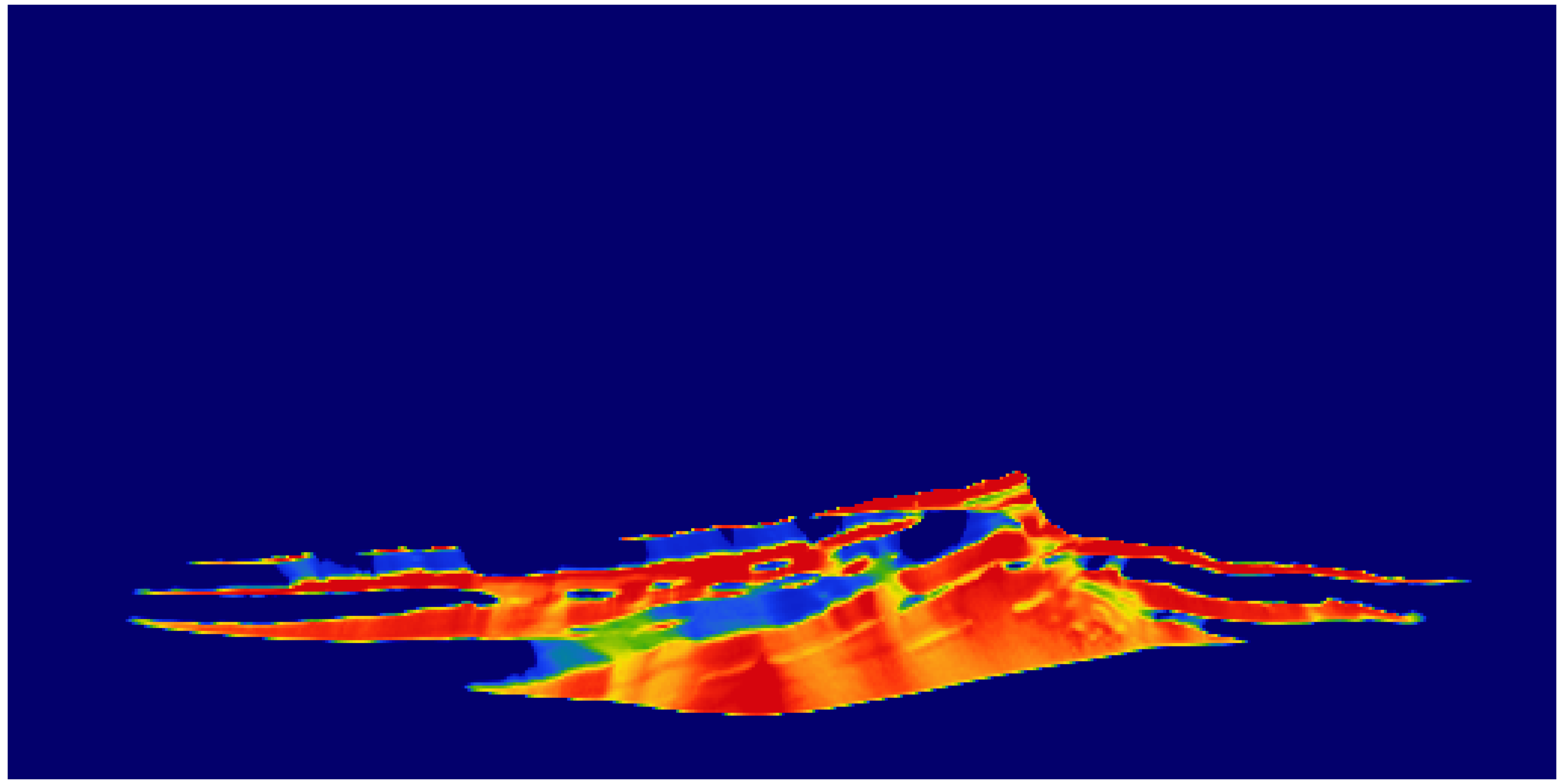


inference variance

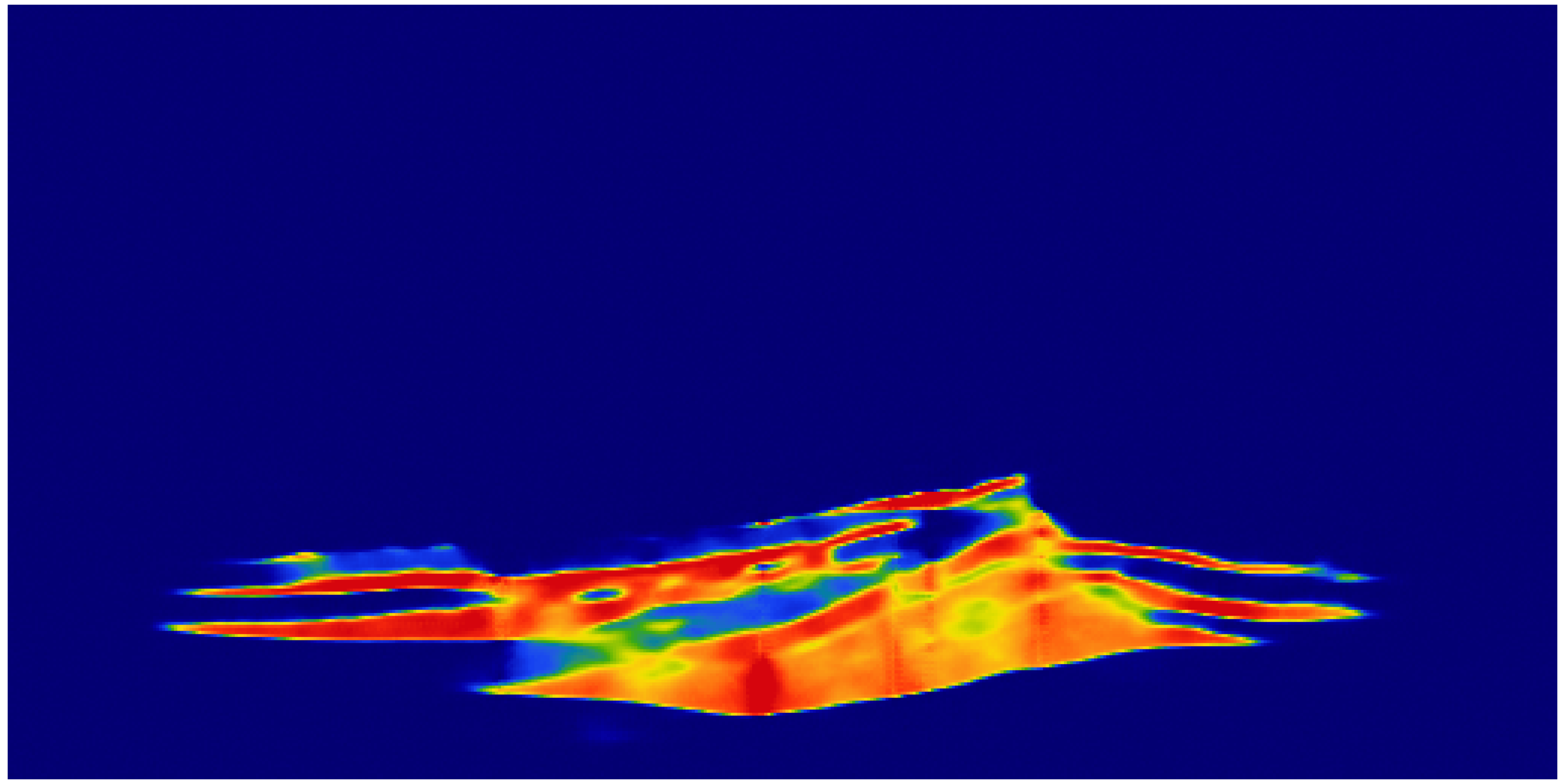


Monitor 4

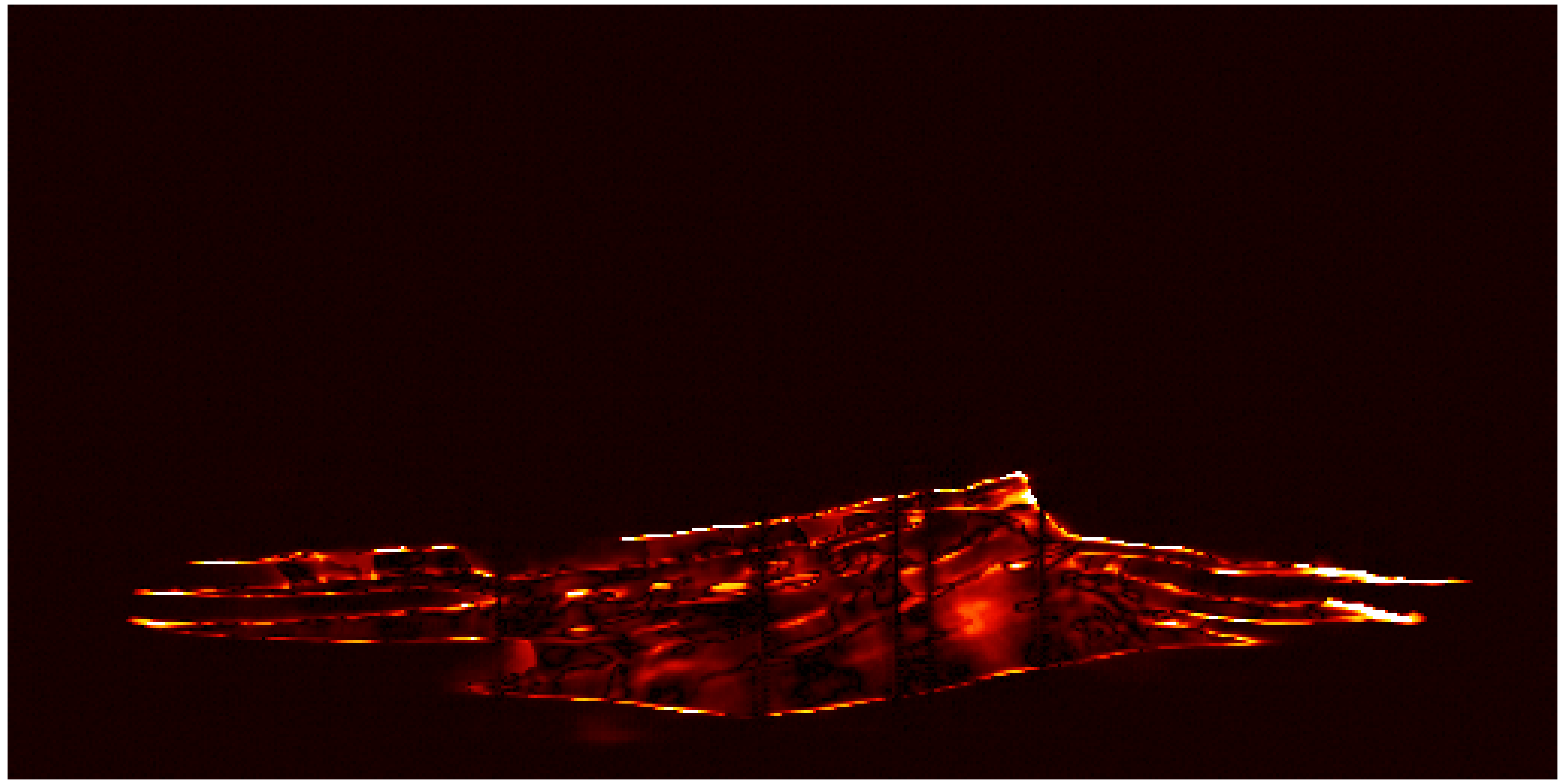
ground-truth CO_2



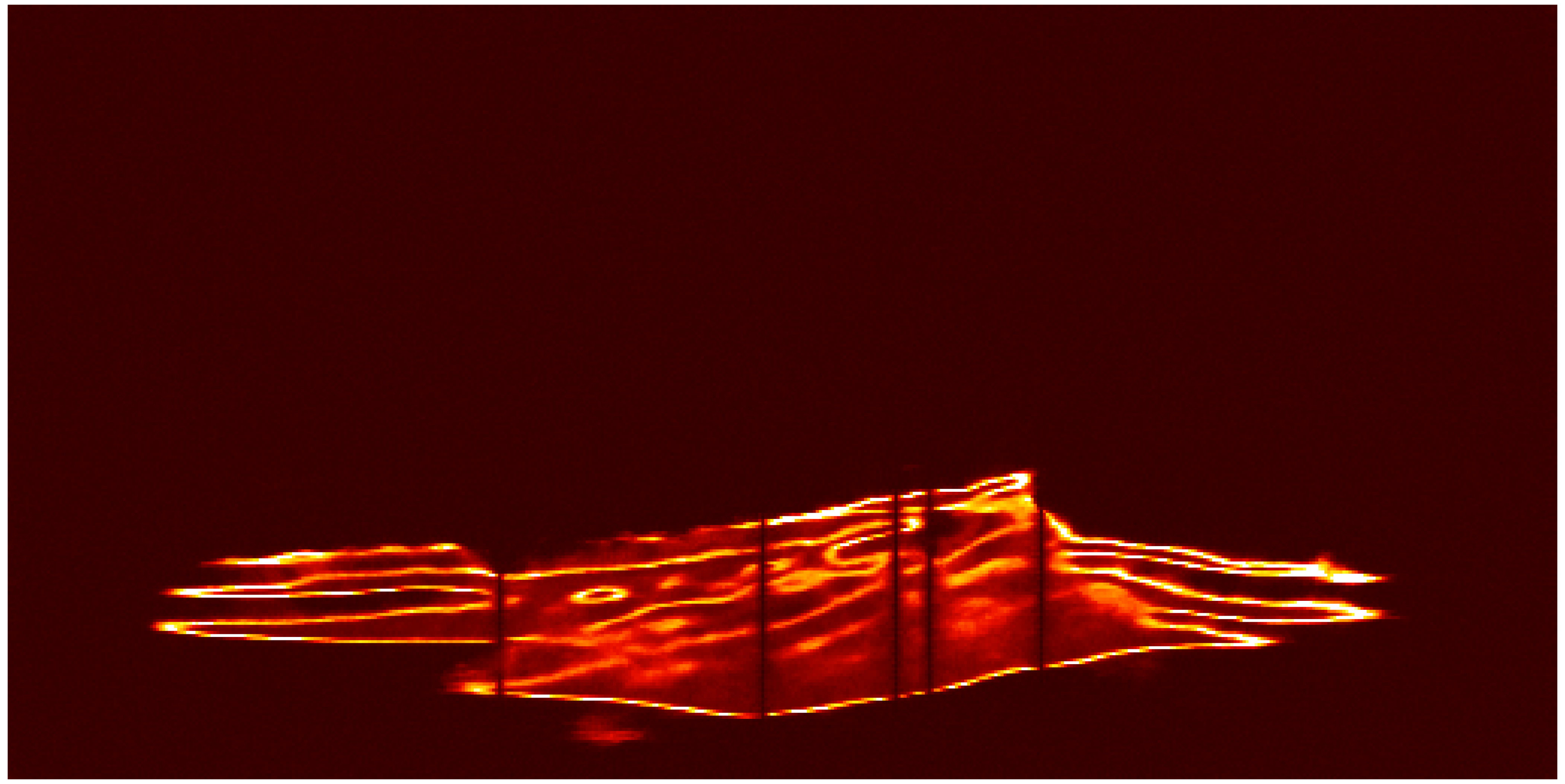
inference mean



inference error

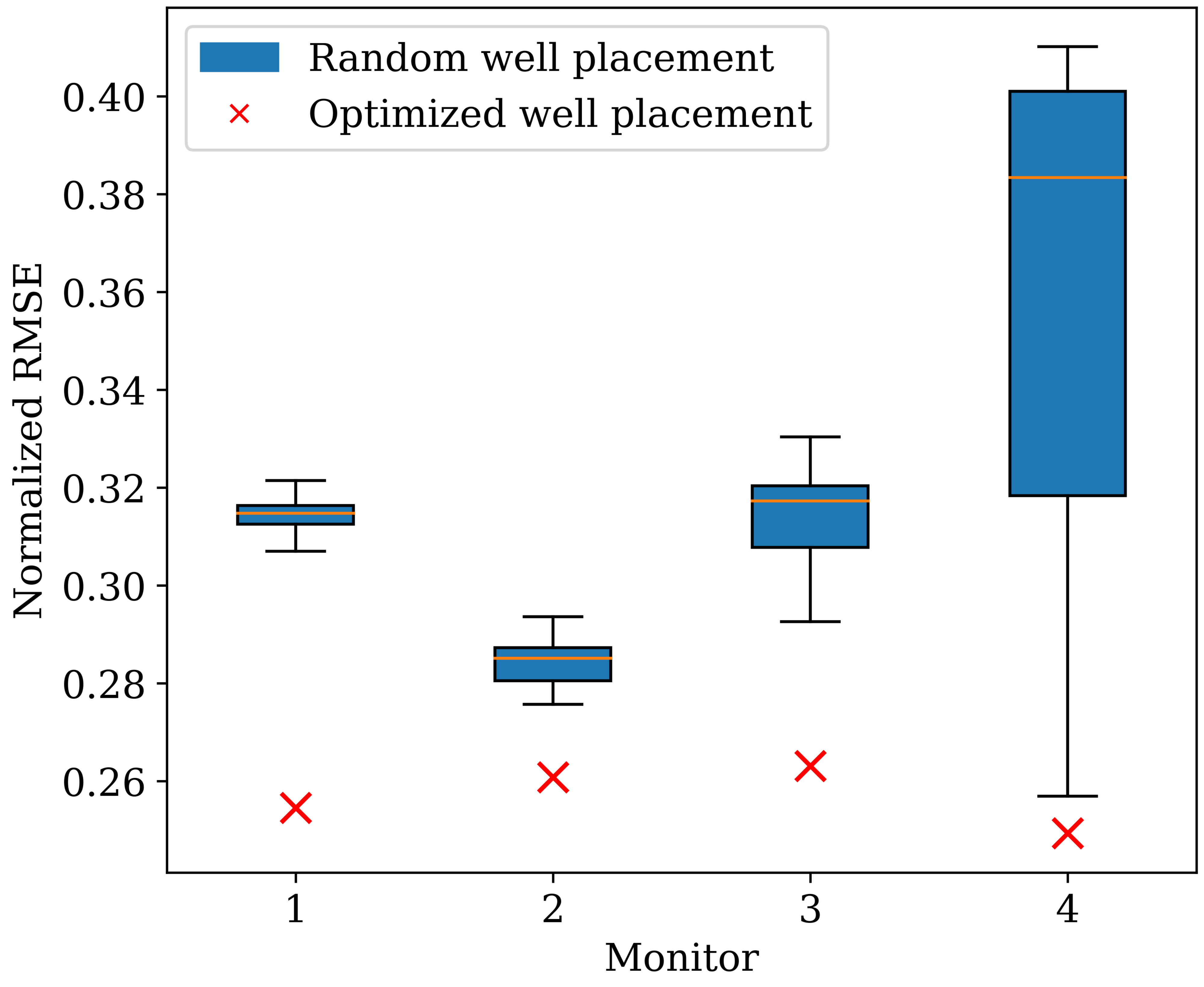


inference variance

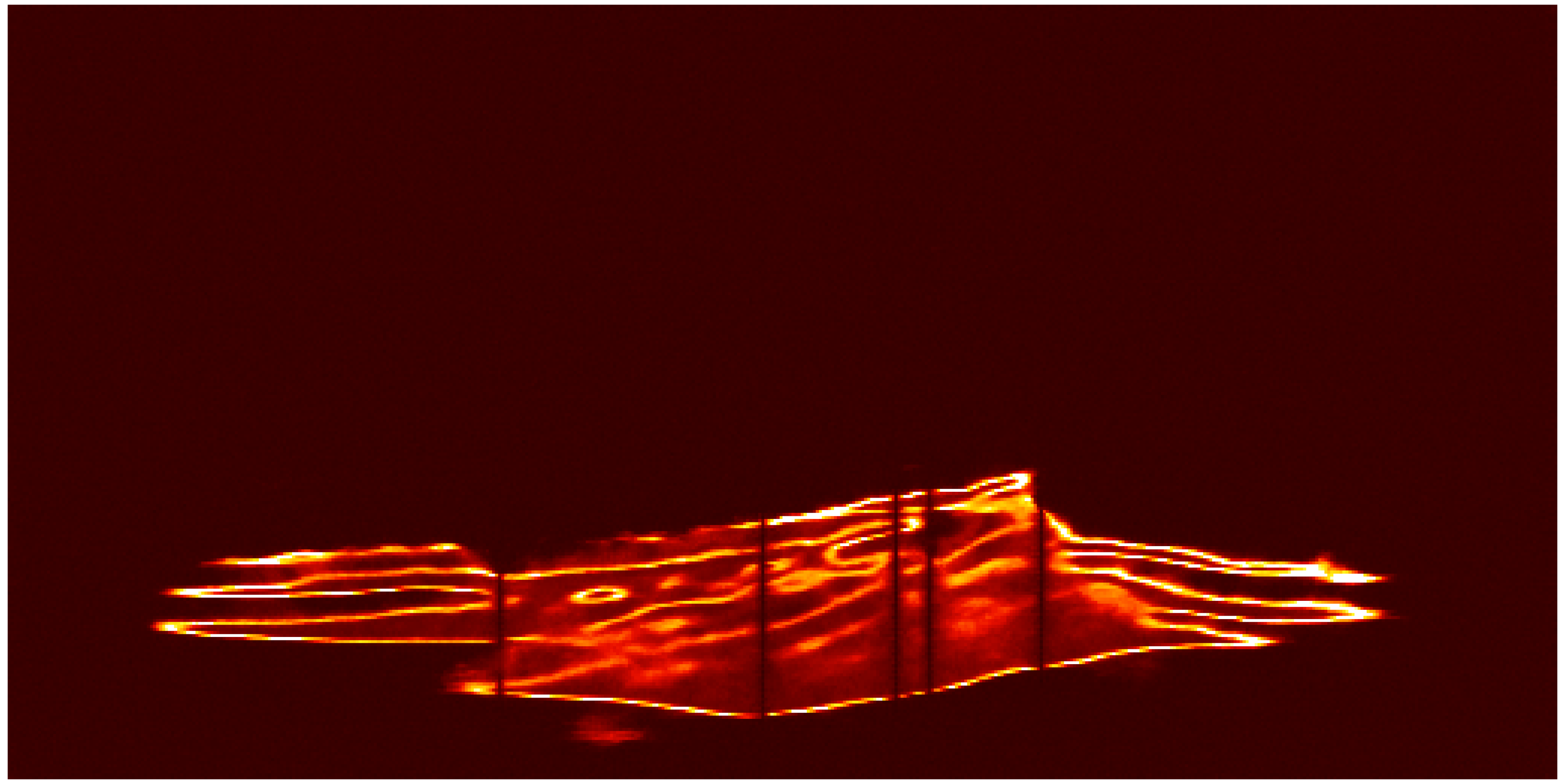


Improvement on baseline

Our algorithm places wells at optimal locations as measured by error



our inference variance

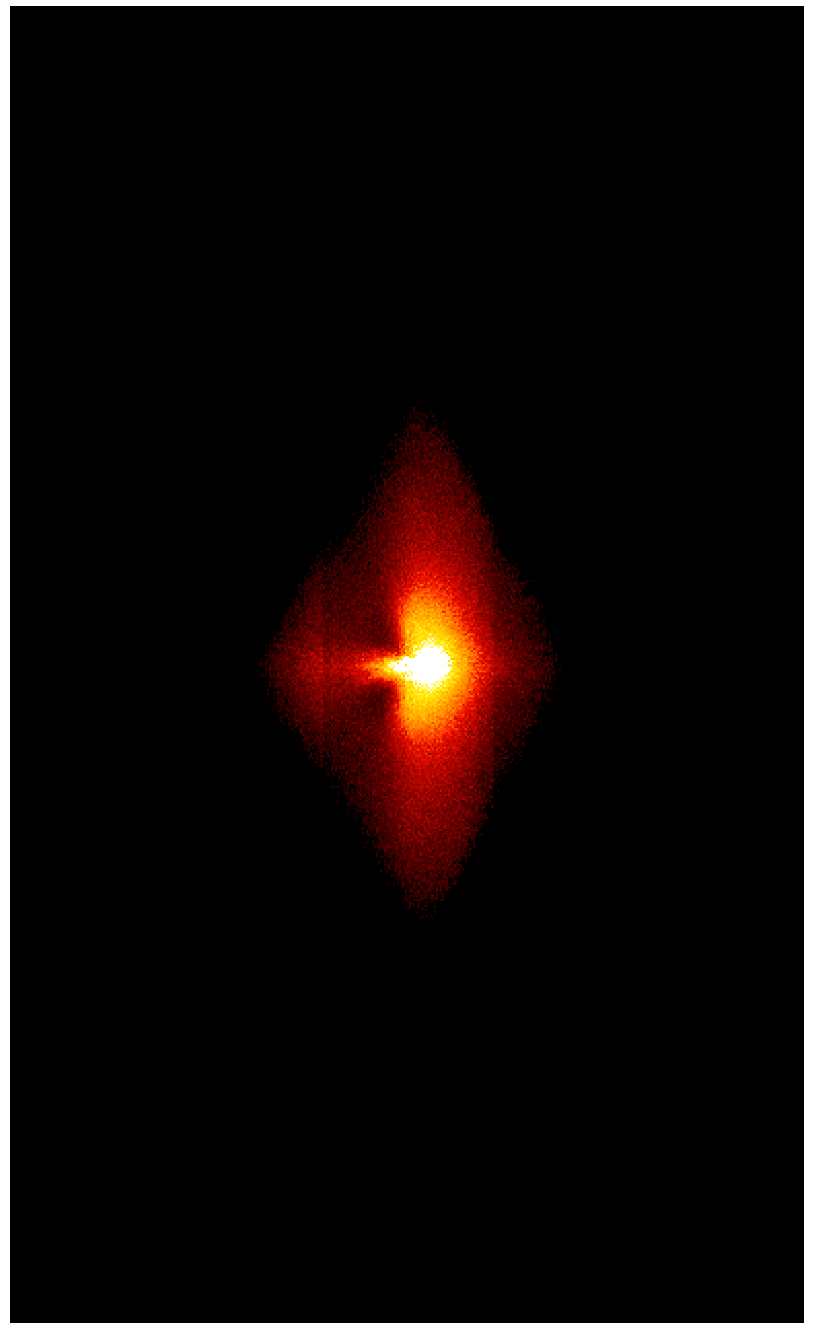


Conclusions

Gradient optimization of designs w.r.t a normalizing flow loss enables experimental design in realistic problems:

MRI

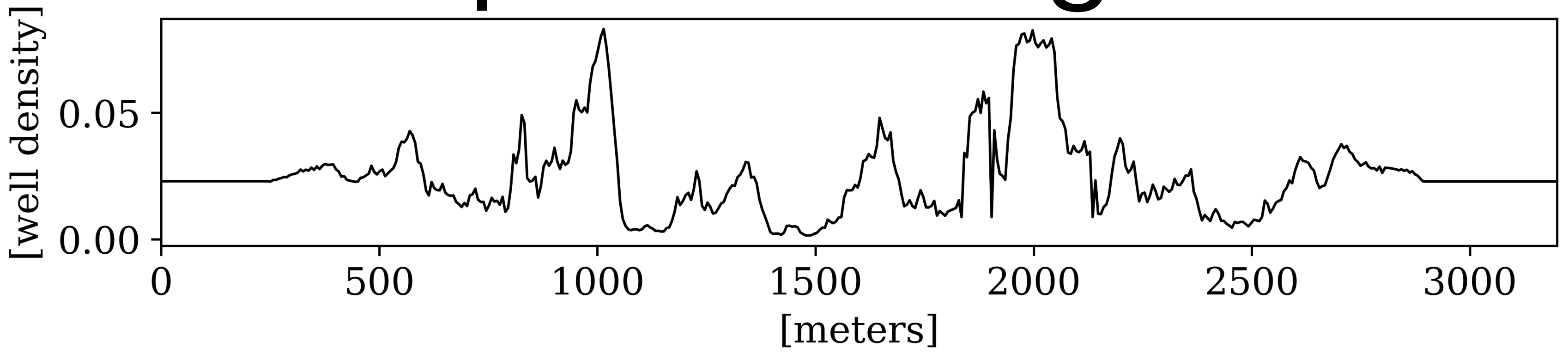
optimal design



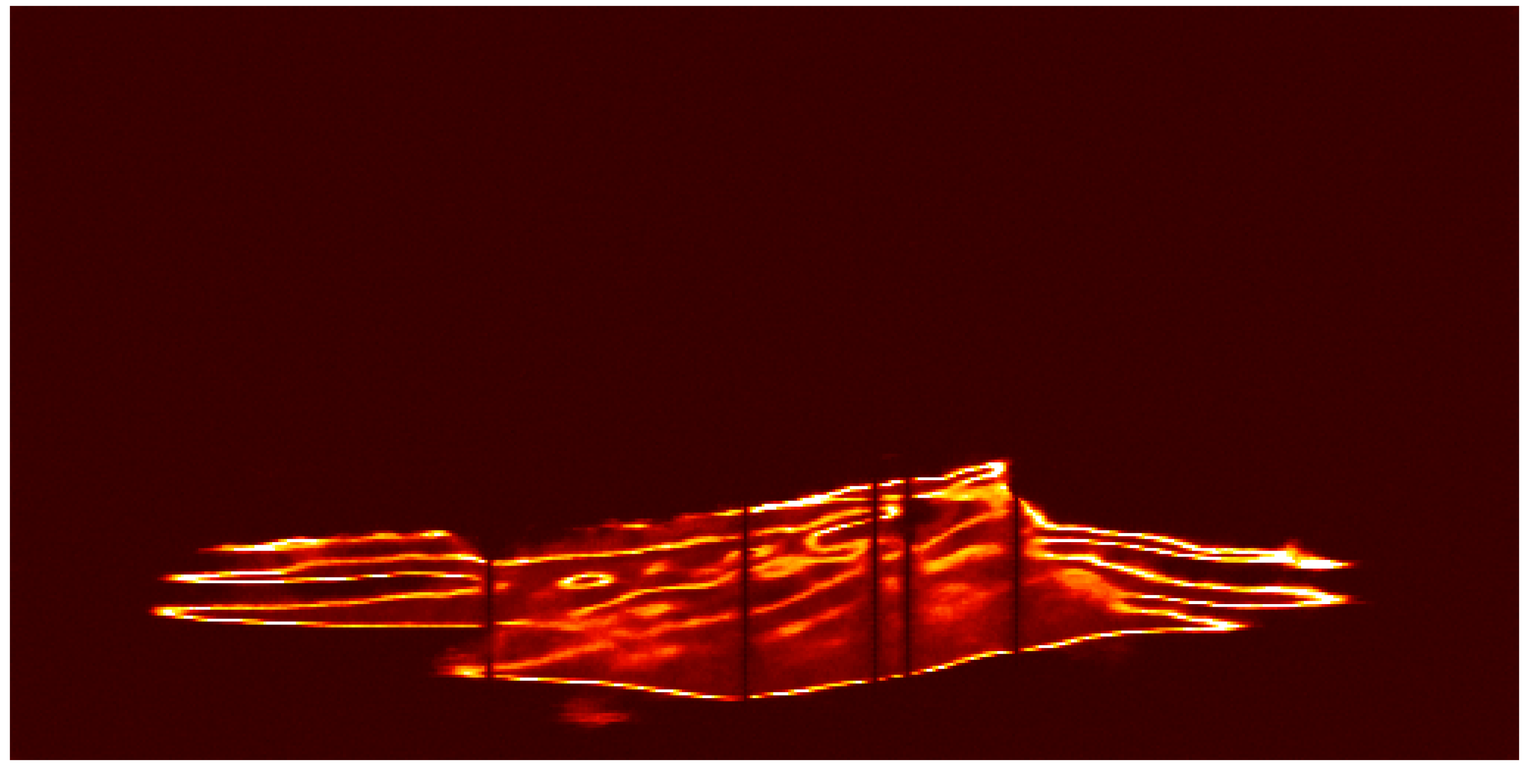
posterior samples

Underground CO2 monitoring

optimal design



posterior variance

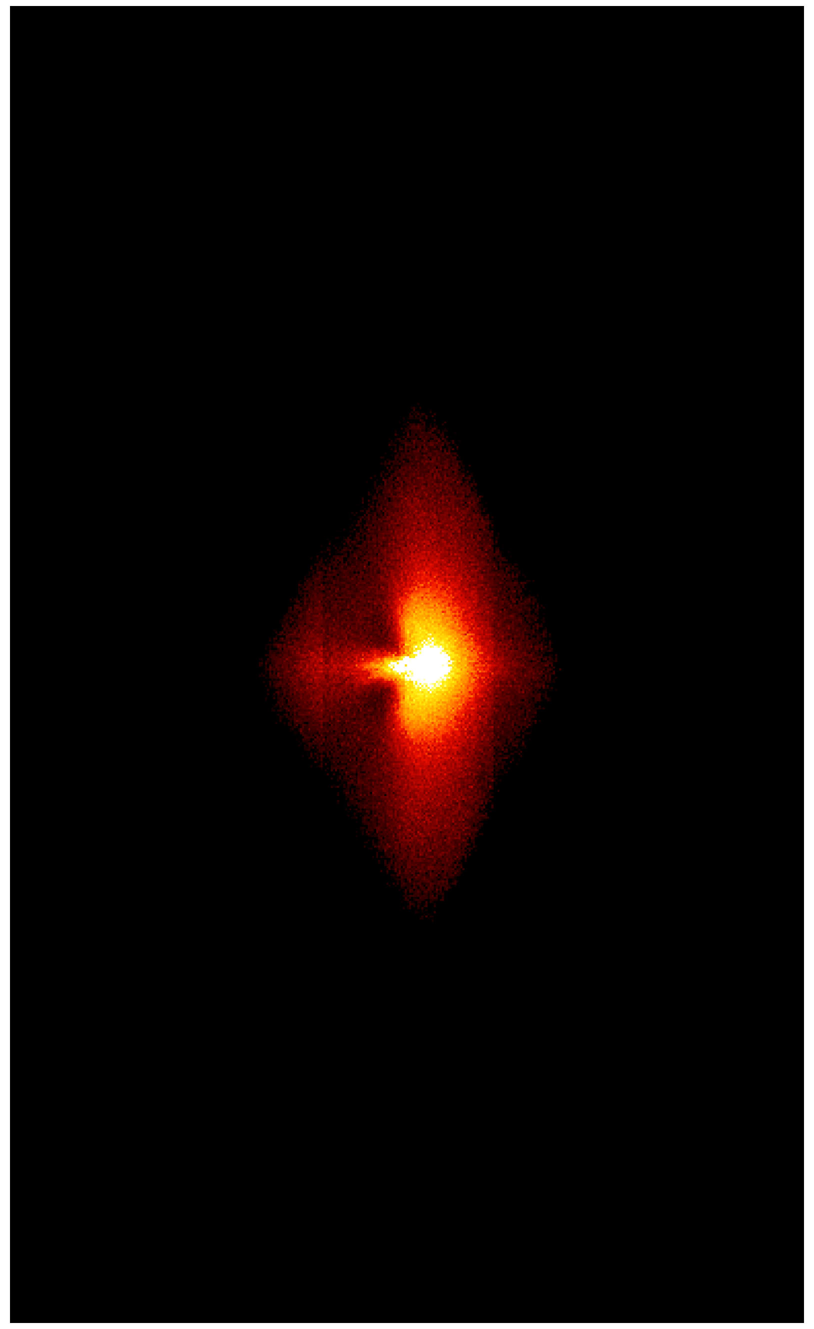


Conclusions

Gradient optimization of designs w.r.t a normalizing flow loss enables experimental design in realistic problems:

MRI

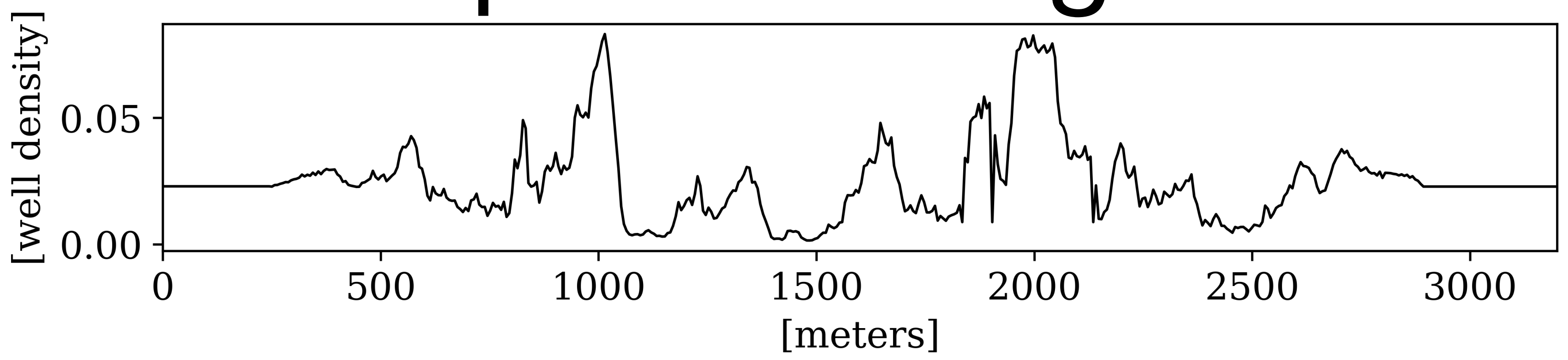
optimal design



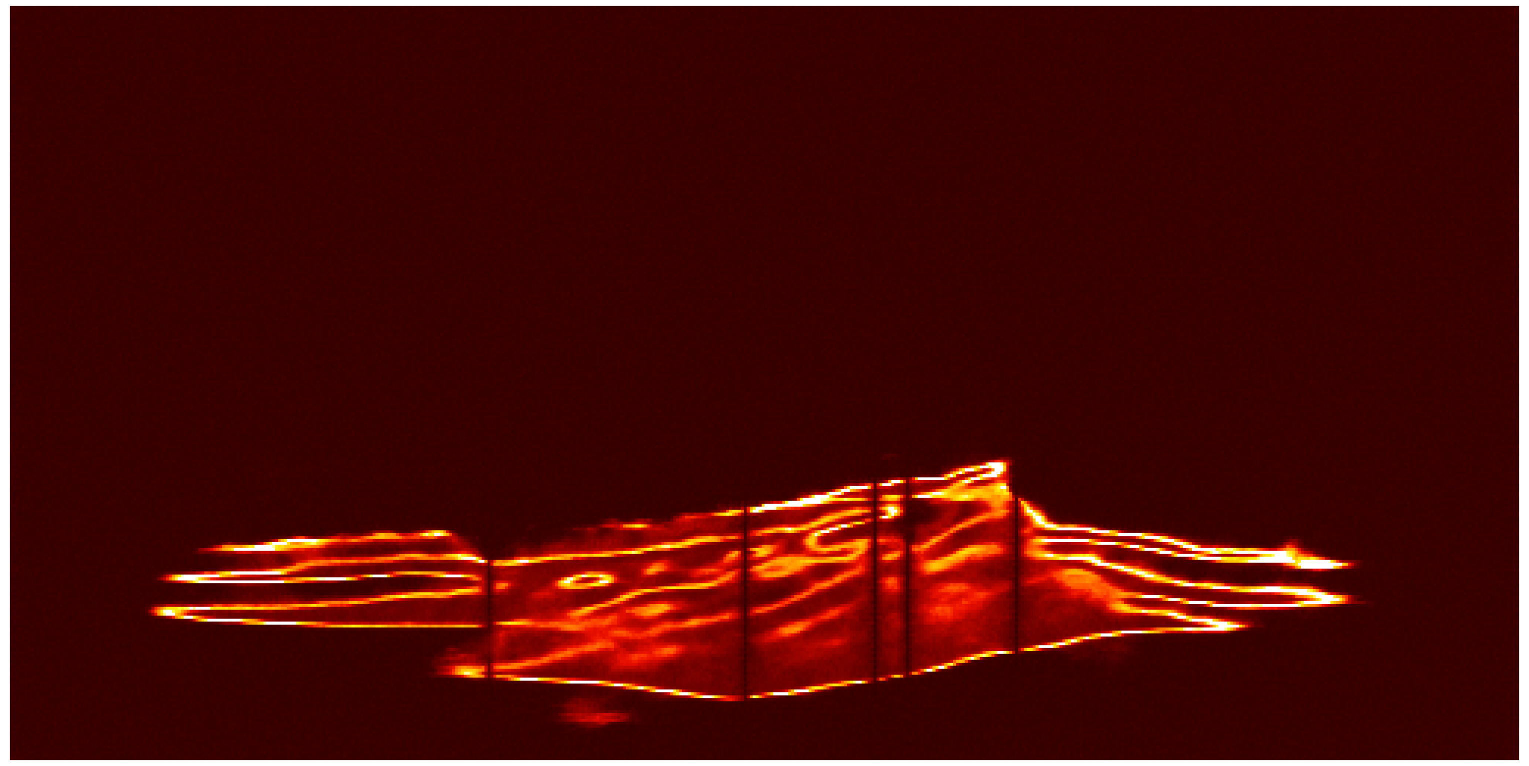
posterior samples

Underground CO2 monitoring

optimal design



posterior variance



and possible because normalizing flows have exact likelihood evaluation.

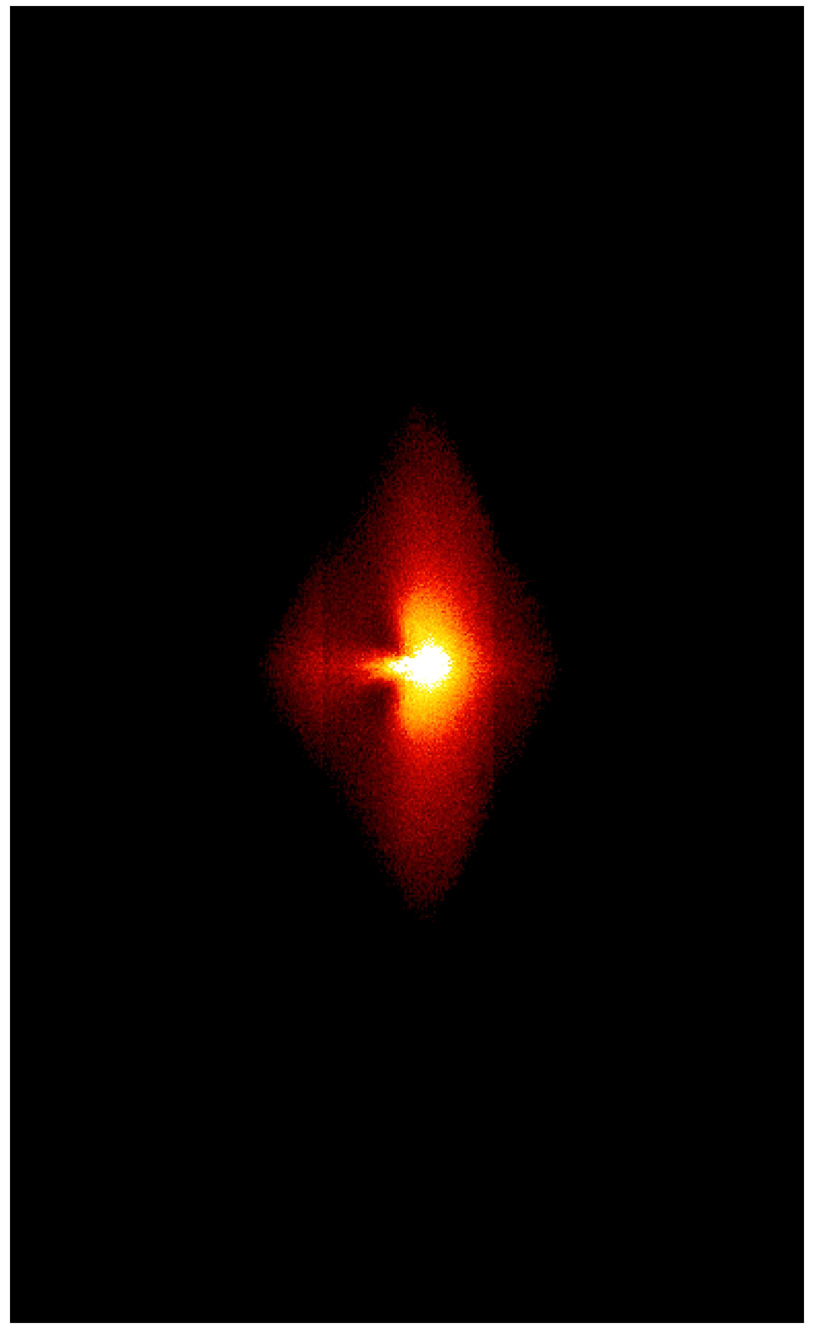
Conclusions

Probabilistic Bayesian optimal experimental design using conditional normalizing flows.
Orozco, Chen, Herrmann arxiv:2402.18337 (2024)

Gradient optimization of designs w.r.t a normalizing flow loss enables experimental design in realistic problems:

MRI

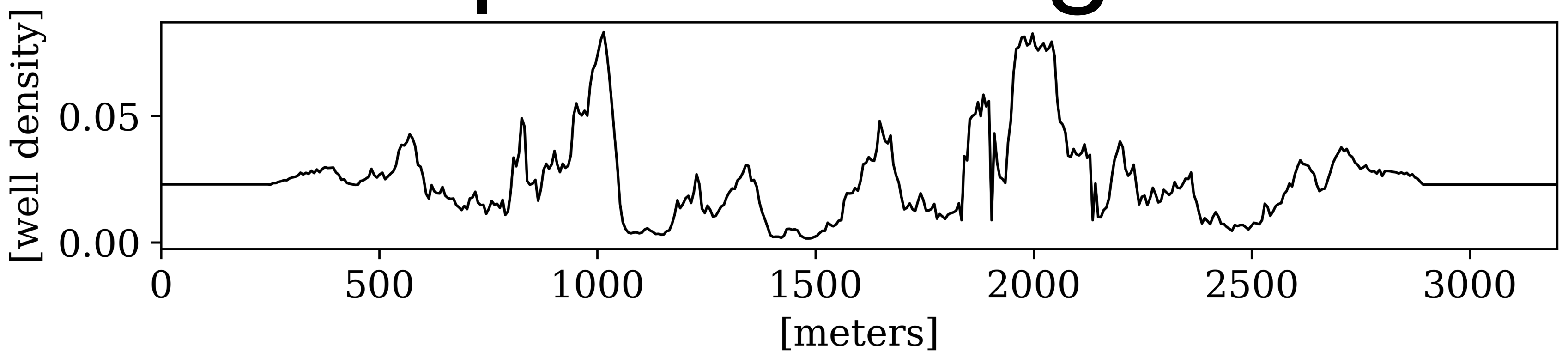
optimal design



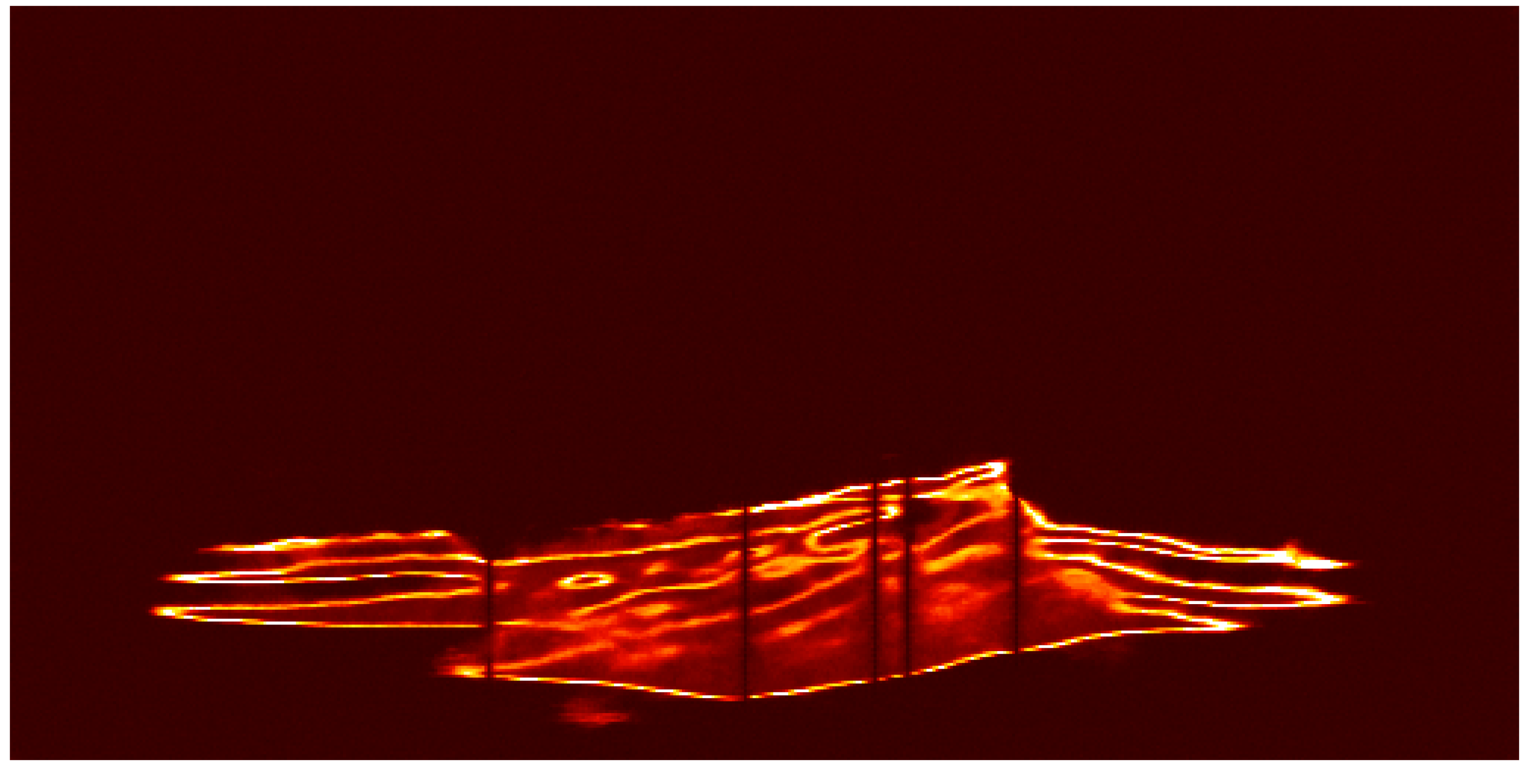
posterior samples

Underground CO2 monitoring

optimal design



posterior variance



and possible because normalizing flows have exact likelihood evaluation.