

Deep and Multi-fidelity learning with Gaussian processes

Andreas Damianou

Amazon Research, Cambridge UK

Talk at the Alan Turing Institute, 7 Aug. 2019



Various parts of this talk come from work with:

Kurt Cutajar

Paris Perdikaris

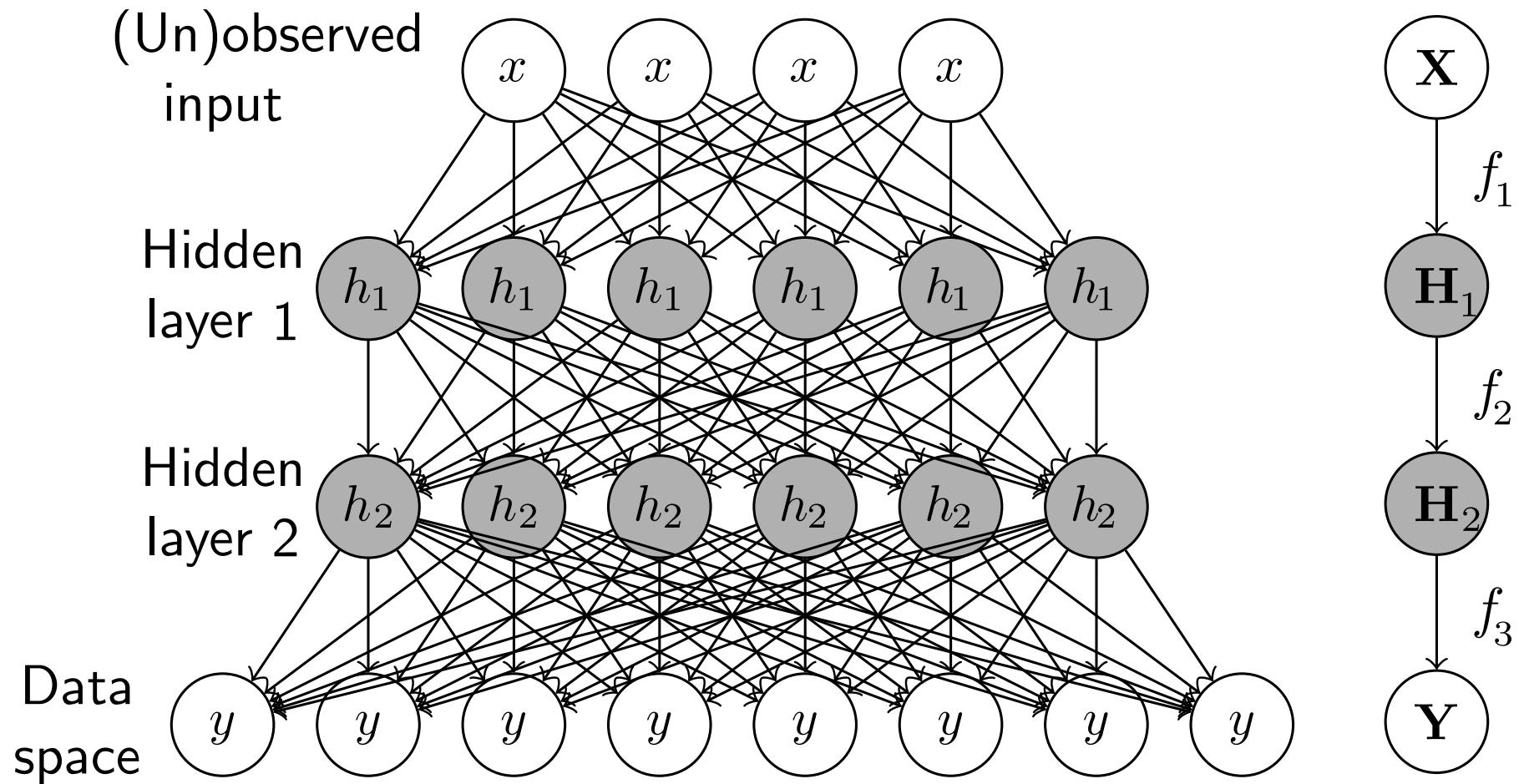
Mark Pullin

Javier Gonzalez

Neil Lawrence

Part 1: Deep Gaussian processes

A general *family* of probabilistic models

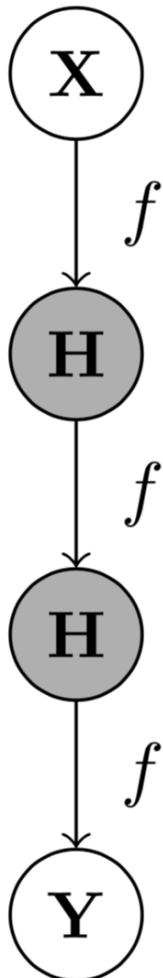


$$\mathbf{Y} = f_3(f_2(\cdots f_1(\mathbf{X}))), \quad 8/7/19$$

$$\mathbf{H}_i = f_i(\mathbf{H}_{i-1})$$

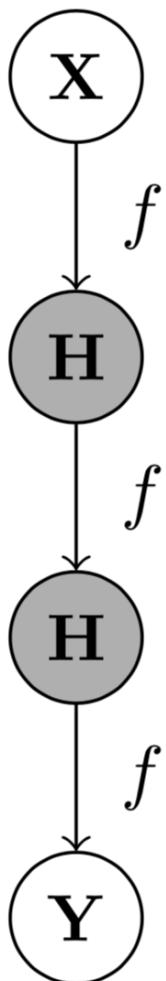
A. Damianou

Deep Gaussian process



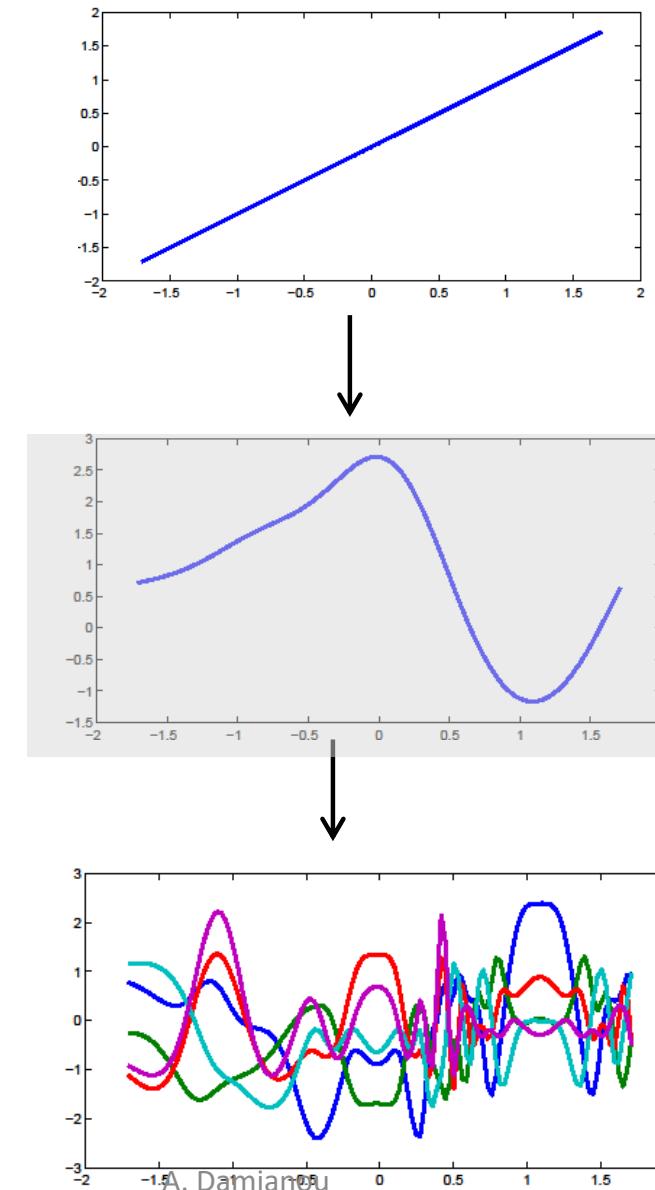
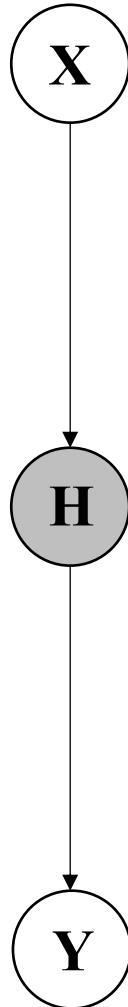
- ▶ Nested function composition
- ▶ Non-parametric, non-linear mappings f
- ▶ Mappings f marginalized out analytically
- ▶ NOT a GP!

Deep Gaussian process



- ▶ Objective: $p(y|x) = \int_{h_2} \left(p(y|h_2) \int_{h_1} p(h_2|h_1)p(h_1|x) \right)$
- ▶ $p(h_2|x) = \int_{h_1, f_2} p(h_2|f_2)p(f_2|h_1)p(h_1|x)$

Sampling from a Deep GP



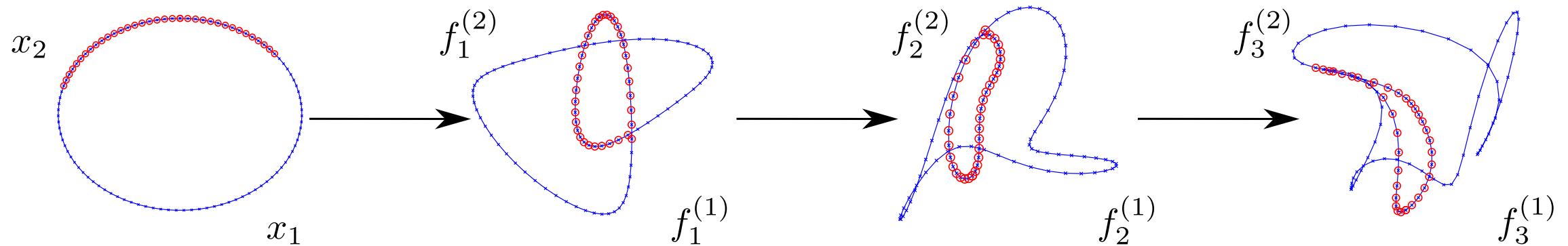
Input

Unobserved

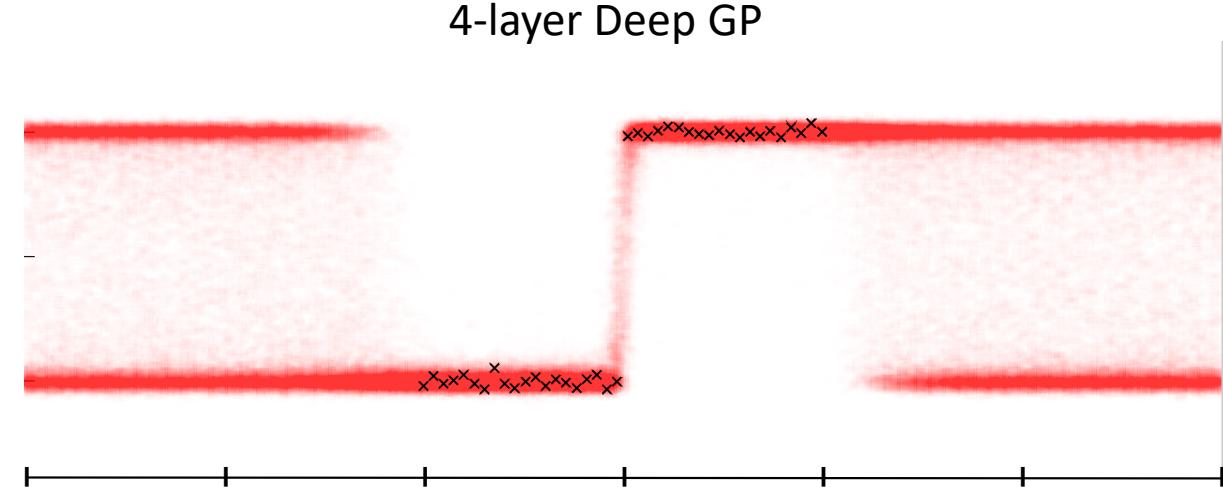
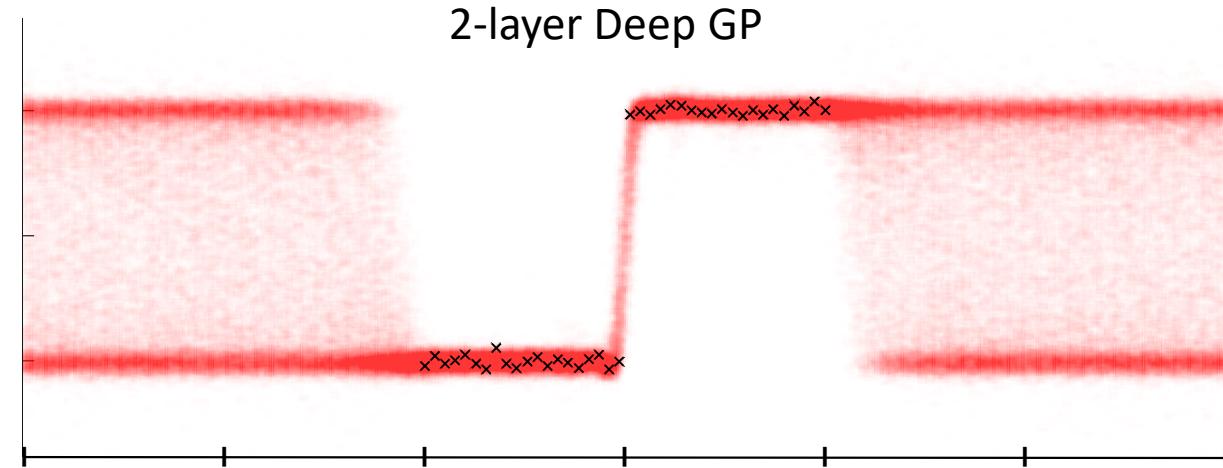
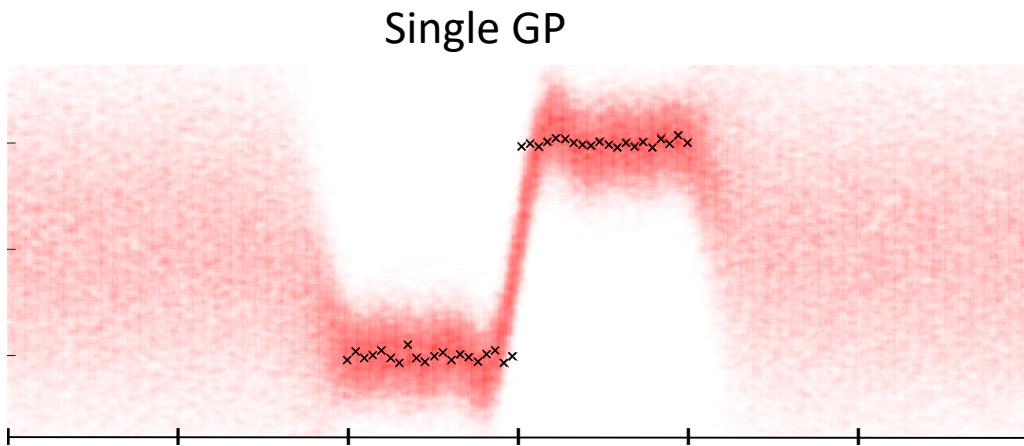
Output

Feature learning

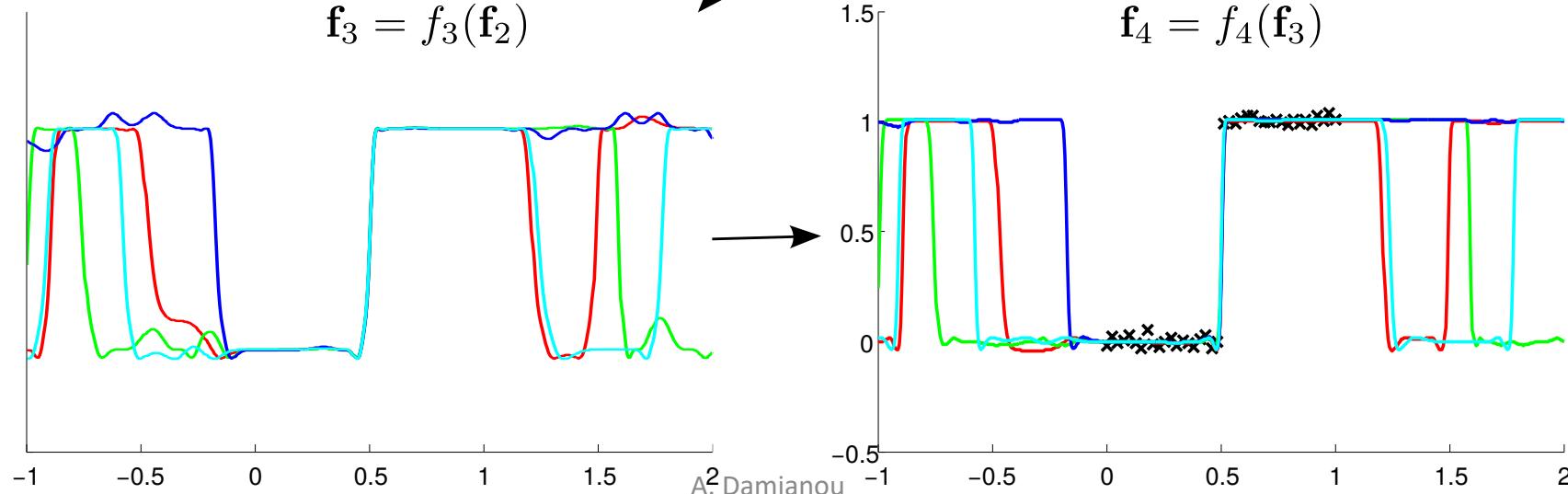
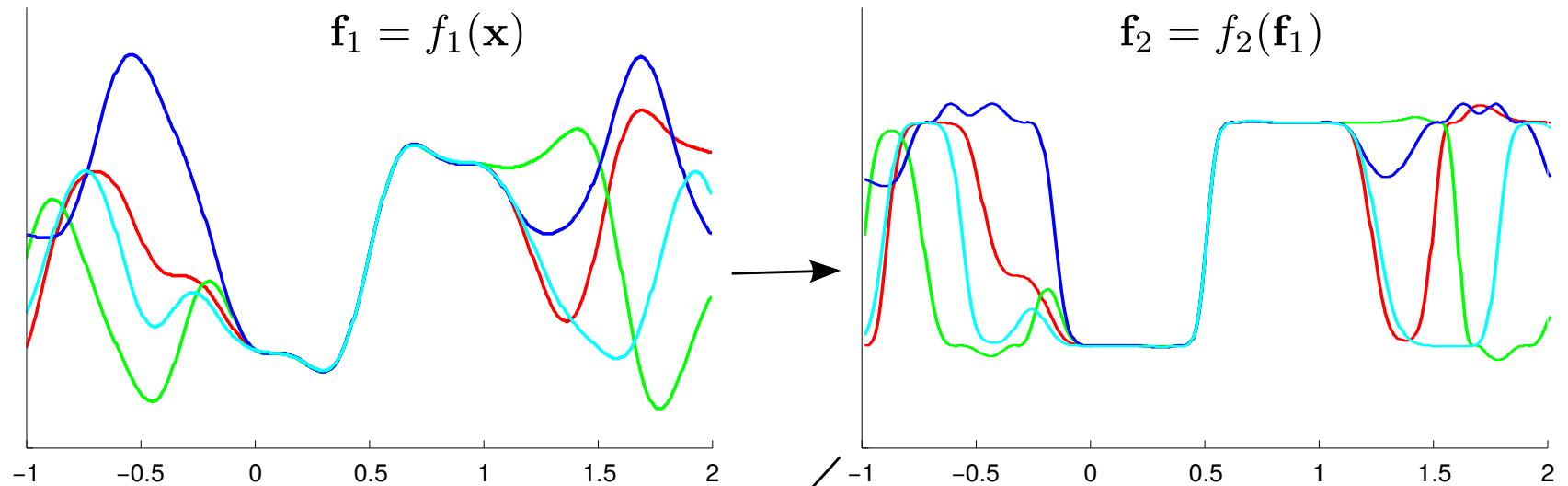
Regularities are learned as “knots” in the latent space, carried over from layer to layer.



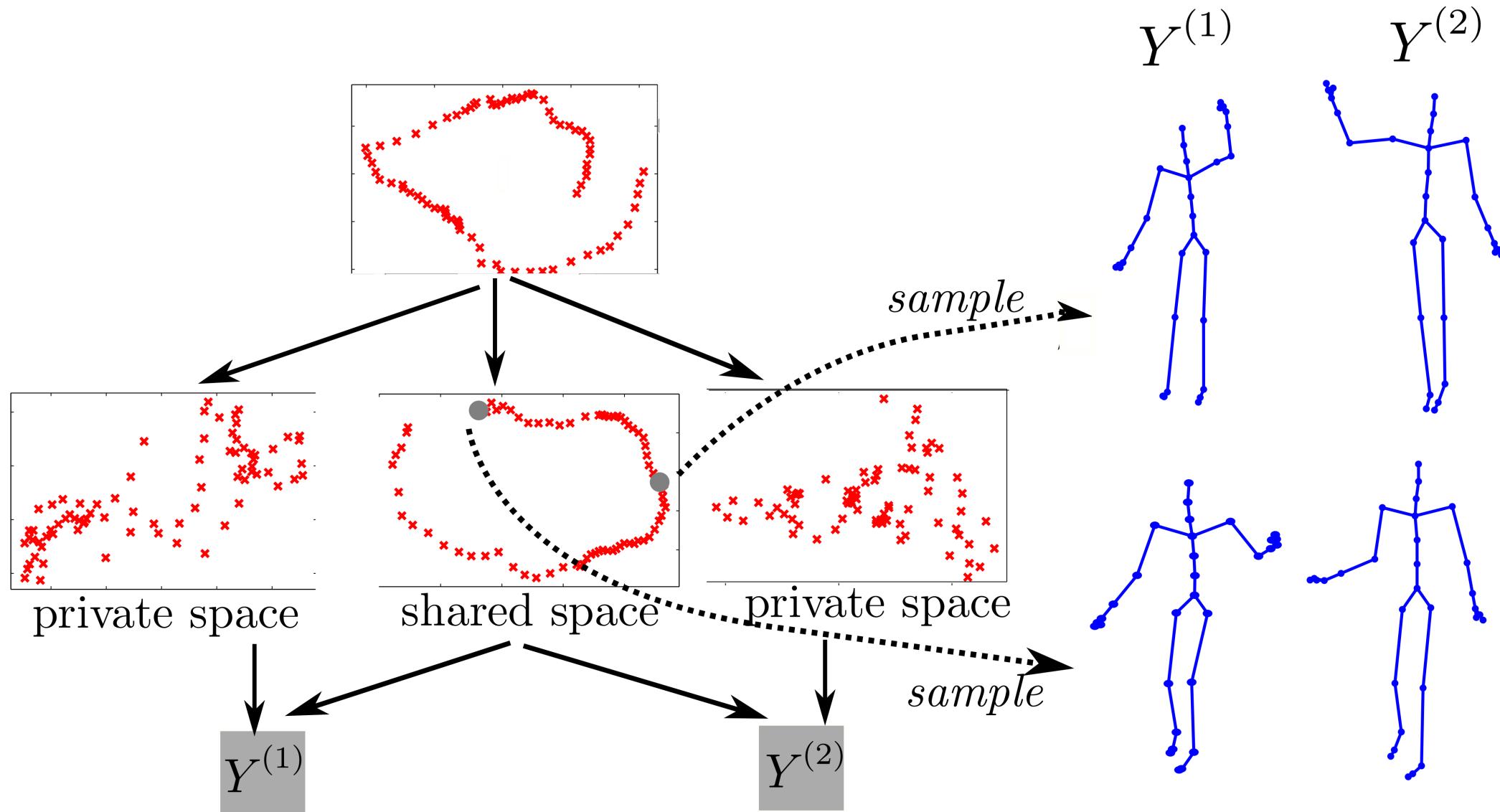
Step function example



Successive warping to learn the step function



Unsupervised learning for multiple data views



Properties

- ▶ Unsupervised learning possible due to Bayesian regularization
 - ▶ Very data efficient
 - ▶ Scalability also possible with newer techniques
-
- ▶ Intractable objective
 - ▶ Classification is more challenging

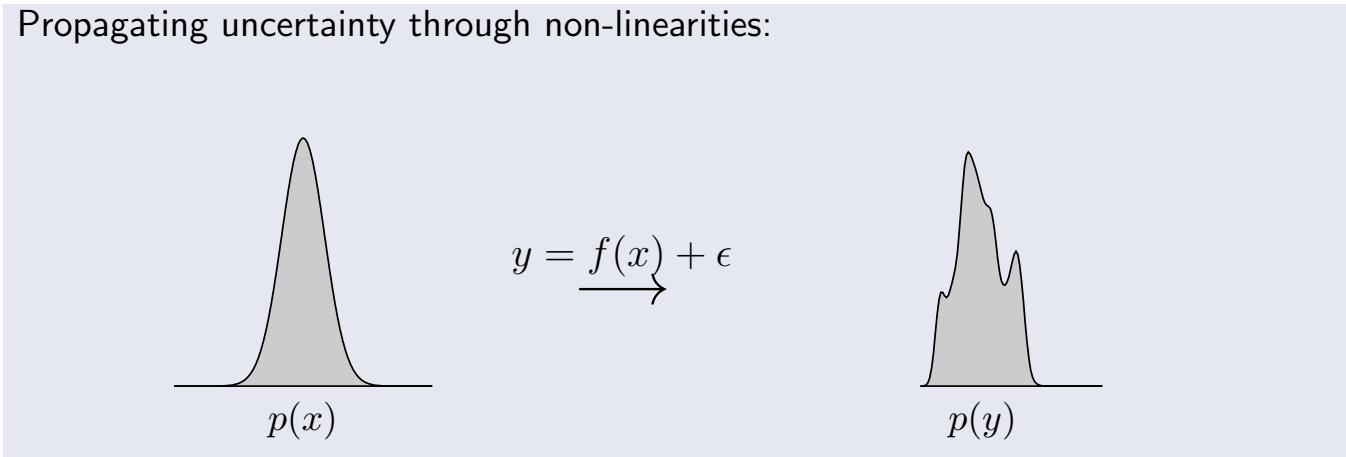
Properties

- ▶ Unsupervised learning possible due to Bayesian regularization
 - ▶ Very data efficient
 - ▶ Scalability also possible with newer techniques
-
- ▶ Intractable objective
 - ▶ Classification is more challenging

Inference in Deep GPs: uncertainty propagation

- Objective: $p(y|x) = \int_{h_2} \left(p(y|h_2) \underbrace{\int_{h_1} p(h_2|h_1)p(h_1|x)}_{\text{contains}} \right)$
- $p(h_2|x) = \int_{h_1, f_2} p(h_2|f_2) \underbrace{p(f_2|h_1)}_{(k(h_1, h_1))^{-1}} p(h_1|x)$

Propagating uncertainty through non-linearities:



Variational bound and its properties

Bound on the log marginal likelihood $\log p(y)$

$$\mathcal{F} = \overbrace{\sum_{l=2}^{L+1} \left\langle \sum_{n=1}^N \mathcal{L}(\mathbf{h}_l^{(n)}, \mathbf{u}_l) \right\rangle_{\mathcal{Q}}}^{\text{Data fit}} - \sum_{l=2}^{L+1} \text{KL}(q(\mathbf{u}_l) \| p(\mathbf{u}_l)) \underbrace{- \text{KL}(q(\mathbf{h}_1) \| p(\mathbf{h}_1))}_{\text{Regularization}} + \sum_{l=2}^L \underbrace{\mathcal{H}(q(\mathbf{h}_l))}_{\text{Regularization}}$$

Part 2: Multi-fidelity modeling

Multi-fidelity data

High fidelity observations



Low fidelity observations



High fidelity simulations



Low fidelity simulations



Multi-fidelity data

High fidelity observations



Low fidelity observations



High fidelity simulations

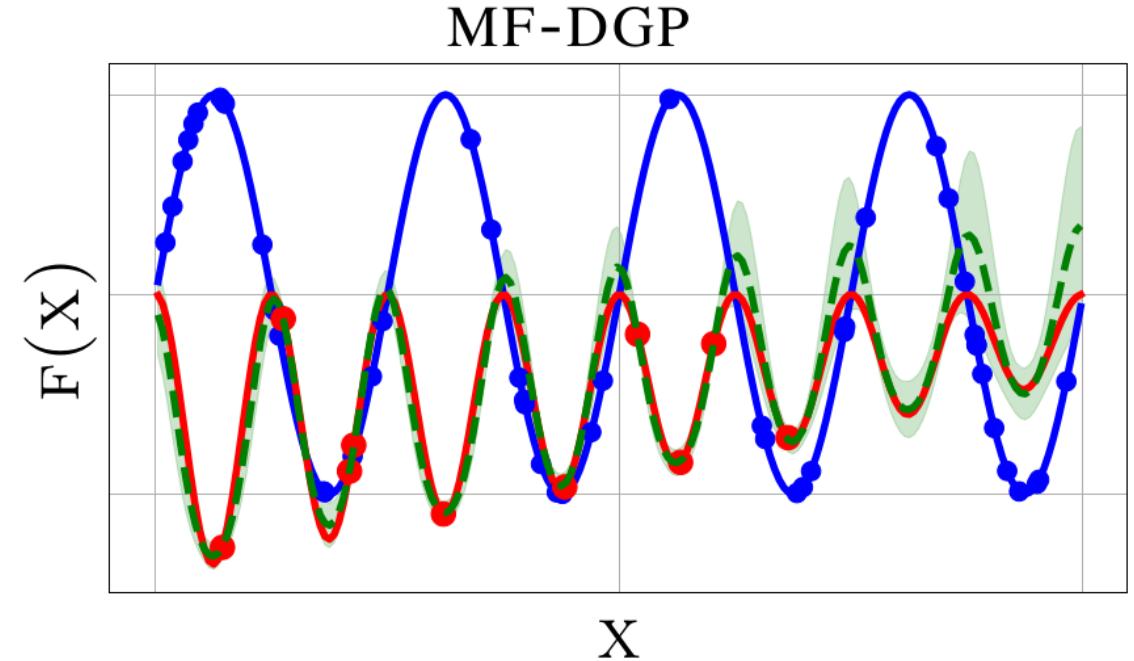
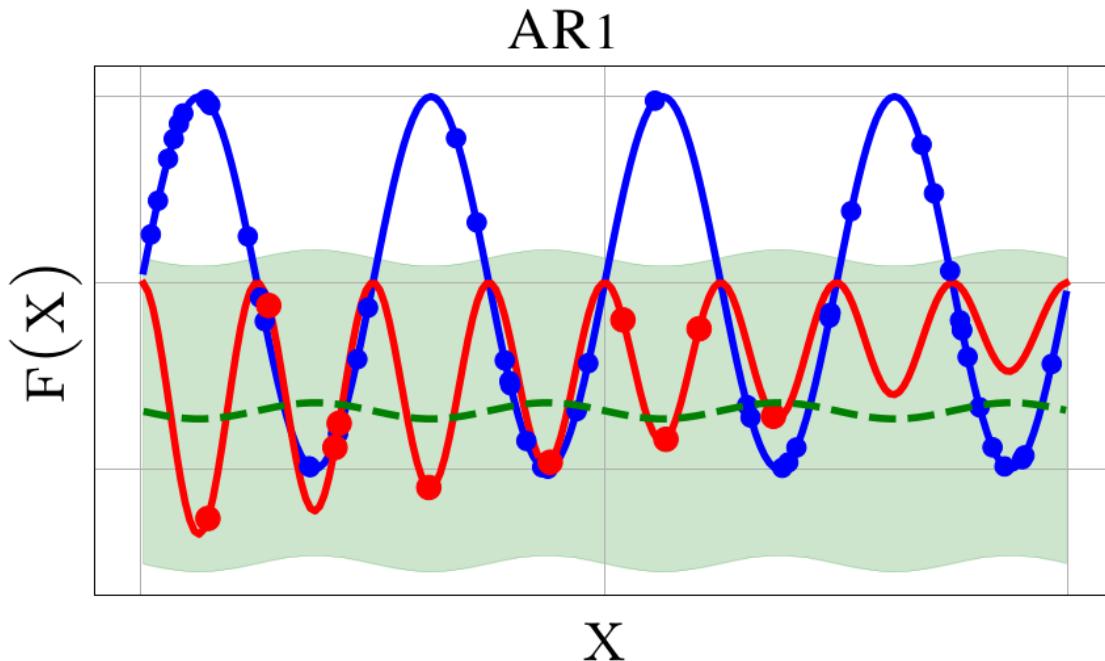


Low fidelity simulations



Fusing information from multiple fidelities

■■■ PREDICTED HIGH-FIDELITY ■■■ HIGH-FIDELITY ■■■ LOW-FIDELITY



We want to trust the high-fidelity data, where we have them, and where we don't have them to learn how to reason based on low-fidelity data.

Linear GP multi-fidelity

$$f_H(x) = \rho_H f_L(x) + \delta_H(x)$$

f_H High fidelity function (GP)

ρ_H Contribution of low fidelity (const)

f_L Low fidelity function (GP)

δ_H Bias between fidelities (GP)

Kennedy & O'Hagan 2000, Le Gratiet & Garnier 2014

Non-linear multi-fidelity GP => Deep GP

$$f_H(x) = \rho_H f_L(x) + \delta_H(x)$$

Linear relationship between fidelities

$$f_H(x) = \rho_H(f_L(x), x) + \delta_H(x)$$

Non-linear relationship between fidelities
(deep GP!)

Non-linear multi-fidelity GP => Deep GP

$$f_H(x) = \rho_H f_L(x) + \delta_H(x)$$

Linear relationship between fidelities

$$f_H(x) = \rho_H(f_L(x), x) + \delta_H(x)$$

Non-linear relationship between fidelities
(deep GP!)

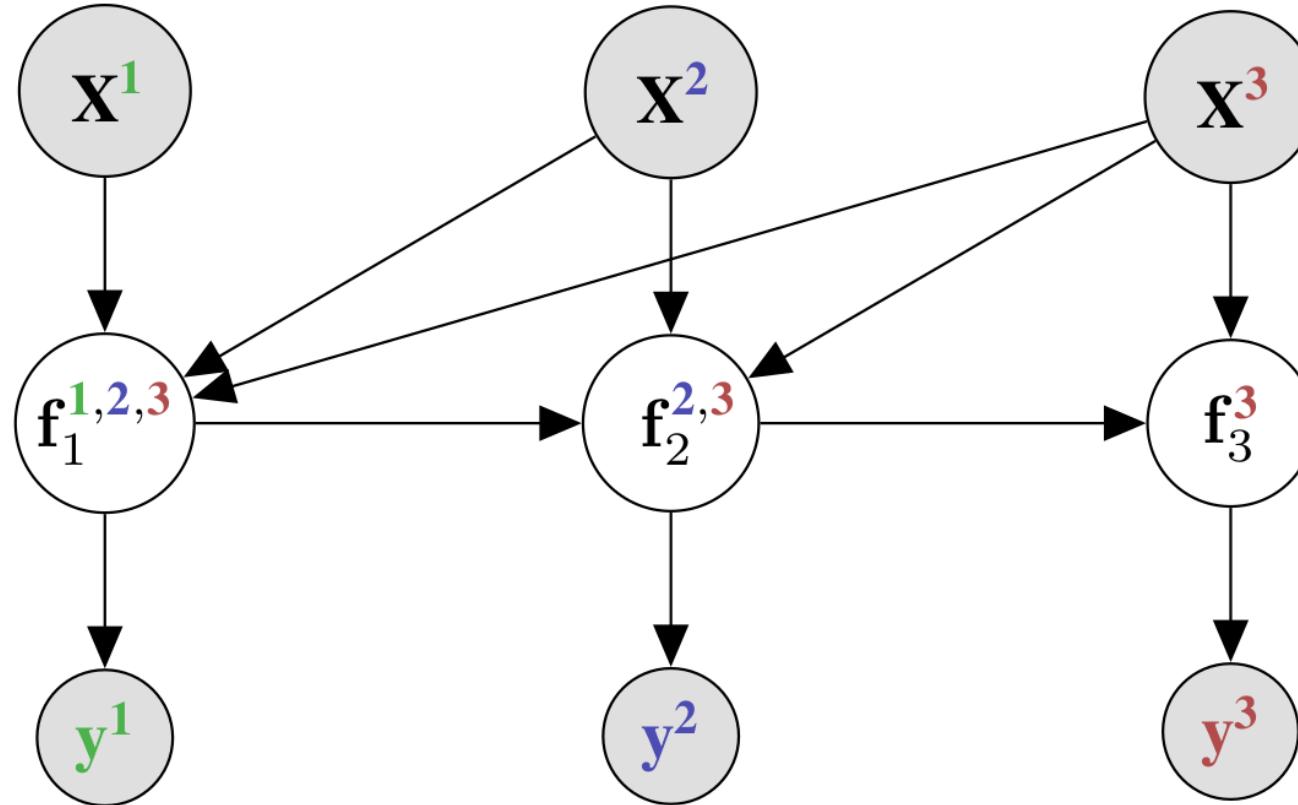


$$f_H(x) = g_H(f_L^*(x), x)$$

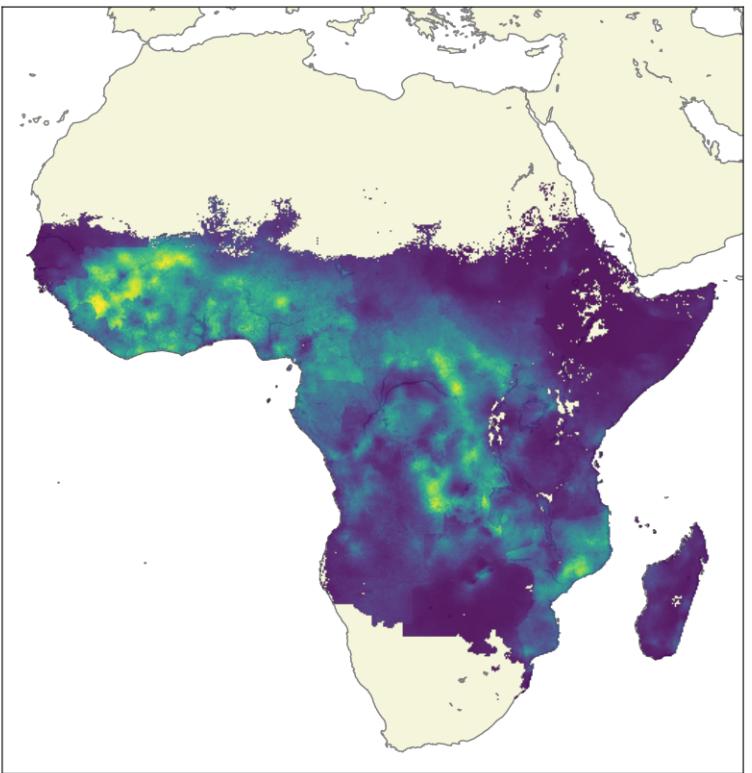
δ absorbed into g

$f_L^*(x)$ denotes a sample from the posterior of the GP modeling the low-fidelity data.

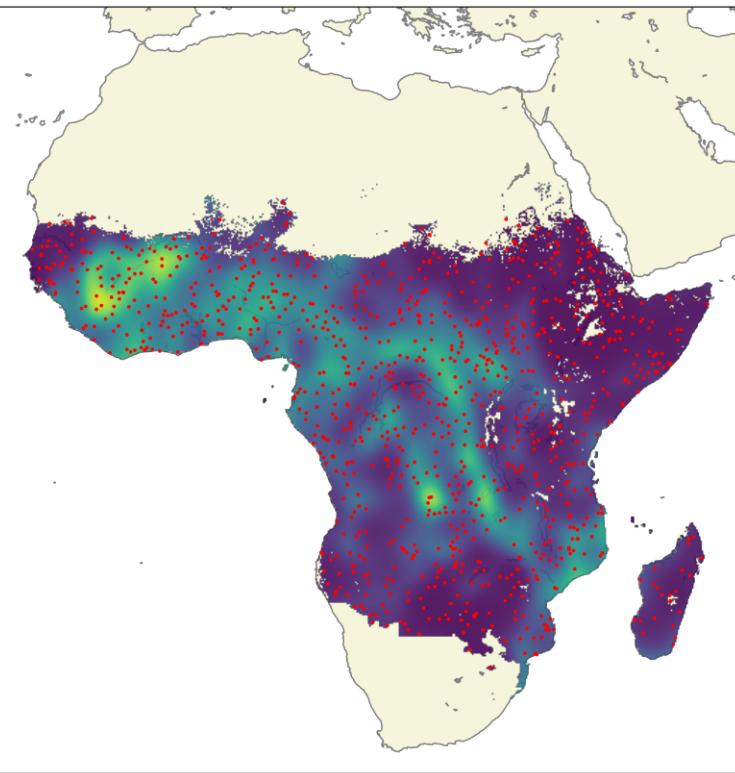
Graphical model



TRUE HIGH-FIDELITY



PREDICTED HIGH-FIDELITY



DPP SAMPLES

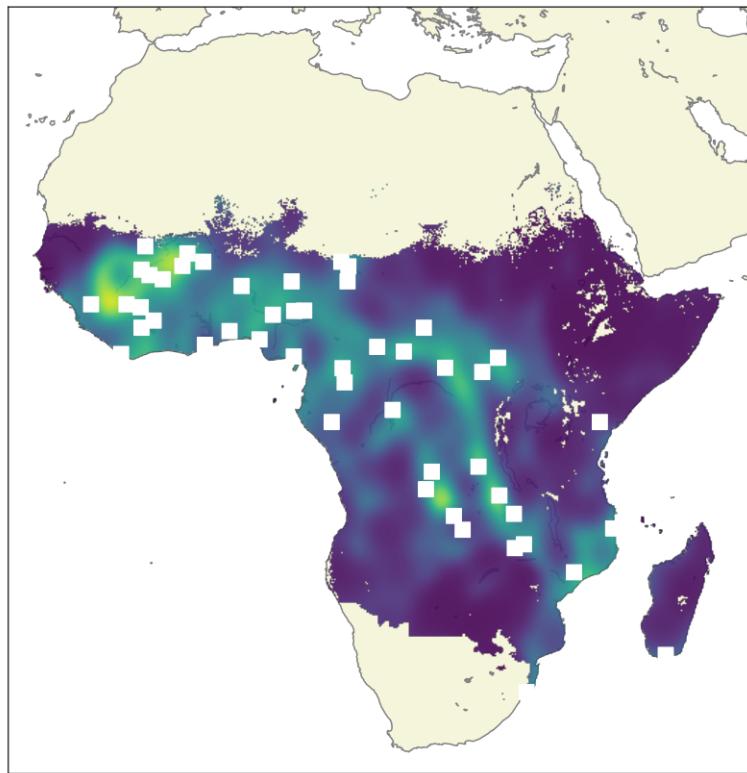


Figure 5: Real-world experiment indicating the infection rate of *Plasmodium falciparum* among African children. Lighter-shaded regions denote higher infection rates in that area of the continent. *Left:* True infection rates recorded for the year 2015. *Center:* MF-DGP predictions given low-fidelity data from 2005 and limited high-fidelity training points (marked in red) from 2015. *Right:* White squares show the samples drawn from a DPP using the posterior covariance of the MF-DGP model as its kernel.

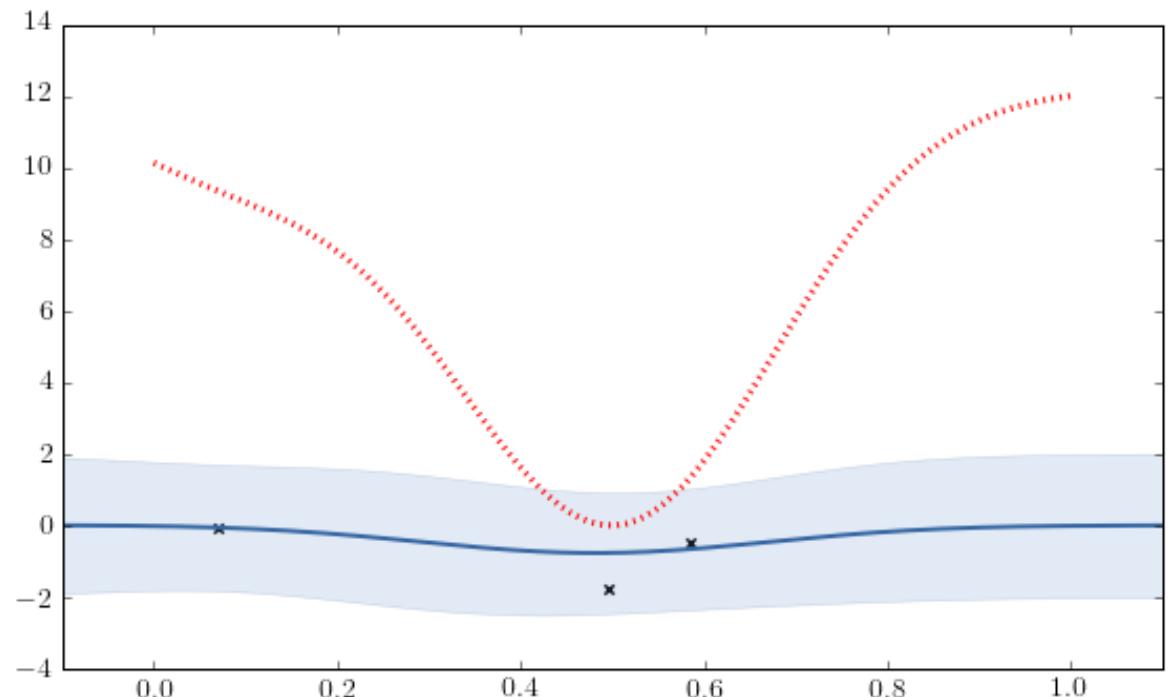
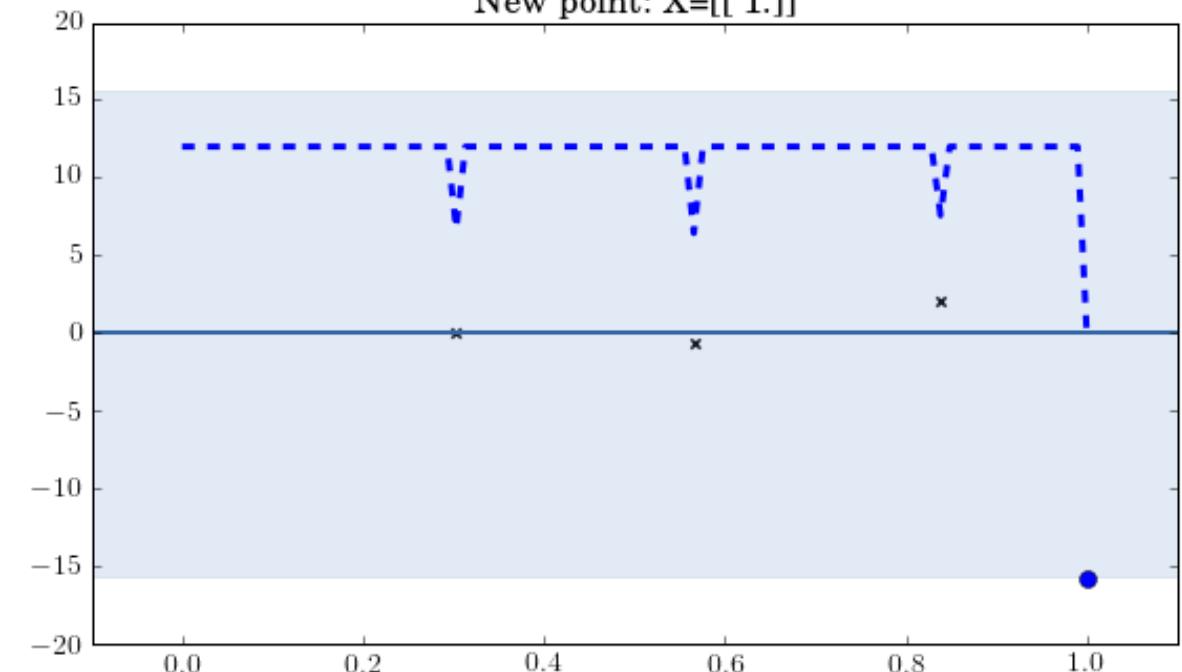
Sequential design for multi-fidelity GP

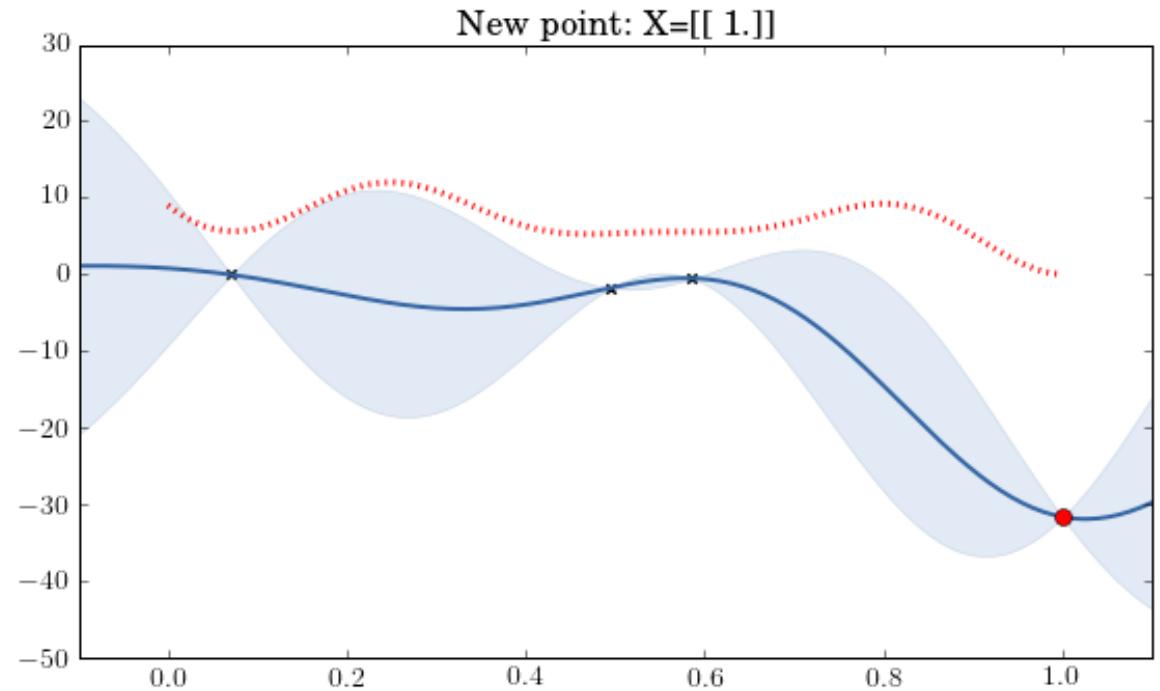
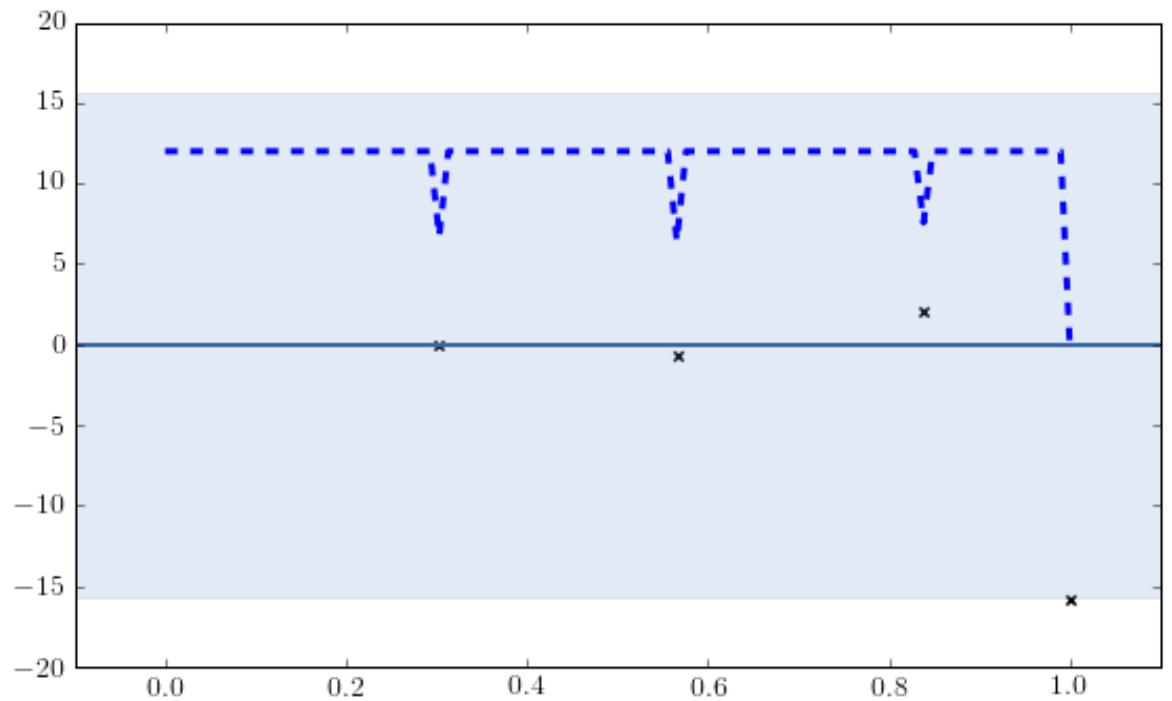
- ▶ **Strategy for collecting data points** across fidelities
- ▶ Each fidelity evaluation comes with a different **cost**, proportional to the level of fidelity
- ▶ Approach: Formulation as a multi-fidelity **bandit** GP problem in the UCB setting (*Kandasamy et al. 2016*)

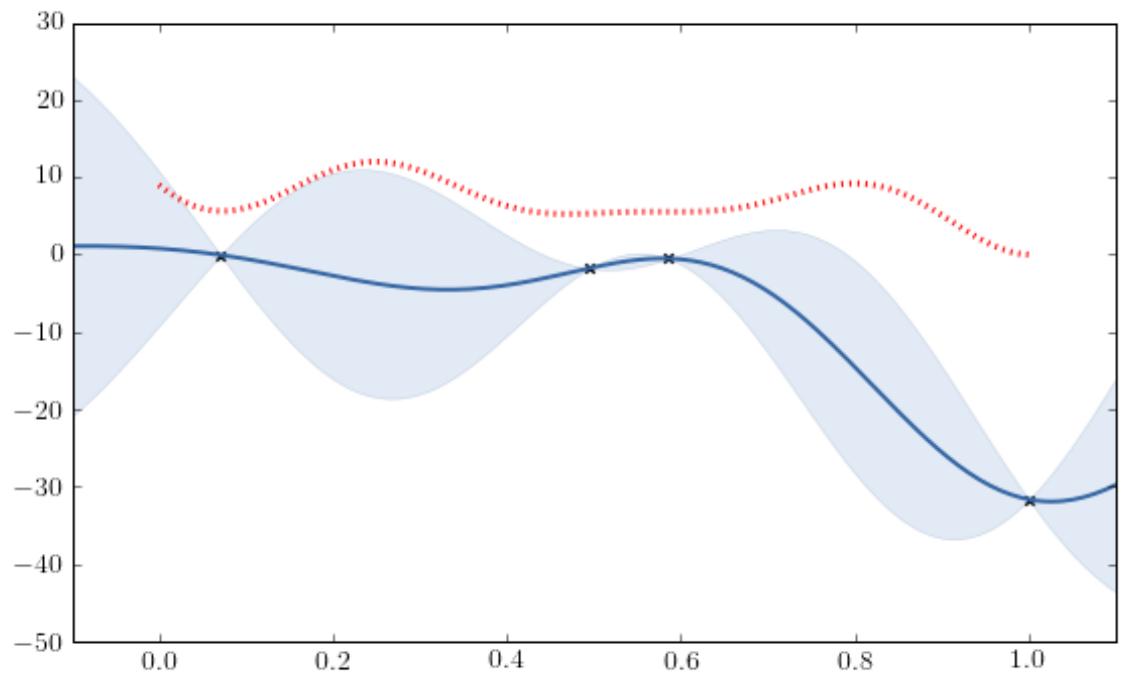
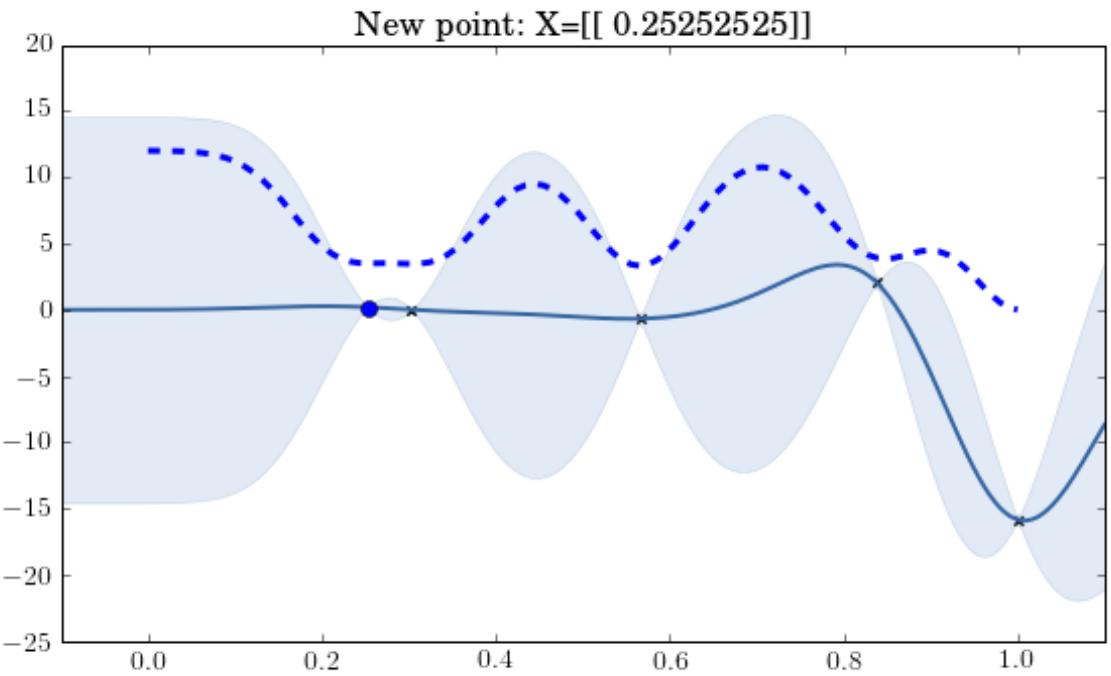
Demonstration

- ▶ Find the maximum of the high-fidelity function
- ▶ Consider cost while collecting points from each fidelity

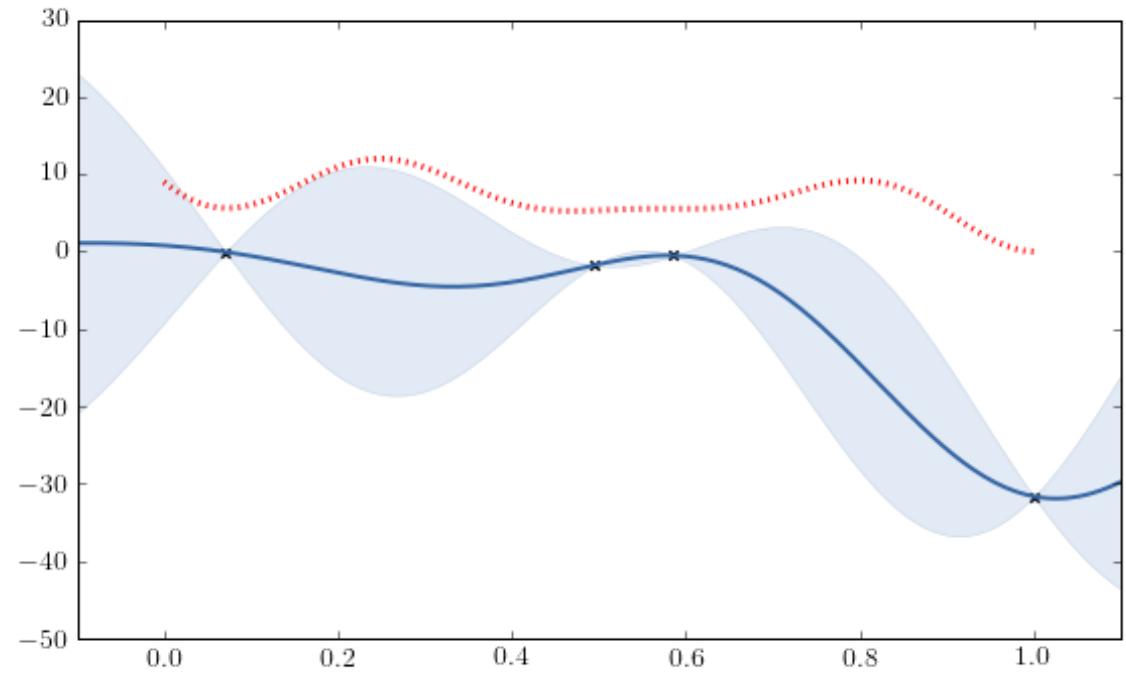
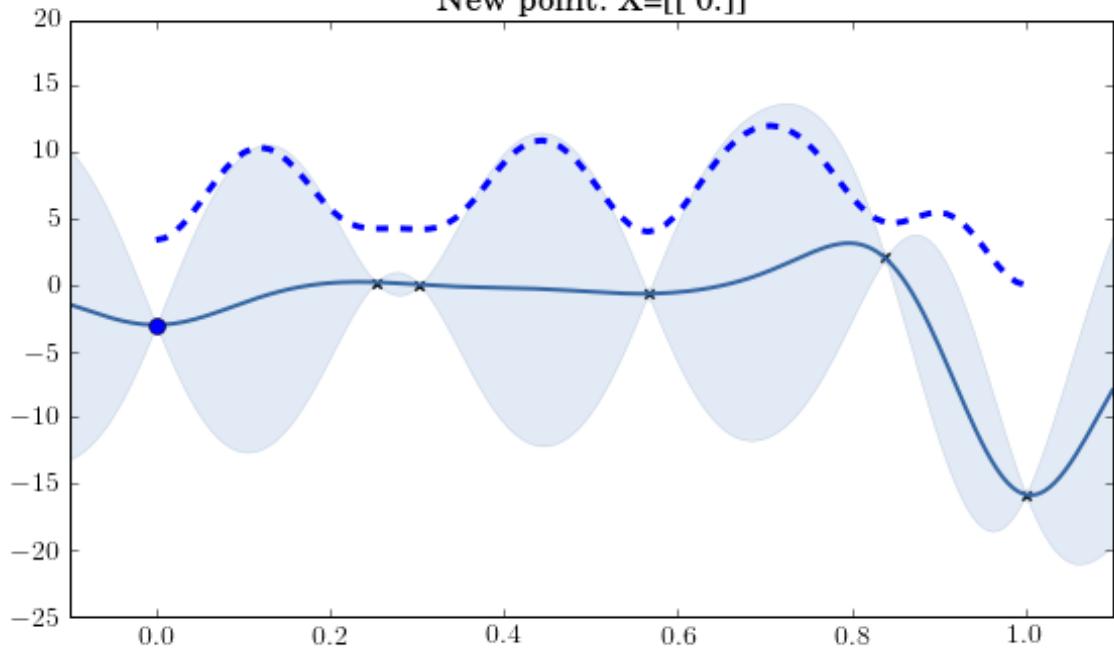
New point: X=[[1.]]

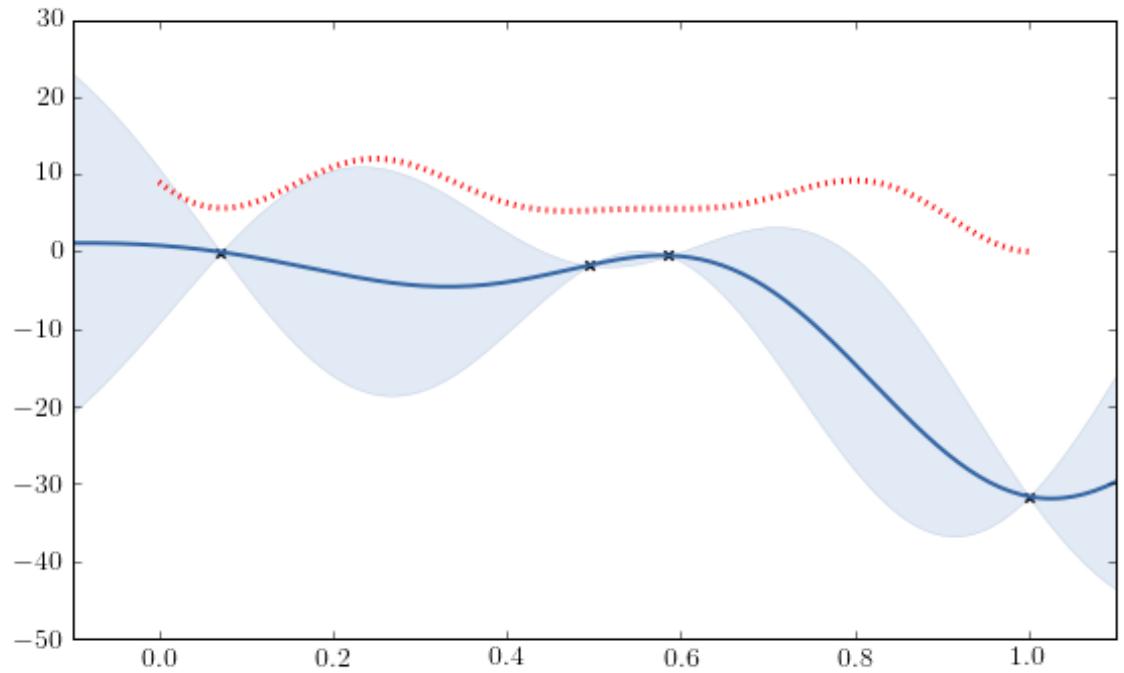
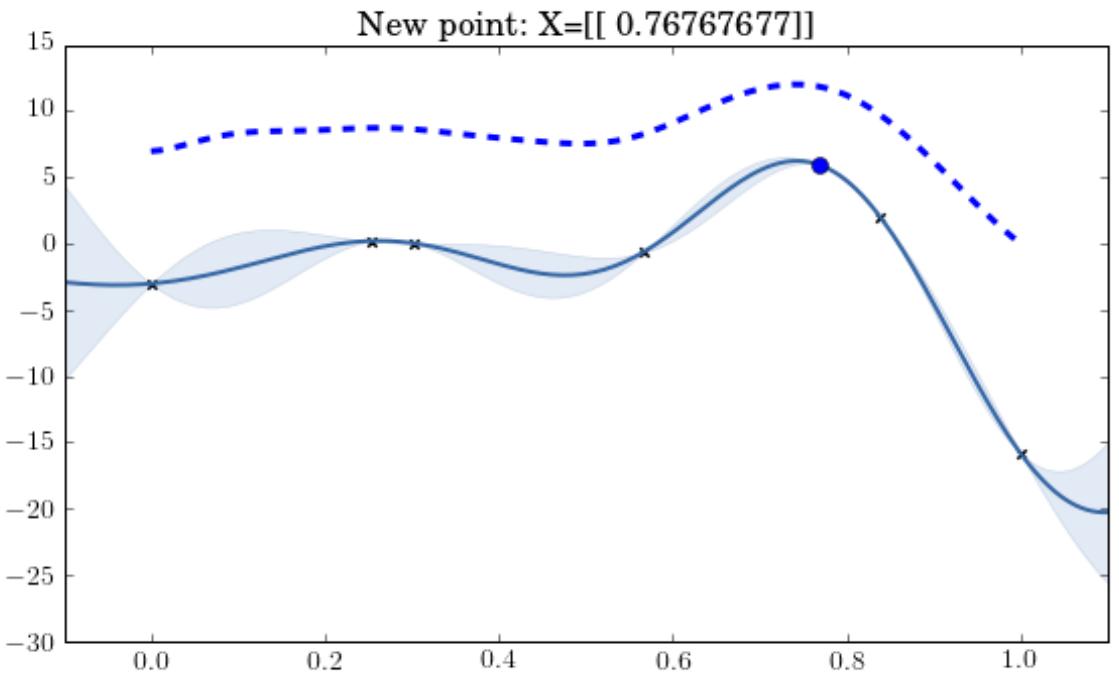


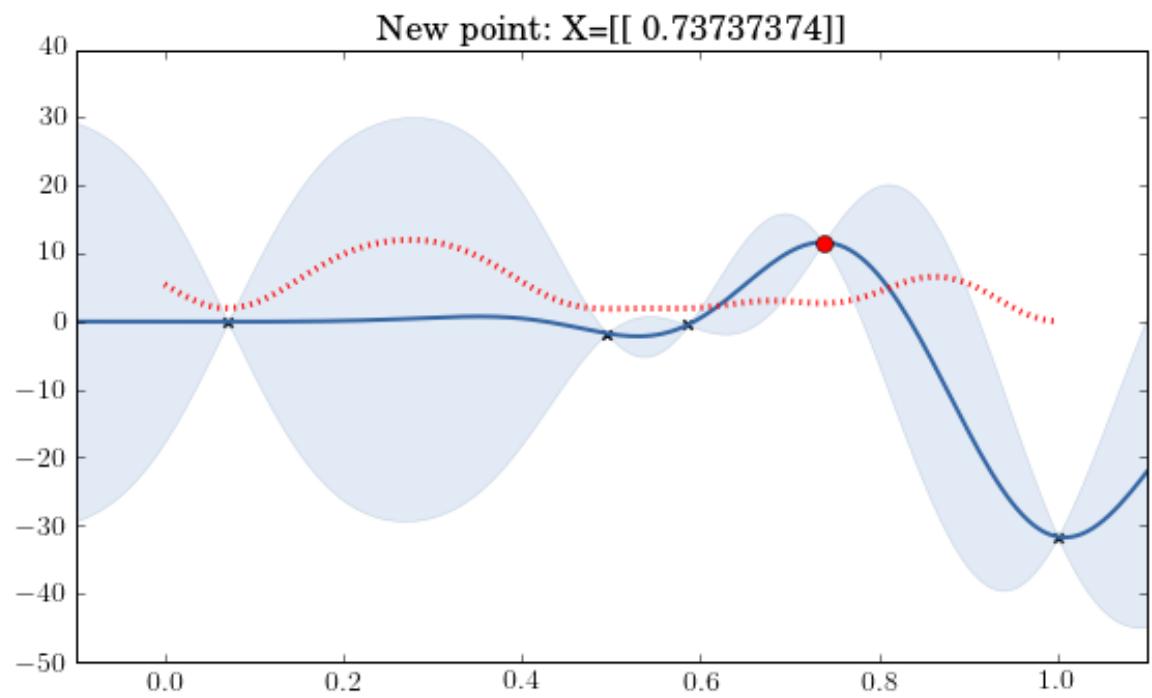
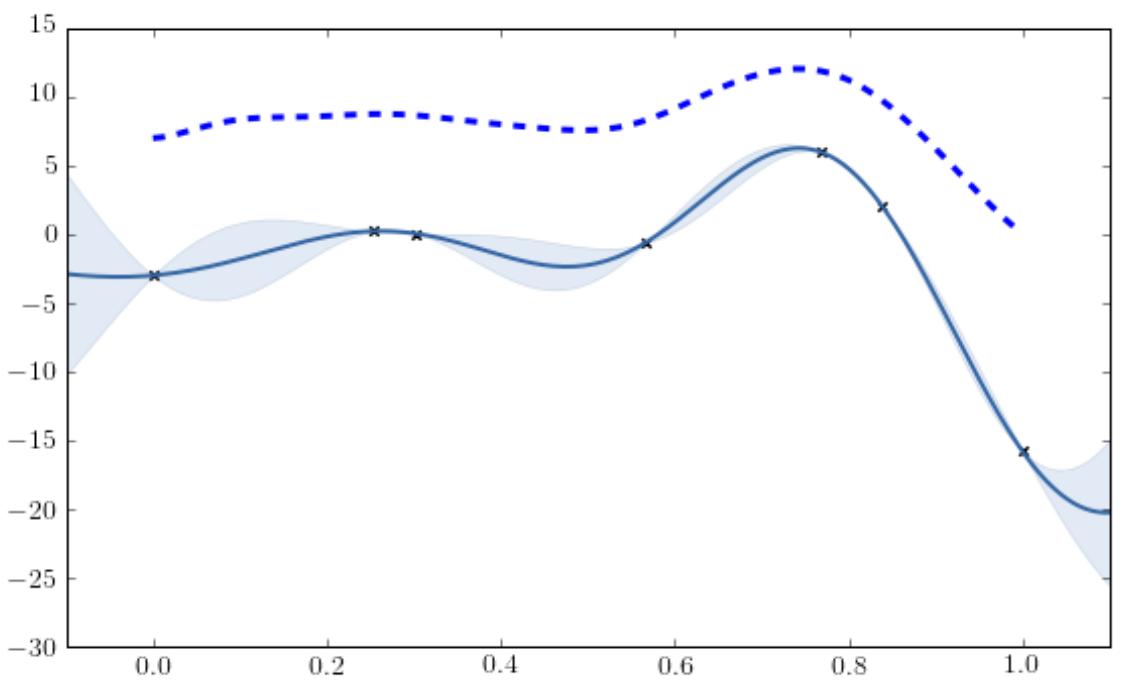


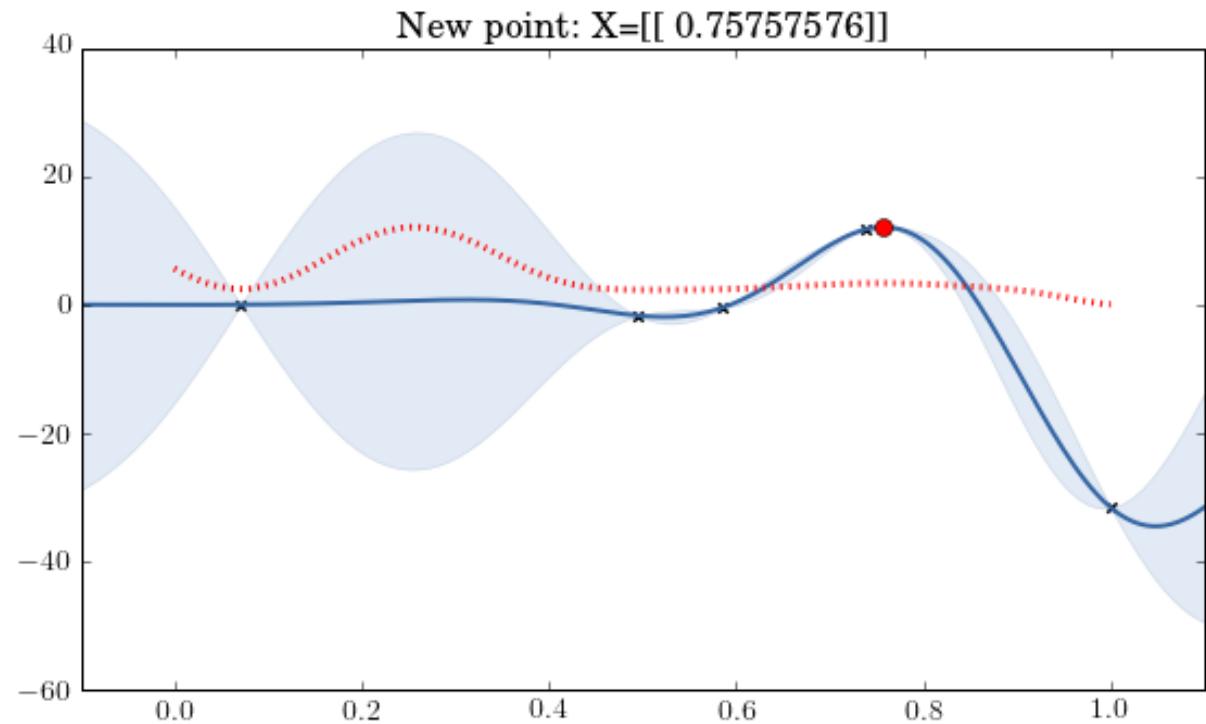
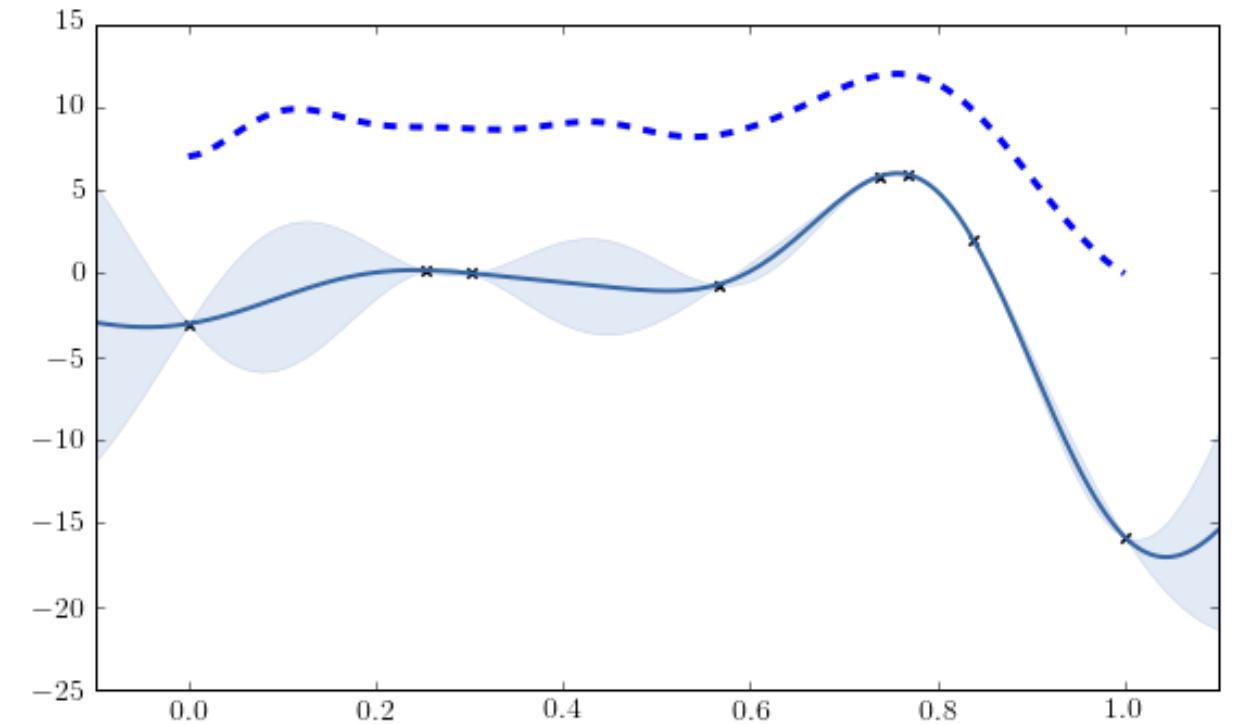


New point: $X=[[0.]]$









Conclusion

- ▶ Deep GPs learn rich mappings in a regularized and data efficient manner
- ▶ Multi-fidelity GPs allow us to fuse multiple fidelity data
- ▶ We can actively acquire data of different fidelities using the GP emulators in each fidelity.

Thanks!

Questions?

Appendix

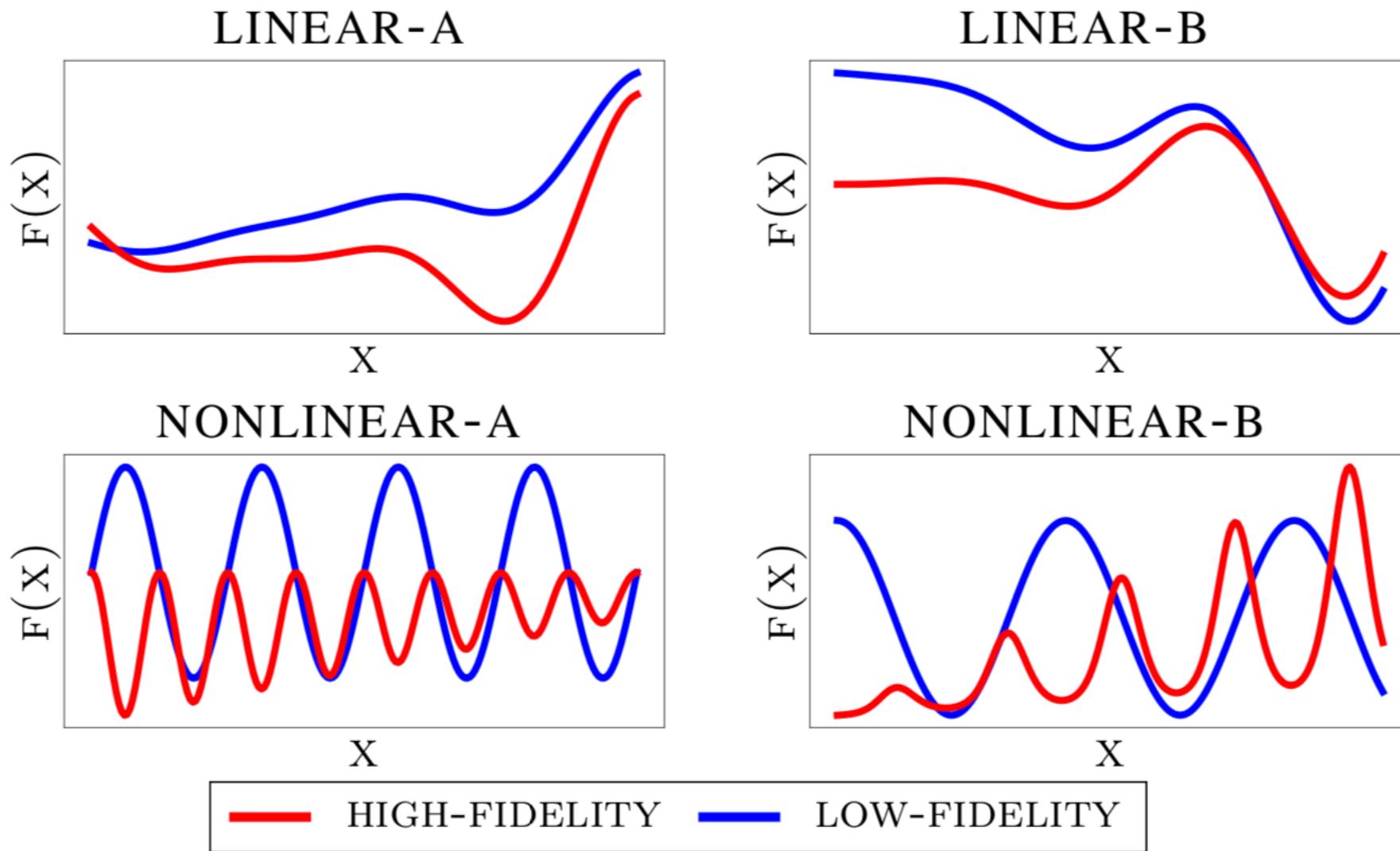
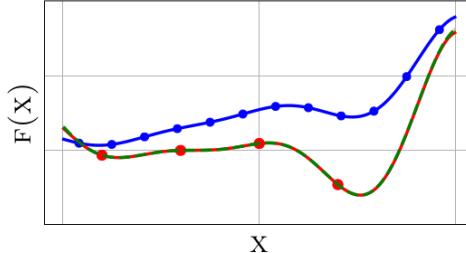


Figure 3: Synthetic examples. *Top:* Linear mapping between fidelities. *Bottom:* Nonlinear mapping.

AR1

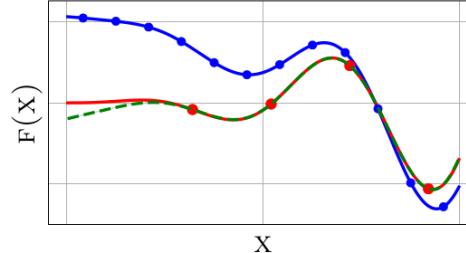
LINEAR-A

MNLL = -4.252



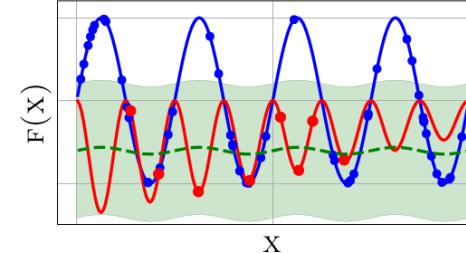
LINEAR-B

MNLL = 2722.070



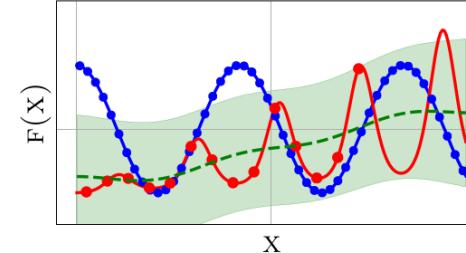
NONLINEAR-A

MNLL = 0.504



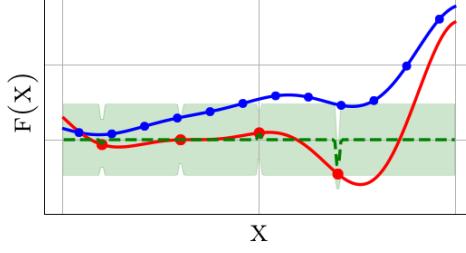
NONLINEAR-B

MNLL = -0.039

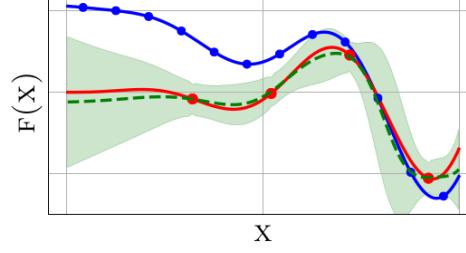


NARGP

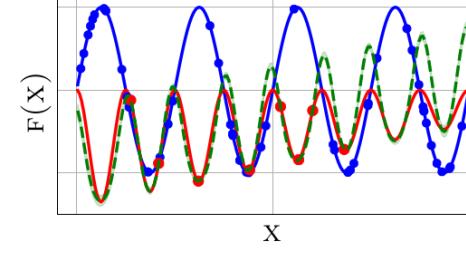
MNLL = 1.231



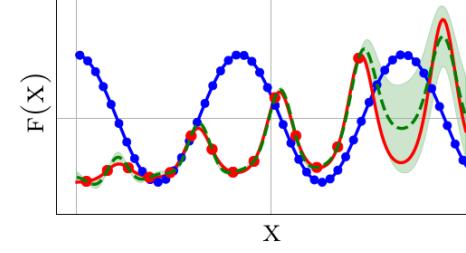
MNLL = -0.089



MNLL = 39.588

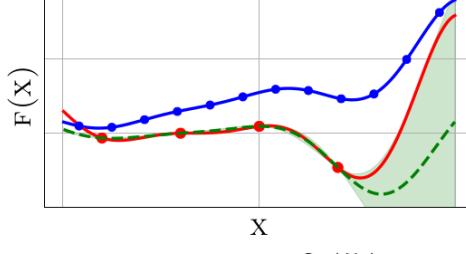


MNLL = -0.363



DEEP-MF

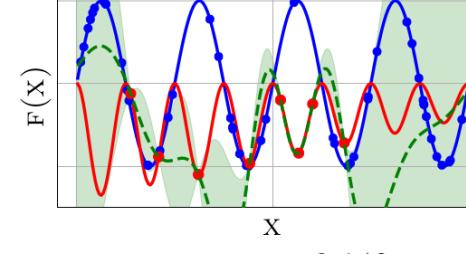
MNLL = 22.644



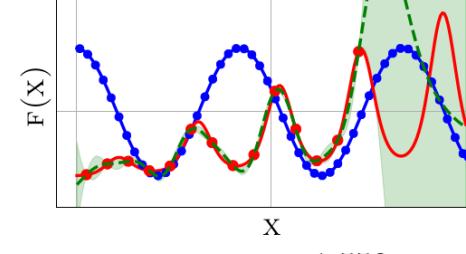
MNLL = 7.239



MNLL = 1.441

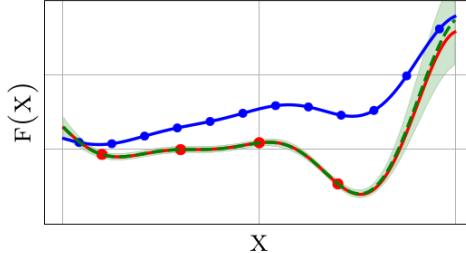


MNLL = 0.568

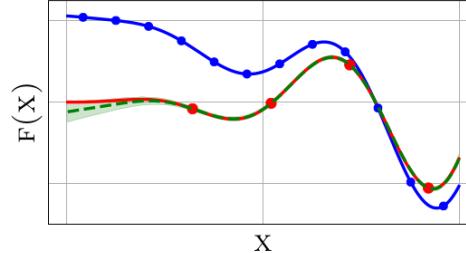


MF-DGP

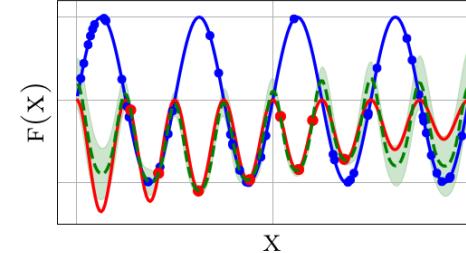
MNLL = -2.454



MNLL = -2.281



MNLL = -0.149



MNLL = -1.558

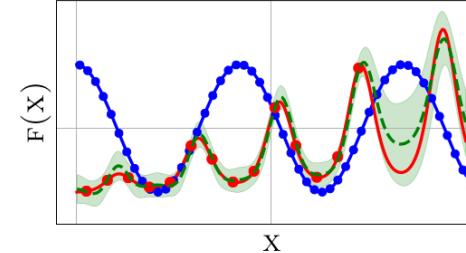


Figure 4: Cross-comparison across methods and synthetic examples for challenging multi-fidelity scenarios. MF-DGP yields conservative uncertainty estimates where few high-fidelity observations are available.

Benchmark examples

Table 1: Model Comparison on Multi-fidelity Benchmark Examples.

BENCHMARK	D_{in}	ALLOCATION	FIDELITY			AR1			NARGP			MF-DGP		
			R ²	RMSE	MNLL	R ²	RMSE	MNLL	R ²	RMSE	MNLL	R ²	RMSE	MNLL
CURRIN	2	12-5	0.913	0.677	20.105	0.903	0.740	20.817	0.935	0.601	0.763			
PARK	4	30-5	0.985	0.575	465.377	0.954	0.928	743.119	0.985	0.565	1.383			
BOREHOLE	8	60-5	1.000	0.005	-3.946	0.973	0.063	-1.054	0.999	0.015	-2.031			
BRANIN	2	80-30-10	0.891	0.044	-1.740	0.929	0.053	-1.223	0.965	0.030	-2.572			
HARTMANN-3D	3	80-40-20	0.998	0.043	0.440	0.305	0.755	0.637	0.994	0.075	-0.731			