Table 6: Ablation study for mini-batch optimal transport, better performance in bold

Metrics	Density-MSE	Autocovariance-MSE	Spectra-MSE	Convergent L^2 Loss	NFE
Independent	$3.4 \cdot 10^{-5}$	$\mathbf{7.4 \cdot 10^{-5}}$	$1.3 \cdot 10^{1}$	$5.1 \cdot 10^{-2}$	168
mini-batch = 64	$\mathbf{9.8 \cdot 10^{-6}}$	$9.8 \cdot 10^{-5}$	$\mathbf{6.3 \cdot 10^{0}}$	$\mathbf{1.9 \cdot 10^{-2}}$	**153**

Table 7: Scaling study for the size of mini-batch given optimal transport plan, best performance in bold

Metrics	Density-MSE	Autocovariance-MSE	Spectra-MSE	Convergent L^2 Loss	NFE
mini-batch = 32	$3.0 \cdot 10^{-5}$	$1.3 \cdot 10^{-4}$	$5.0 \cdot 10^{0}$	$2.3 \cdot 10^{-2}$	**141**
mini-batch = 64	$\mathbf{9.8 \cdot 10^{-6}}$	$9.8 \cdot 10^{-5}$	$6.3 \cdot 10^{0}$	$1.9 \cdot 10^{-2}$	153
mini-batch = 128	$2.4 \cdot 10^{-5}$	$\mathbf{1.7 \cdot 10^{-5}}$	$\mathbf{3.6 \cdot 10^{0}}$	$\mathbf{1.5 \cdot 10^{-2}}$	182

Table 8: Scaling study for the number of runs of Hutchinson trace estimator

	$n_{\text{run}} = 4$	$n_{\text{run}} = 8$	$n_{\text{run}} = 16$	$n_{\text{run}} = 32$	$n_{\text{run}} = 64$	$n_{\text{run}} = 128$	exact
sample 1	-481.7 ± 9.4	-482.9 ± 7.7	-480.6 ± 4.9	-481.4 ± 3.8	-481.1 ± 2.5	-480.8 ± 1.8	-482.7
sample 2	-482.8 ± 9.8	-481.8 ± 6.7	-481.3 ± 4.9	-480.8 ± 3.5	-481.1 ± 2.7	-481.2 ± 1.7	-482.0
sample 3	-479.7 ± 10.2	-479.7 ± 7.6	-478.2 ± 4.7	-478.6 ± 3.4	-478.8 ± 2.6	-478.8 ± 1.7	-479.1
sample 4	-476.9 ± 10.0	-477.0 ± 6.6	-477.9 ± 5.6	-478.5 ± 4.0	-478.1 ± 2.7	-478.0 ± 1.8	-477.4
sample 5	-479.4 ± 10.7	-479.2 ± 6.6	-479.2 ± 5.3	-479.4 ± 3.6	-479.6 ± 2.8	-479.3 ± 2.0	-478.5

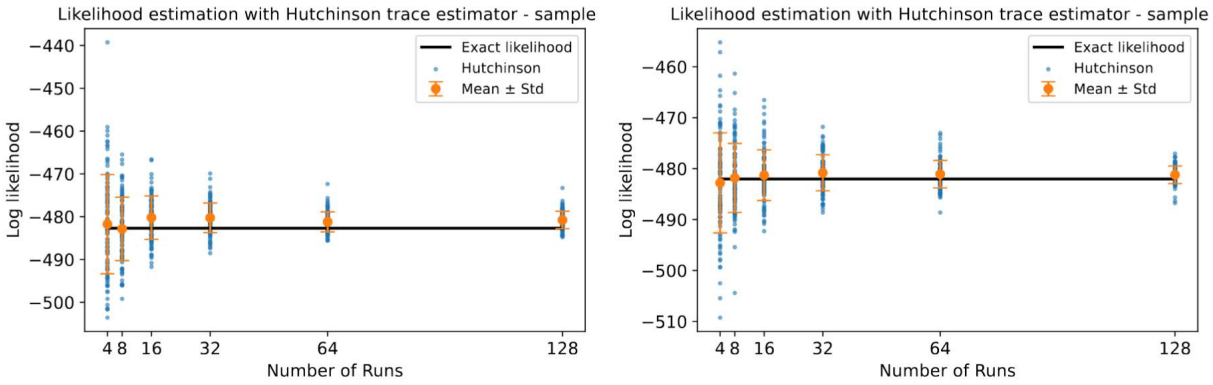
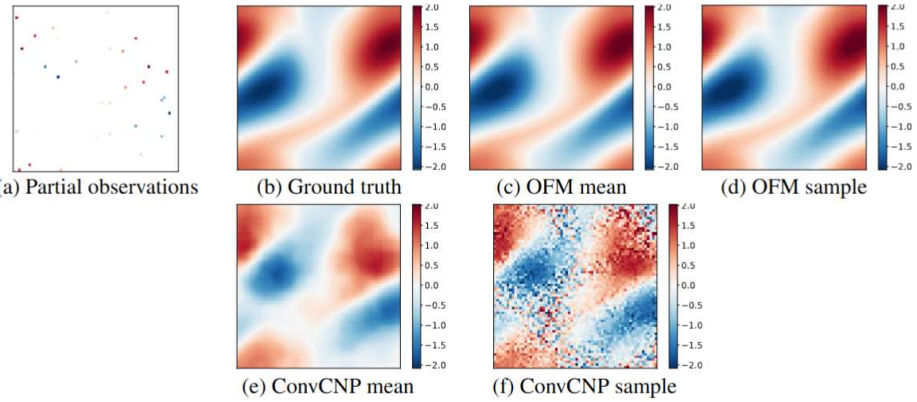


Figure 24: Log likelihood by the integration of divergence, (**left**) plot for first GP sample. (**right**) plot for second GP sample

Table 1: Comparison of OFM with baseline models: GP regression; OpFlow (Shi et al., 2024a); Neural Processes ((Garnelo et al., 2018), NP); Attentive NP ((Kim et al., 2019), ANP); Convolutional Conditional NP ((Gordon et al., 2020), ConvCNP); Deep variational GP ((Salimbeni & Deisenroth, 2017), DGP); Deep Sigma Point Process ((Jankowiak et al., 2020), DSSP); Metrics SMSE and MSLL used for 1D and 2D GP example. Mean squared error for the predicted mean (μ) and standard deviation (σ) are used for TGP example. Performance of GP regression for 1D and 2D GP are removed (marked with '$-$'), which are taken as the ground truth. Best performance in bold.

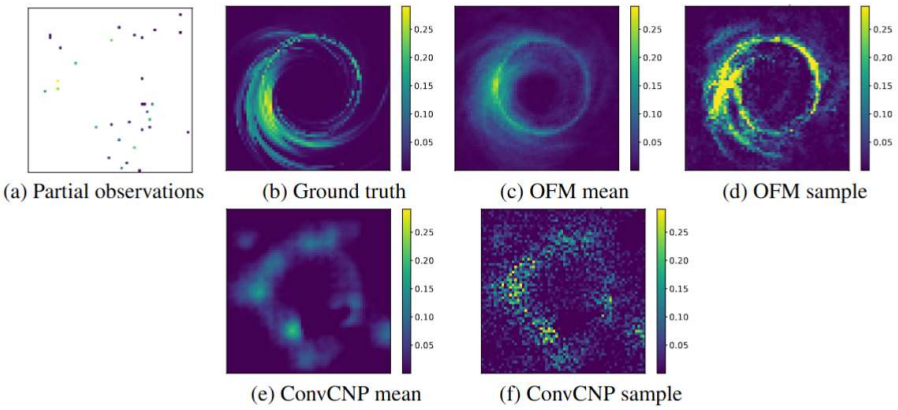
Dataset $\rightarrow$	1D GP		2D GP		1D TGP	
Algorithm $\downarrow$ Metric $\rightarrow$	SMSE	MSLL	SMSE	MSLL	μ	σ
GP prior	-	-	-	-	$6.4 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$
OpFlow	$5.0 \cdot 10^{-1}$	$2.0 \cdot 10^{-1}$	$1.4 \cdot 10^{-1}$	$\mathbf{1.1 \cdot 10^{-1}}$	$1.3 \cdot 10^{-2}$	$3.9 \cdot 10^{-3}$
NP	$6.1 \cdot 10^{-1}$	$4.5 \cdot 10^{0}$	$1.7 \cdot 10^{-1}$	$2.1 \cdot 10^{0}$	$1.0 \cdot 10^{-1}$	$1.9 \cdot 10^{-2}$
ANP	$5.1 \cdot 10^{-1}$	$9.8 \cdot 10^{-1}$	$1.6 \cdot 10^{-1}$	$1.1 \cdot 10^{0}$	$1.4 \cdot 10^{-1}$	$1.7 \cdot 10^{-2}$
ConvCNP	$5.6 \cdot 10^{-1}$	$2.7 \cdot 10^{-1}$	$1.7 \cdot 10^{-1}$	$4.5 \cdot 10^{-1}$	$1.6 \cdot 10^{-2}$	$2.1 \cdot 10^{-3}$
DGP	$4.1 \cdot 10^{-1}$	$6.8 \cdot 10^{-2}$	$1.8 \cdot 10^{0}$	$4.2 \cdot 10^{0}$	$4.9 \cdot 10^{-1}$	$1.4 \cdot 10^{-2}$
DSPP	$4.7 \cdot 10^{-1}$	$6.5 \cdot 10^{0}$	$1.9 \cdot 10^{-1}$	$6.6 \cdot 10^{0}$	$1.1 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$
OFM	$\mathbf{4.1 \cdot 10^{-1}}$	$\mathbf{5.5 \cdot 10^{-2}}$	$\mathbf{1.3 \cdot 10^{-1}}$	$1.6 \cdot 10^{-1}$	$\mathbf{5.2 \cdot 10^{-3}}$	$\mathbf{9.5 \cdot 10^{-4}}$



(a) Partial observations (b) Ground truth (c) OFM mean (d) OFM sample

(e) ConvCNP mean (f) ConvCNP sample

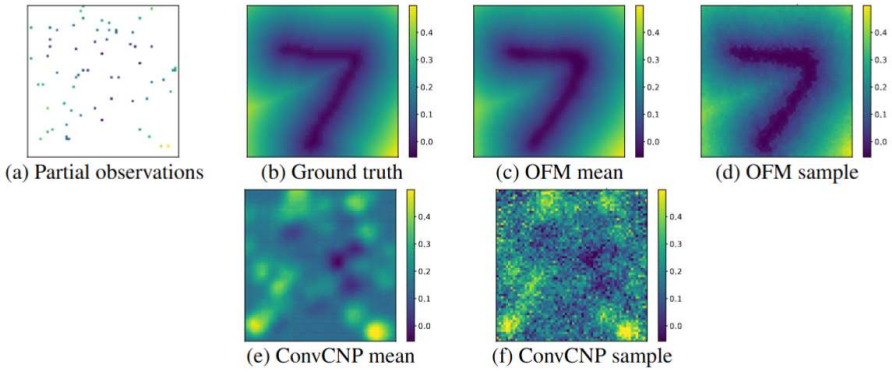
Figure 25: Operator Flow Matching (OFM) regression on Navier-Stokes functional data with resolution 64×64. (a) 32 random observations. (b) Ground truth sample (c) Predicted mean from OFM. (d) One posterior sample from OFM. (e) Predicted mean from ConvCNP. (f) One posterior sample from ConvCNP.



(a) Partial observations (b) Ground truth (c) OFM mean (d) OFM sample

(e) ConvCNP mean (f) ConvCNP sample

Figure 26: OFM regression on black hole data with resolution 64×64. (a) 32 random observations. (b) Ground truth sample. (c) Predicted mean from OFM. (d) One posterior sample from OFM. (e) Predicted mean from ConvCNP. (f) One posterior sample from ConvCNP.



(a) Partial observations (b) Ground truth (c) OFM mean (d) OFM sample

(e) ConvCNP mean (f) ConvCNP sample

Figure 27: OFM regression on MNIST-SDF with resolution 64×64. (a) 64 random observations. (b) Ground truth sample. (c) Predicted mean from OFM. (d) One posterior sample from OFM. (e) Predicted mean from ConvCNP. (f) One posterior sample from ConvCNP.