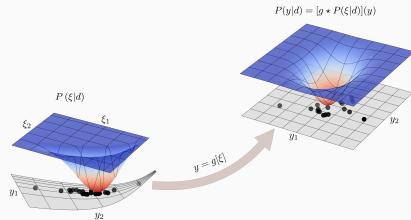


Variational Inference

APPROXIMATE BAYESIAN INFERENCE IN HIGH DIMENSIONS



Philipp Frank¹, Torsten Ensslin^{1,2}

iid2022, Lake Guntersville, AL, November 18, 2022

(1) Max-Planck Institute for Astrophysics MPA, Garching, Germany

(2) Ludwig-Maximilians University LMU, Munich, Germany



Posterior expectation

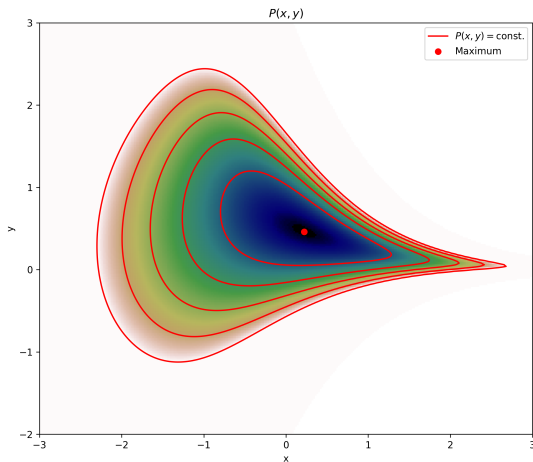
$$\mathbb{E}[f(\xi)]_{\mathcal{P}} = \int f \, d\mathcal{P} = \int f(\xi) \mathcal{P}(\xi|d) \, d\xi$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ ; data: d .

Posterior expectation

$$\mathbb{E}[f(\xi)]_{\mathcal{P}} = \int f \, d\mathcal{P} = \int f(\xi) \mathcal{P}(\xi|d) \, d\xi$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ ; data: d .

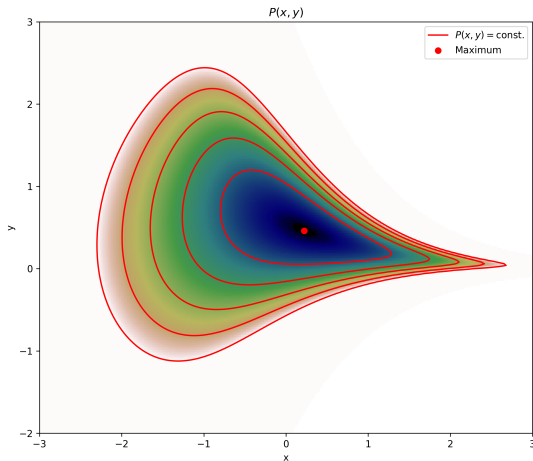


Probability measures

Posterior expectation

$$\mathbb{E}[f(\xi)]_{\mathcal{P}} = \int f \, d\mathcal{P} = \int f(\xi) \mathcal{P}(\xi|d) \, d\xi \approx f(\xi_{\text{MAX}})$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ ; data: d .



Variational Inference

Posterior expectation

$$\mathbb{E}[f(\xi)]_P = \int f \, d\mathcal{P} = \int f(\xi) \mathcal{P}(\xi|d) \, d\xi \approx f(\xi_{\text{MAX}})$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ ; data: d .

Posterior expectation

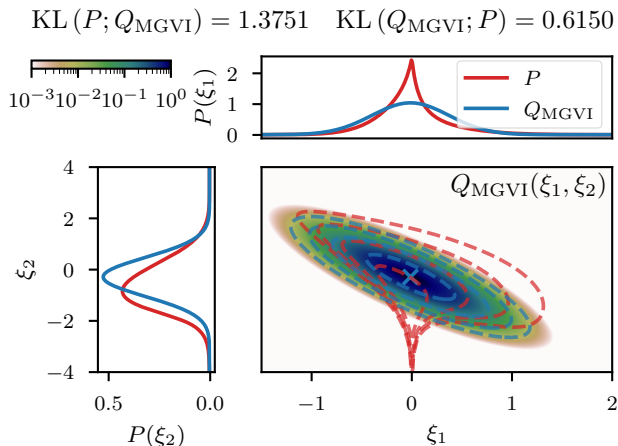
$$\mathbb{E}[f(\xi)]_P = \int f \, d\mathcal{P} = \int f(\xi) \mathcal{P}(\xi|d) \, d\xi \approx \int f(\xi) \mathcal{Q}(\xi|d) \, d\xi$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ ; data: d .

Posterior expectation

$$\mathbb{E}[f(\xi)]_P = \int f \, dP = \int f(\xi) \mathcal{P}(\xi|d) \, d\xi \approx \int f(\xi) \mathcal{Q}(\xi|d) \, d\xi$$

Function: $f(\xi)$; Posterior: $\mathcal{P}(\xi|d)$; parameters: ξ ; data: d .

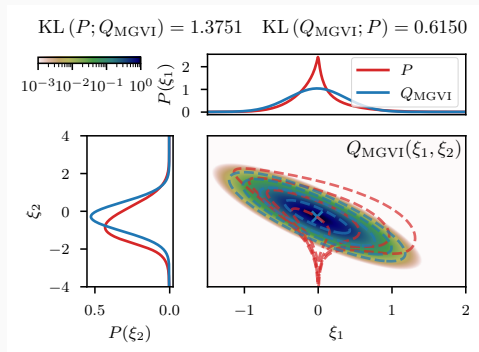


Kullback-Leibler divergence

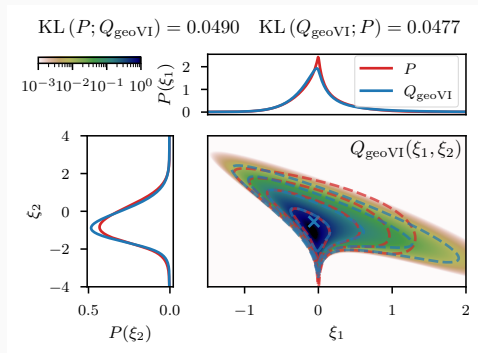
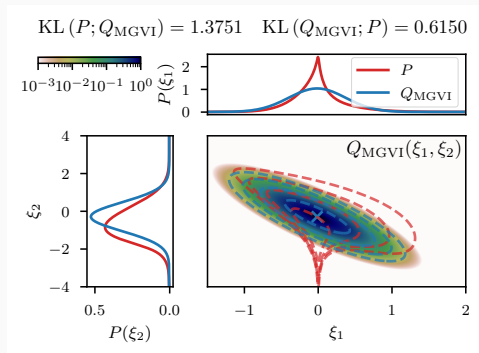
$$\text{KL}[\mathcal{Q}_\sigma || \mathcal{P}] = - \int \log \left(\frac{\mathcal{P}(\xi|d)}{\mathcal{Q}_\sigma(\xi)} \right) \mathcal{Q}_\sigma(\xi) \, \text{d}\xi$$

Posterior: $\mathcal{P}(\xi|d)$; Approximation: $\mathcal{Q}_\sigma(\xi)$; Variational parameters: σ .

Variational Inference (VI)



Variational Inference (VI)



Variational Inference (VI)

Kullback-Leibler divergence

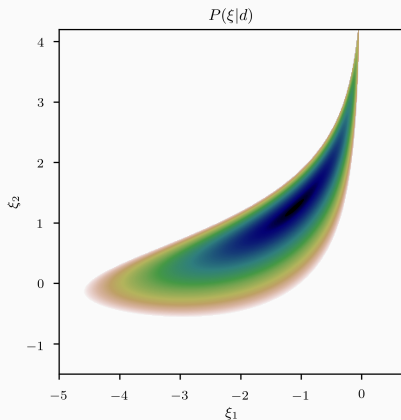
$$\text{KL}[\mathcal{Q}_\sigma || \mathcal{P}] = - \int \log \left(\frac{\mathcal{P}(\xi|d)}{\mathcal{Q}_\sigma(\xi)} \right) \mathcal{Q}_\sigma(\xi) \, d\xi$$

Posterior: $\mathcal{P}(\xi|d)$; Approximation: $\mathcal{Q}_\sigma(\xi)$; Variational parameters: σ .

Approximate distribution $\mathcal{Q}$: $\mathcal{Q}(\mathbf{y}) = \mathcal{N}(\mathbf{y}; 0, \mathbf{1})$

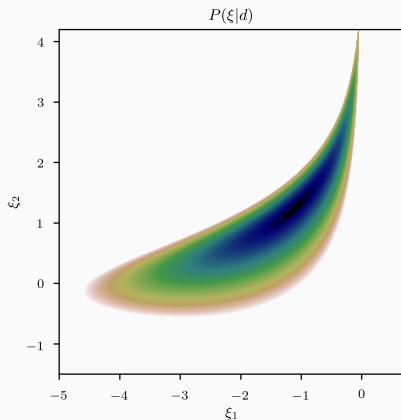
Coordinate system $\mathbf{y} = g_\sigma(\xi)$ such that the *posterior* is close to Normal.

Geometric Variational Inference (geoVI) [FLE21]



Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log(\mathcal{P}(\xi|d))$

Geometric Variational Inference (geoVI) [FLE21]

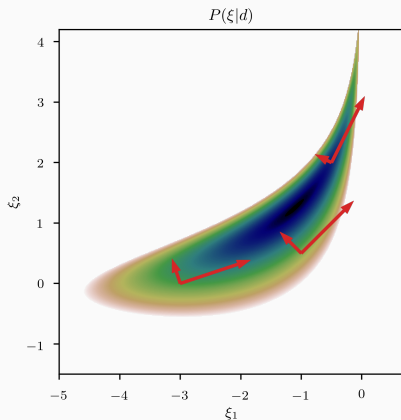


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log(\mathcal{P}(\xi|d))$

Posterior metric $\mathcal{M}(\xi)$: $\mathcal{M}_{\text{lh}}(\xi) + \mathbb{1}$

Fisher information metric $\mathcal{M}_{\text{lh}}(\xi)$: $\left\langle \frac{\partial^2 \mathcal{H}(d|\xi)}{\partial \xi \partial \xi'} \right\rangle_{\mathcal{P}(d|\xi)}$

Geometric Variational Inference (geoVI) [FLE21]

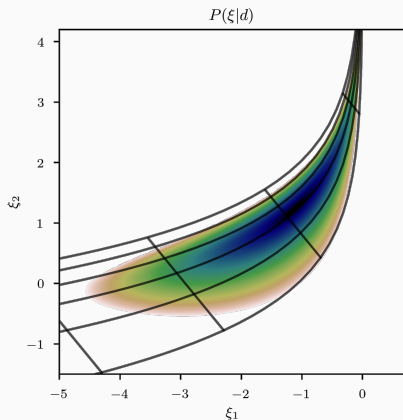


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log(\mathcal{P}(\xi|d))$

Posterior metric $\mathcal{M}(\xi)$: $\mathcal{M}_{\text{lh}}(\xi) + \mathbb{1}$

Fisher information metric $\mathcal{M}_{\text{lh}}(\xi)$: $\left\langle \frac{\partial^2 \mathcal{H}(d|\xi)}{\partial \xi \partial \xi'} \right\rangle_{\mathcal{P}(d|\xi)}$

Geometric Variational Inference (geoVI) [FLE21]

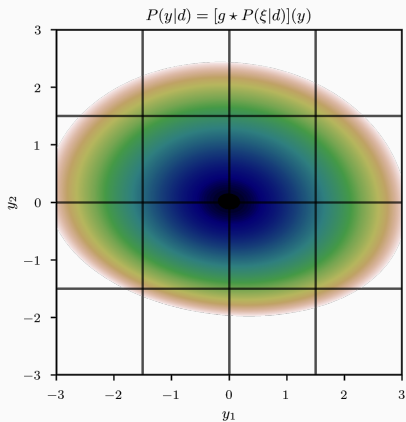
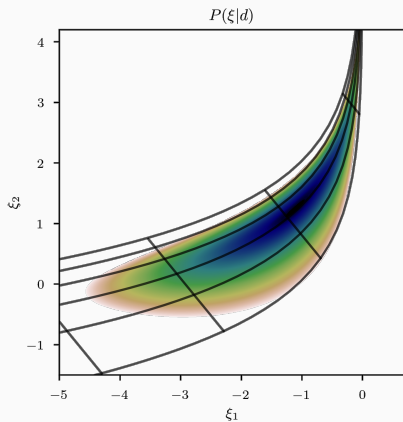


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log(\mathcal{P}(\xi|d))$

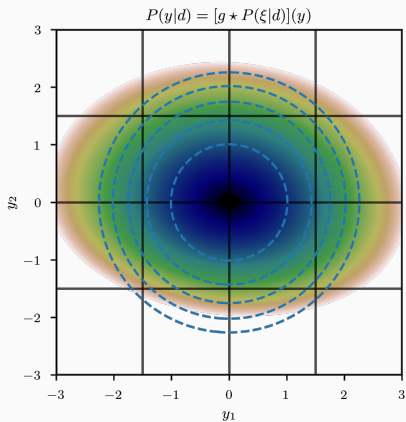
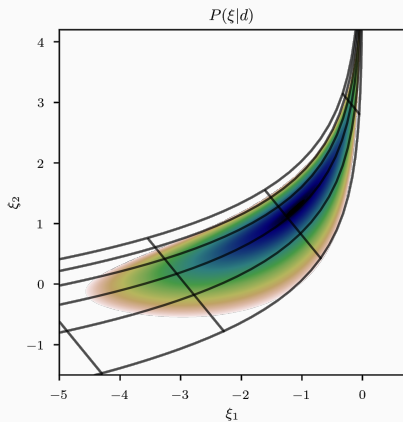
Posterior metric $\mathcal{M}(\xi)$: $\mathcal{M}_{\text{lh}}(\xi) + \mathbb{1}$

Fisher information metric $\mathcal{M}_{\text{lh}}(\xi)$: $\left\langle \frac{\partial^2 \mathcal{H}(d|\xi)}{\partial \xi \partial \xi'} \right\rangle_{\mathcal{P}(d|\xi)}$

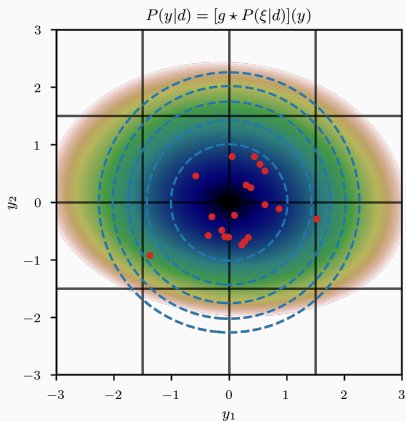
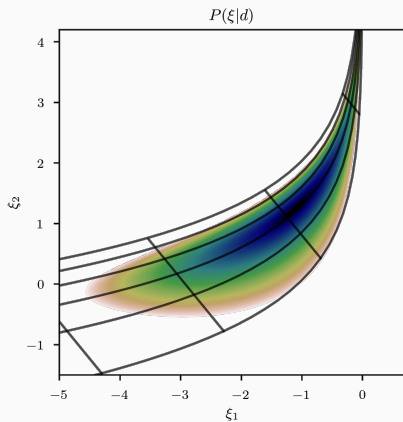
Geometric Variational Inference (geoVI) [FLE21]



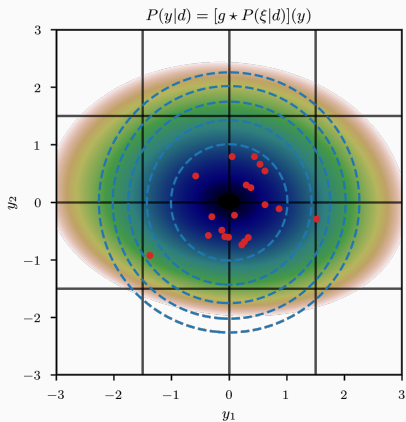
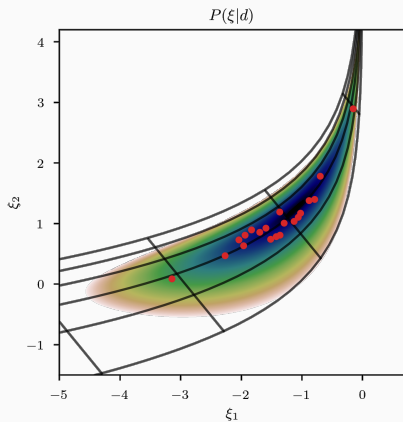
Geometric Variational Inference (geoVI) [FLE21]



Geometric Variational Inference (geoVI) [FLE21]



Geometric Variational Inference (geoVI) [FLE21]



Local Euclidean isometry around $\bar{\xi}$

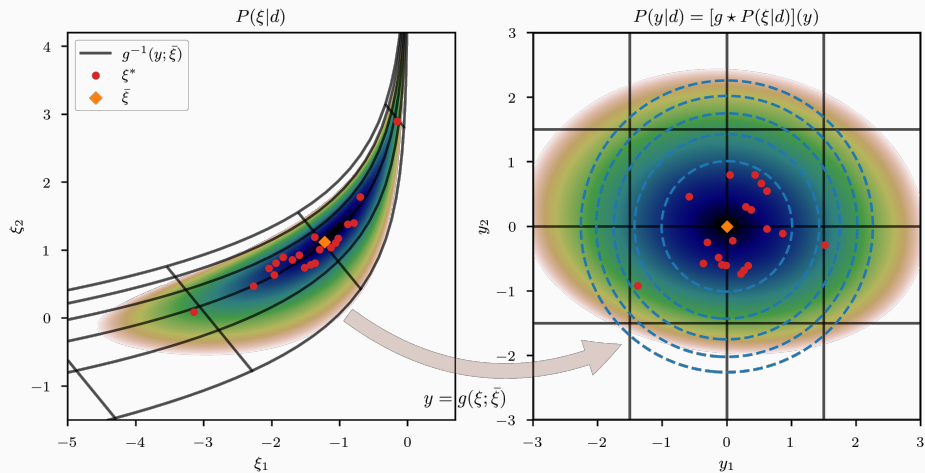
$$y = g(\xi; \bar{\xi}) = \mathcal{M}(\bar{\xi})^{-1/2} \left[\xi - \bar{\xi} + \left(\frac{\partial x}{\partial \xi} \right)^T \Big|_{\xi=\bar{\xi}} (x(\xi) - x(\bar{\xi})) \right]$$

Likelihood transformation: $x(\xi) = x(s(\xi))$, expansion point: $\bar{\xi}$.

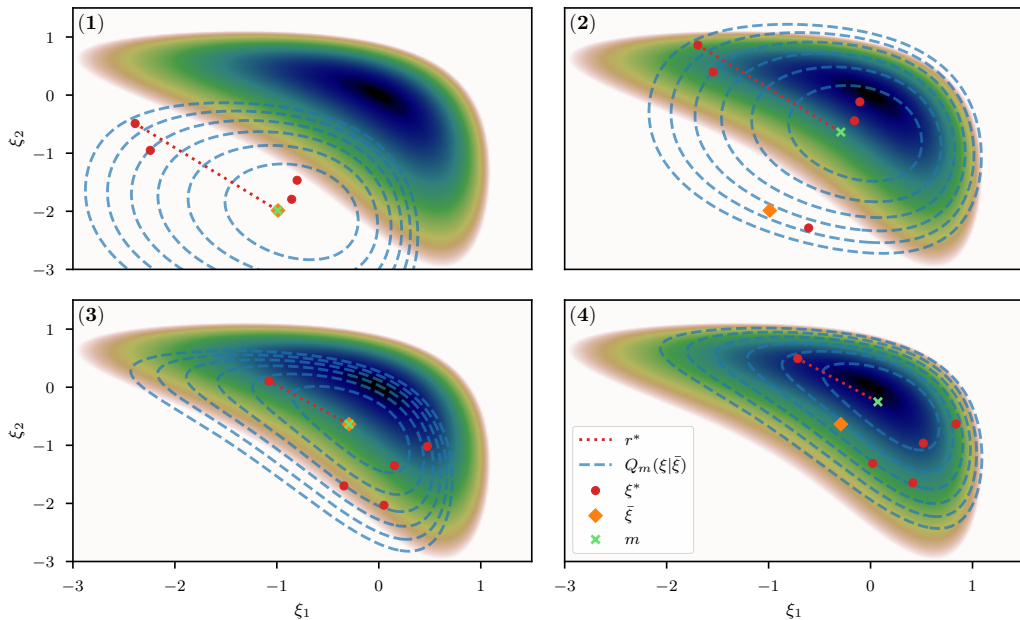
Variational approximation with transformed distribution $\mathcal{Q}$

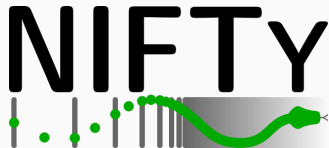
$$\mathcal{Q}_{\bar{\xi}}(\xi) = \mathcal{N}(y|0, \mathbb{1})|_{y=g(\xi; \bar{\xi})} \left\| \frac{\partial g(\xi; \bar{\xi})}{\partial \xi} \right\|$$

Geometric Variational Inference (geoVI) [FLE21]



Geometric Variational Inference (geoVI) [FLE21]





Numerical information field theory

<https://gitlab.mpcdf.mpg.de/ift/nifty>

NIFTy tutorial - this afternoon!

https://gitlab.mpcdf.mpg.de/ift/tutorial_nifty_resolve



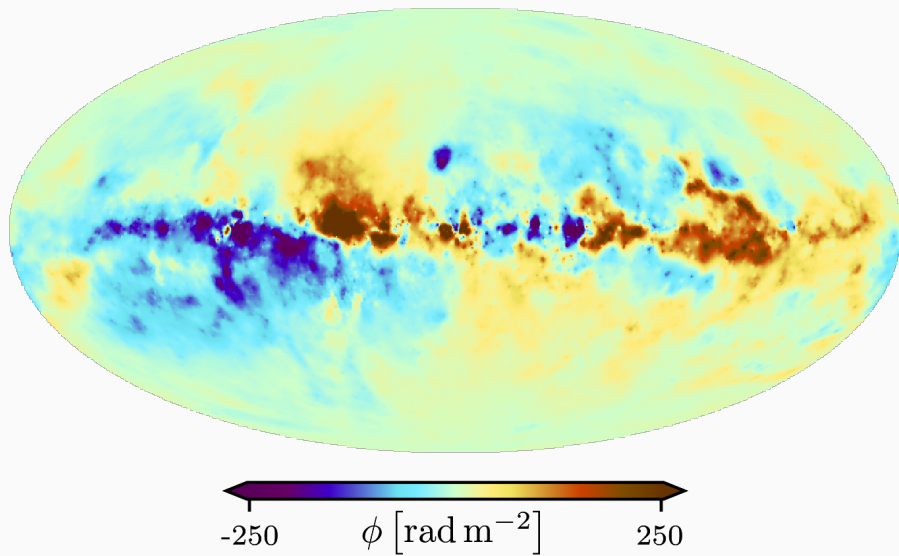
Philipp Frank, Reimar Leike, and Torsten A. Enßlin.

Geometric variational inference.

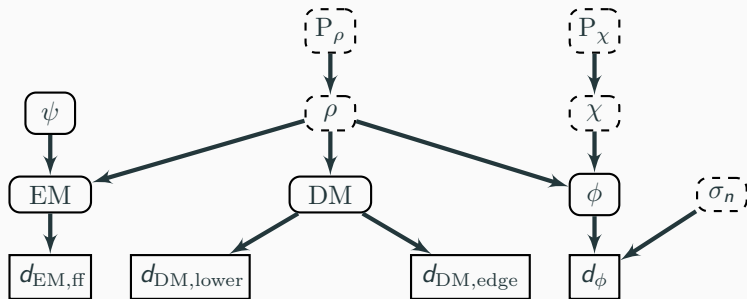
Entropy, 23(7), 2021.

Appendix

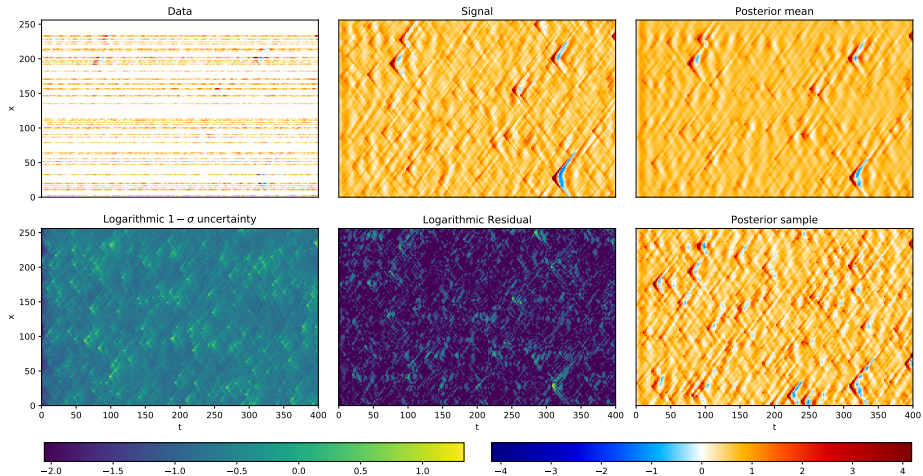
Application



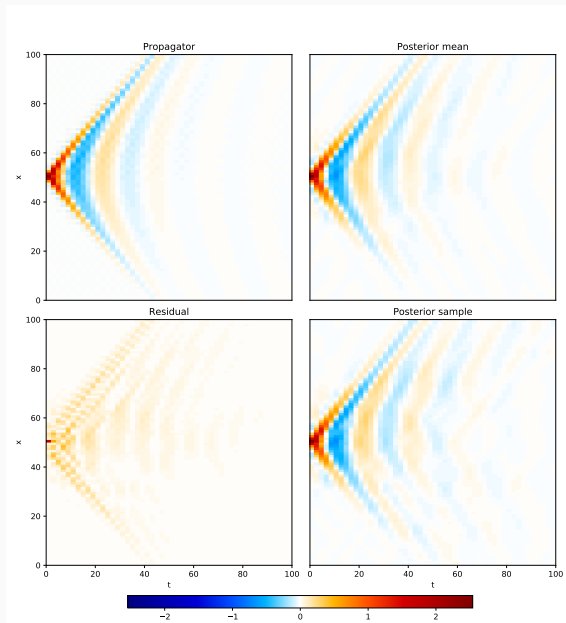
Application



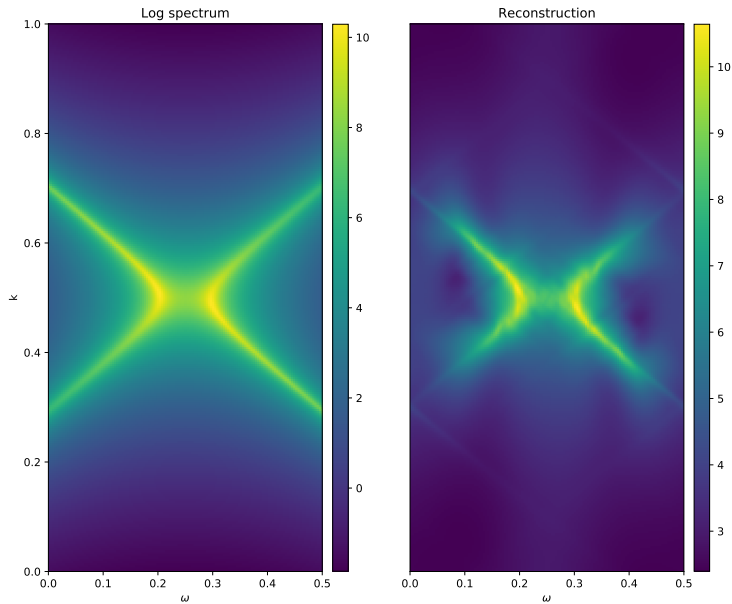
Application



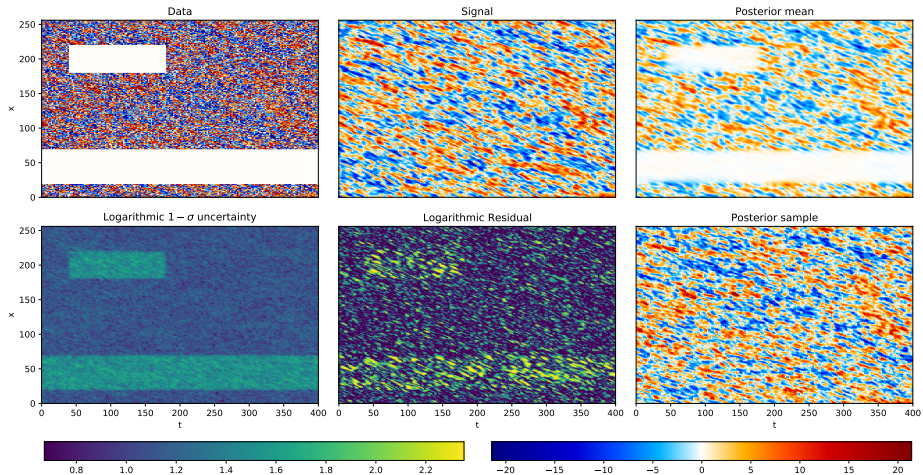
Application



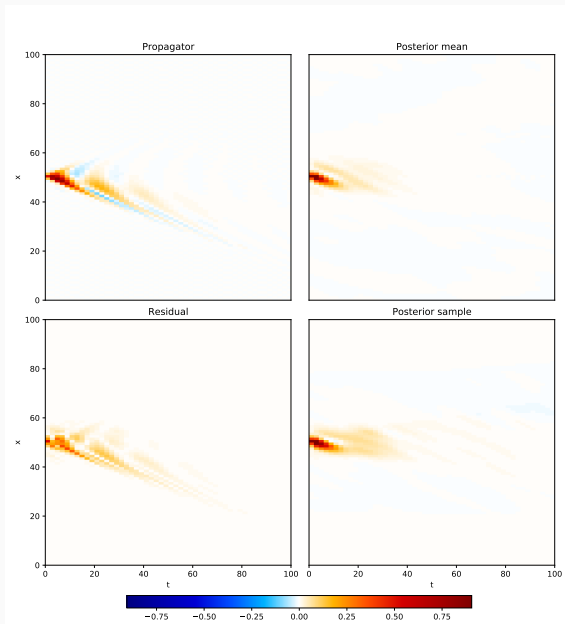
Application



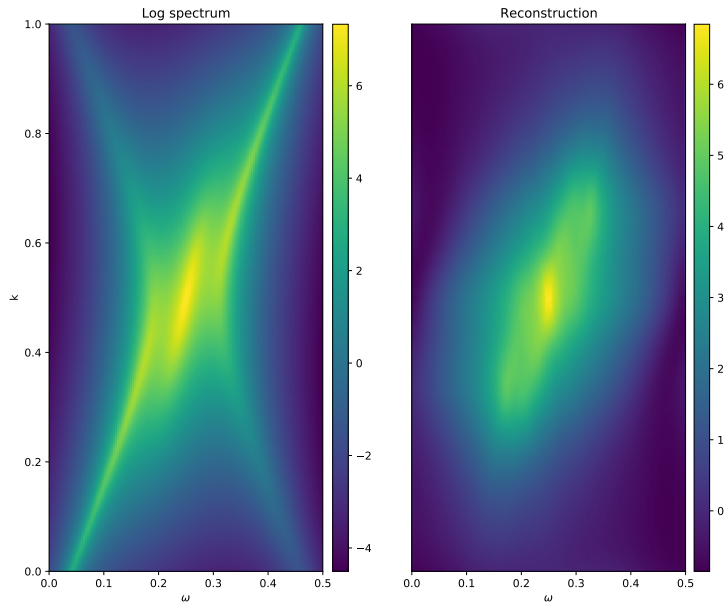
Application



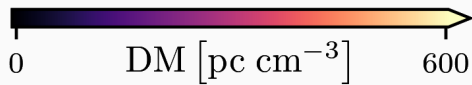
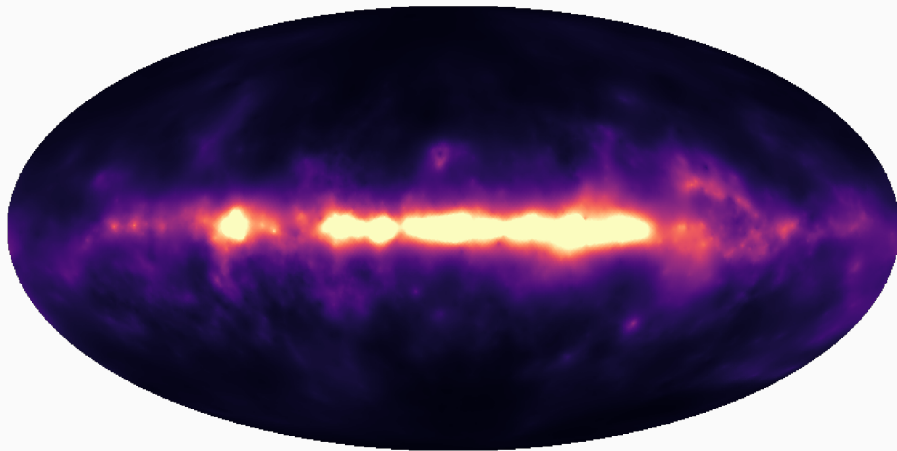
Application



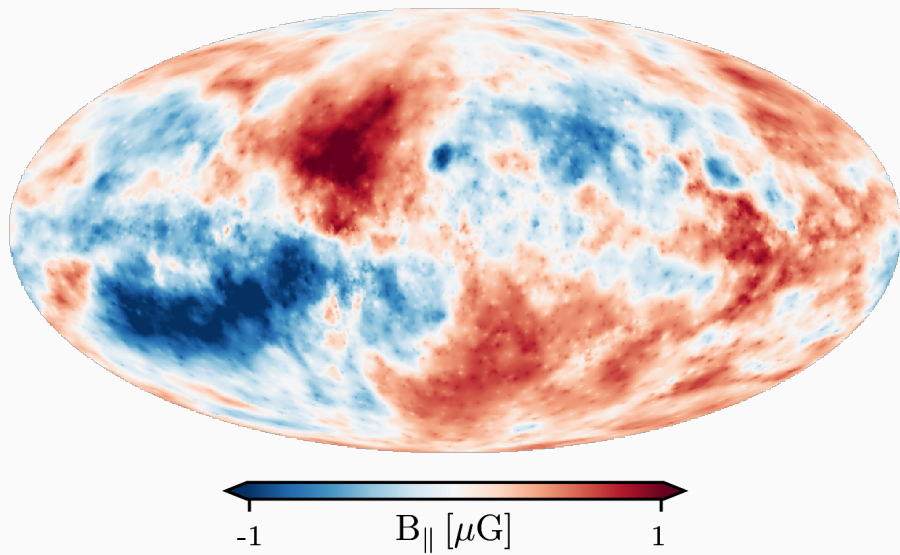
Application



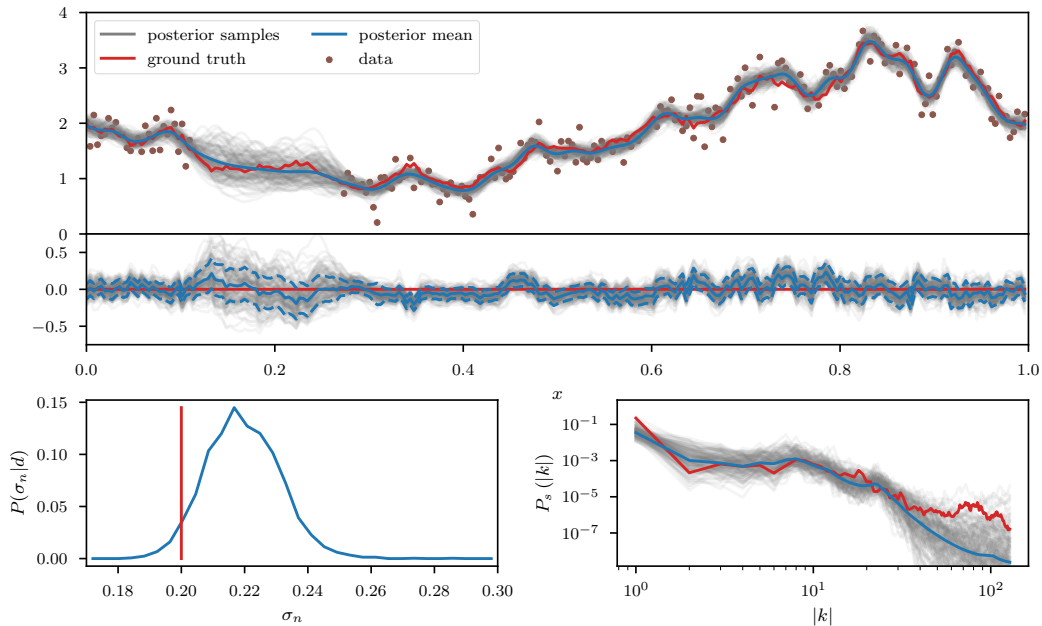
Appendix



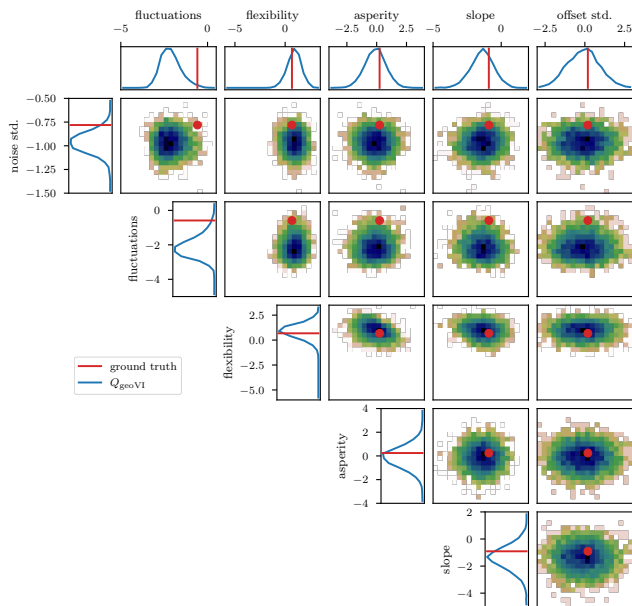
Appendix



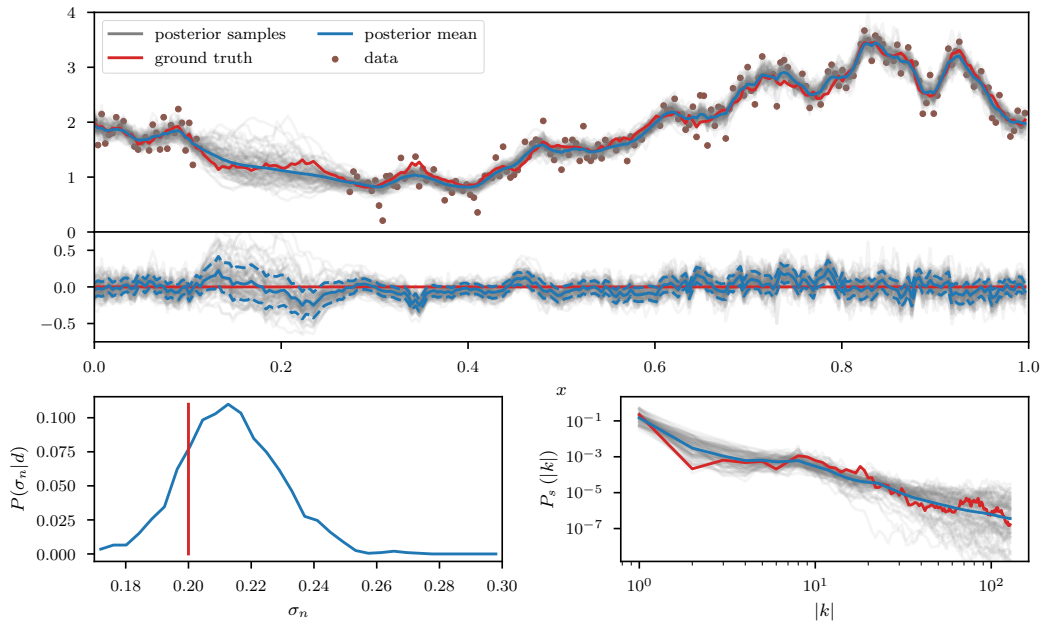
Appendix



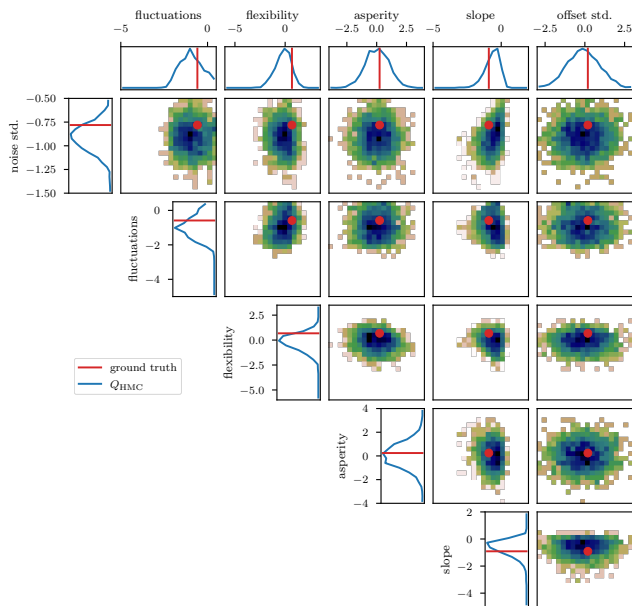
Appendix



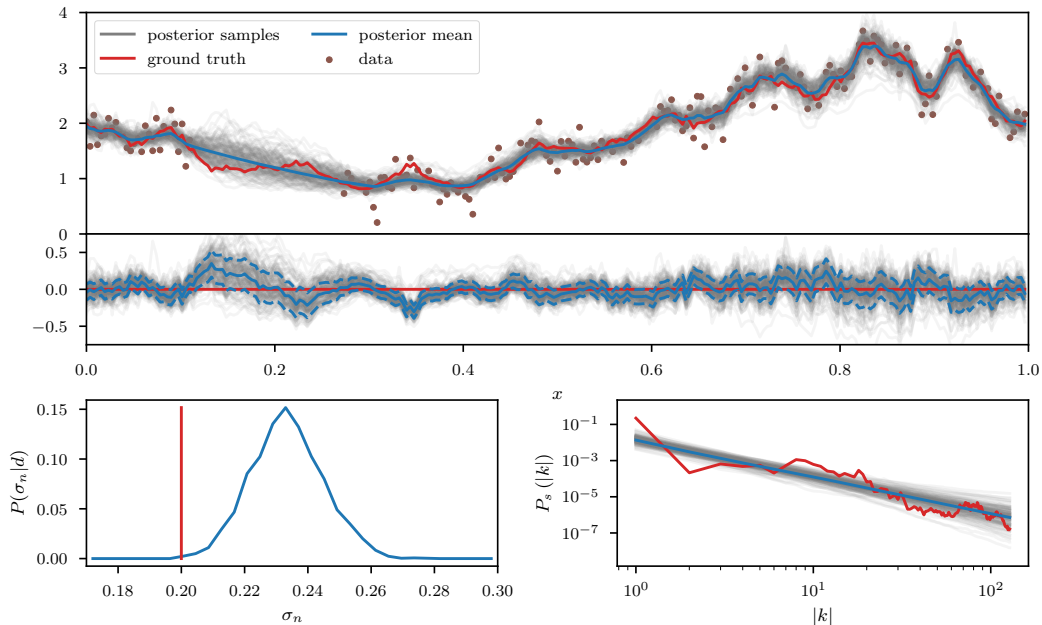
Appendix



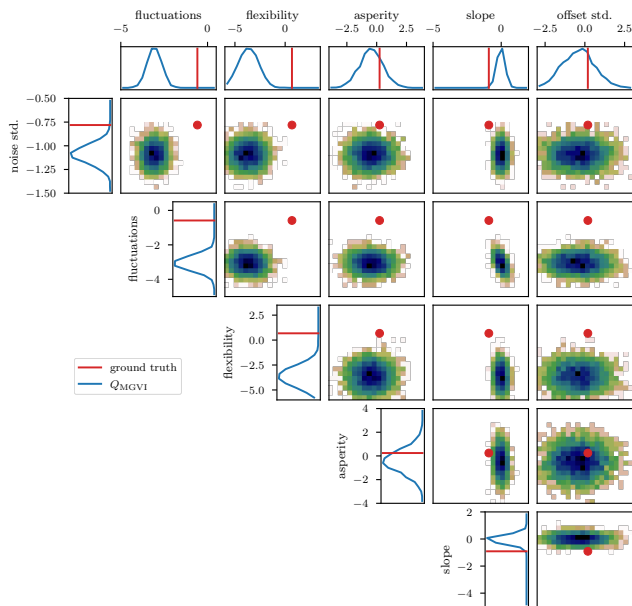
Appendix



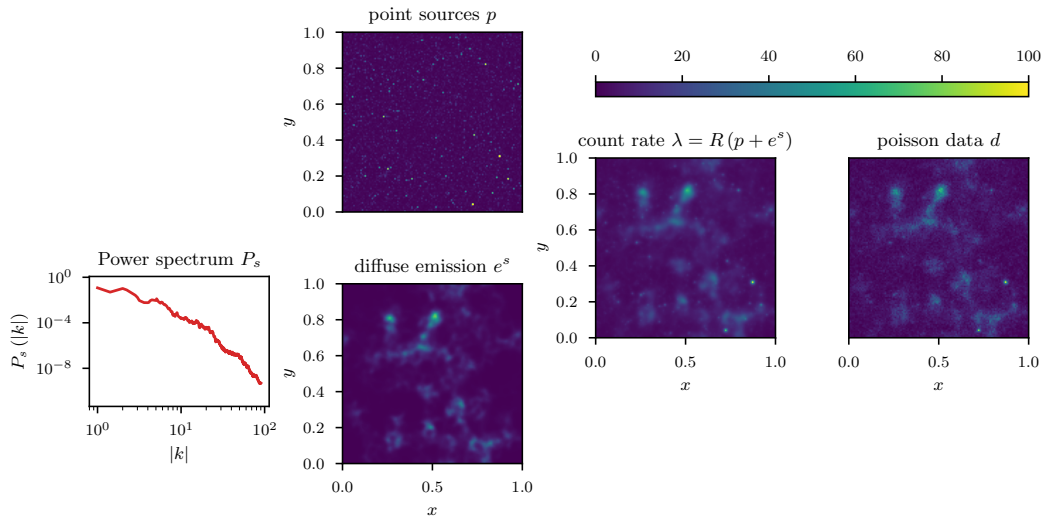
Appendix



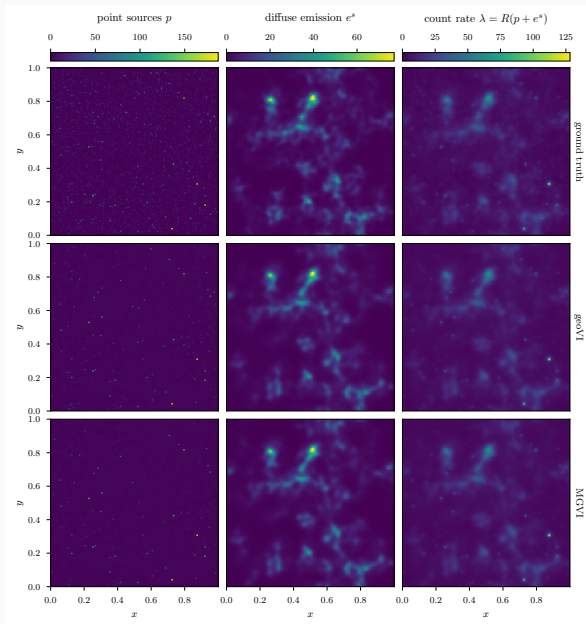
Appendix



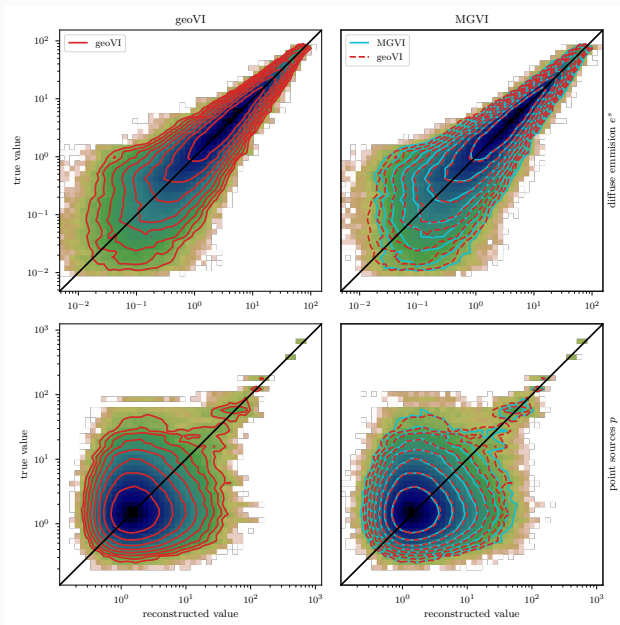
Appendix



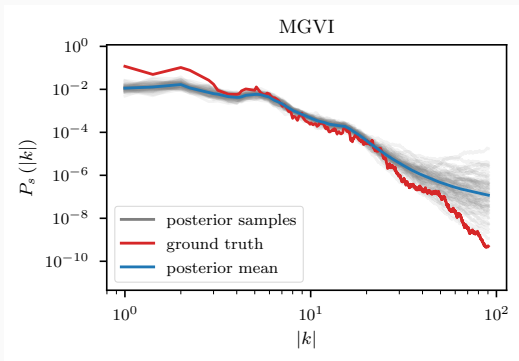
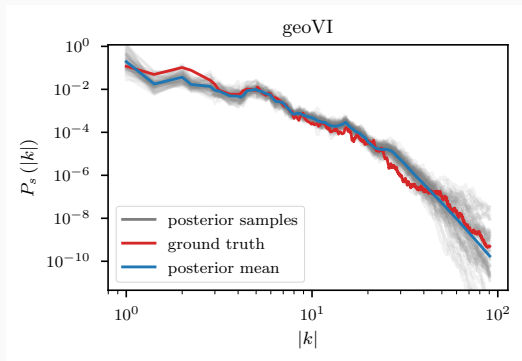
Appendix



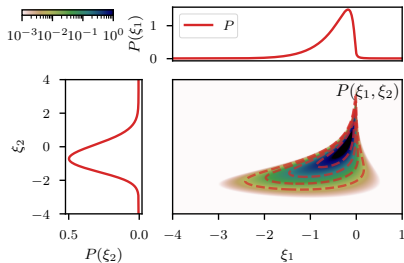
Appendix



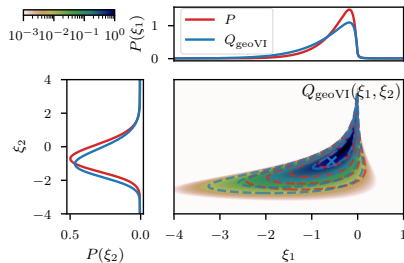
Appendix



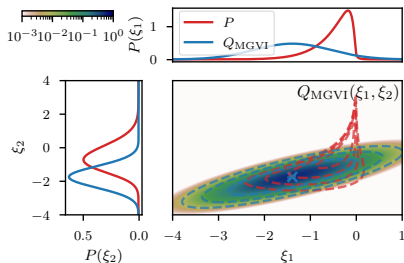
Appendix



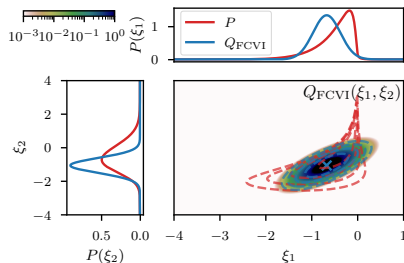
$$\text{KL}(P; Q_{\text{geoVI}}) = 0.0655 \quad \text{KL}(Q_{\text{geoVI}}; P) = 0.0888$$



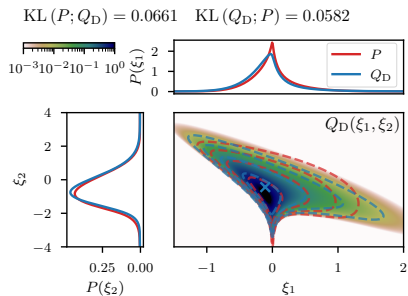
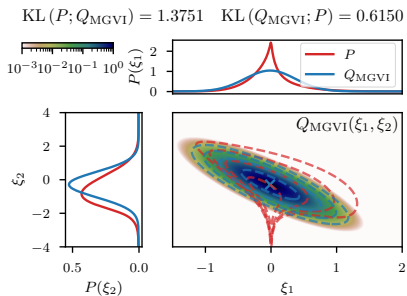
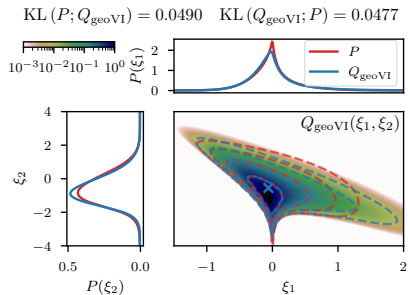
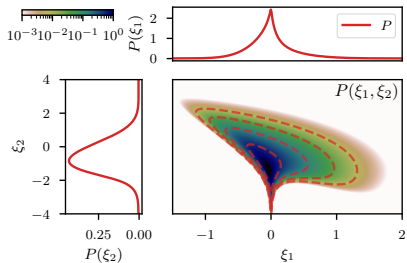
$$\text{KL}(P; Q_{\text{MGVI}}) = 3.3487 \quad \text{KL}(Q_{\text{MGVI}}; P) = 2.0713$$



$$\text{KL}(P; Q_{\text{FCVI}}) = 1.9748 \quad \text{KL}(Q_{\text{FCVI}}; P) = 0.5765$$

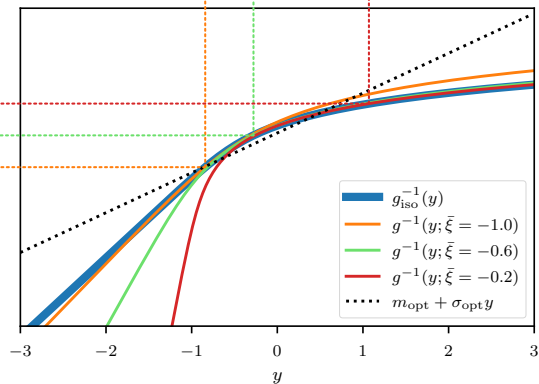
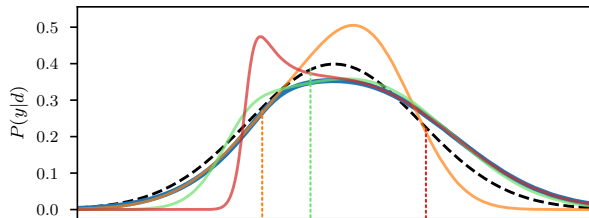
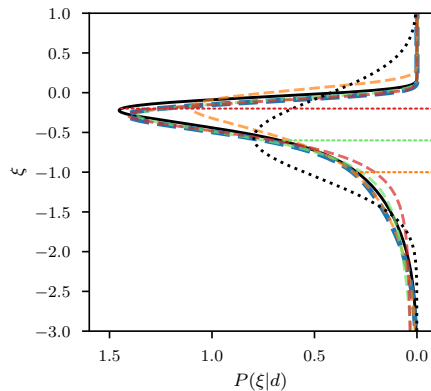


Appendix



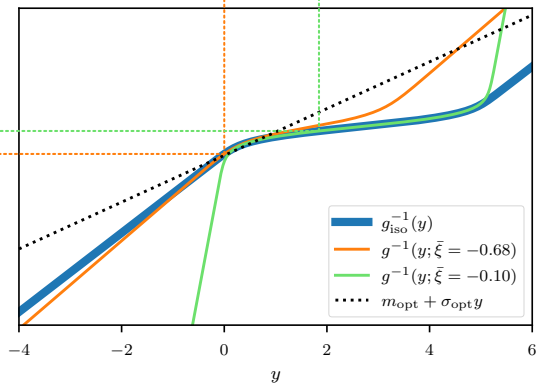
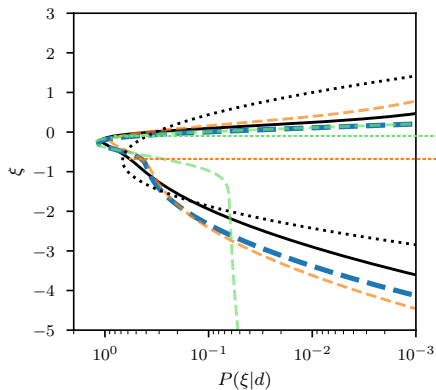
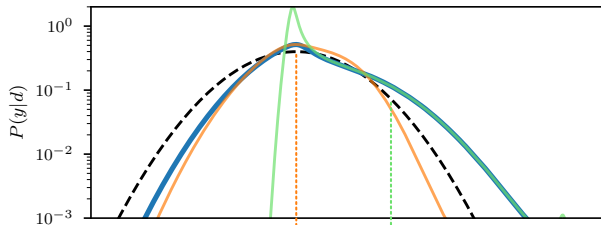
Appendix

$$\begin{aligned}
 \text{KL}(P; Q_{\text{iso}}) &= 0.0333 \\
 \text{KL}(P; Q_{\bar{\xi}=-1.0}) &= 0.0582 \\
 \text{KL}(P; Q_{\bar{\xi}=-0.6}) &= 0.0489 \\
 \text{KL}(P; Q_{\bar{\xi}=-0.2}) &= 0.1557 \\
 \text{KL}(P; Q_{\text{Normal}}) &= 0.2864
 \end{aligned}$$



Appendix

$$\begin{aligned}\text{KL}(P; Q_{\text{iso}}) &= 0.0806 \\ \text{KL}(P; Q_{\bar{\xi}=-0.68}) &= 0.0642 \\ \text{KL}(P; Q_{\bar{\xi}=-0.10}) &= 0.5656 \\ \text{KL}(P; Q_{\text{Normal}}) &= 0.1817\end{aligned}$$



Appendix

