Discovering and forecasting extreme events via active learning in neural operators

Ethan Pickering , Stephen Guth, George Em Karniadakis, and Themistoklis P. Sapsis

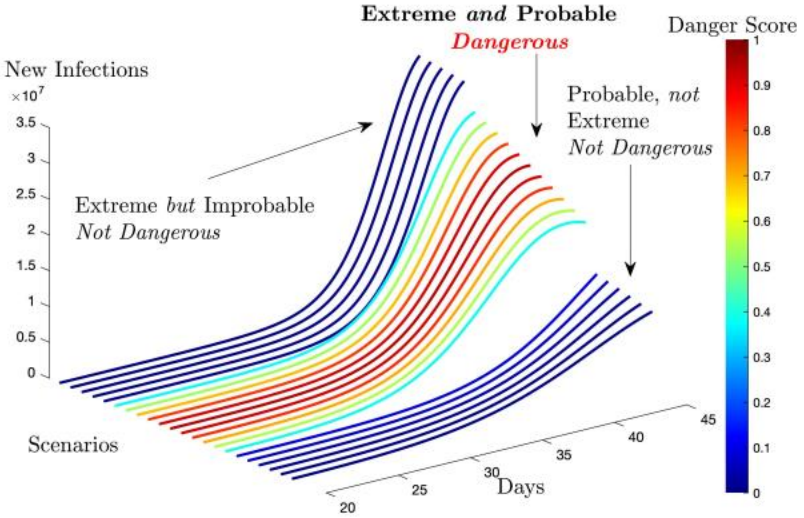
Extreme Phenomena

SIR

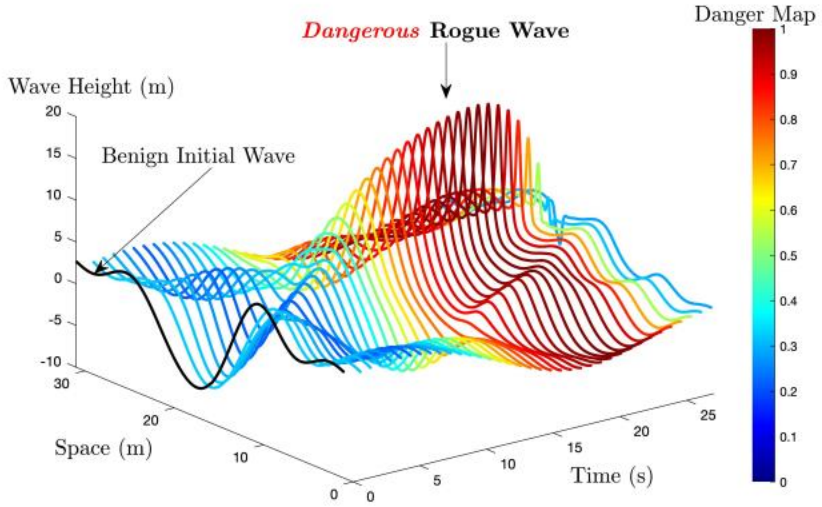
MMT

LAMP

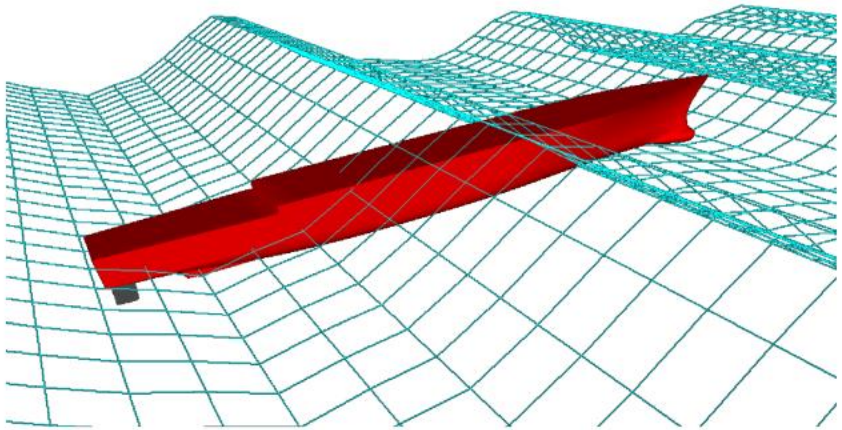
a) Pinpoint Dangerous Pandemic Spikes



b) Discover and Predict Rogue Ocean Waves



c) Estimate Peak Structural Ship Stress



$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta IS + \delta R$$
$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta IS - \gamma I$$
$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I - \delta R,$$

$$iu_t = |\partial_x|^{\alpha} u + \lambda |\partial_x|^{-\beta/4} \left(\left| |\partial_x|^{-\beta/4} u \right|^2 |\partial_x|^{-\beta/4} u \right) + iDu,$$

Deep Neural Operator

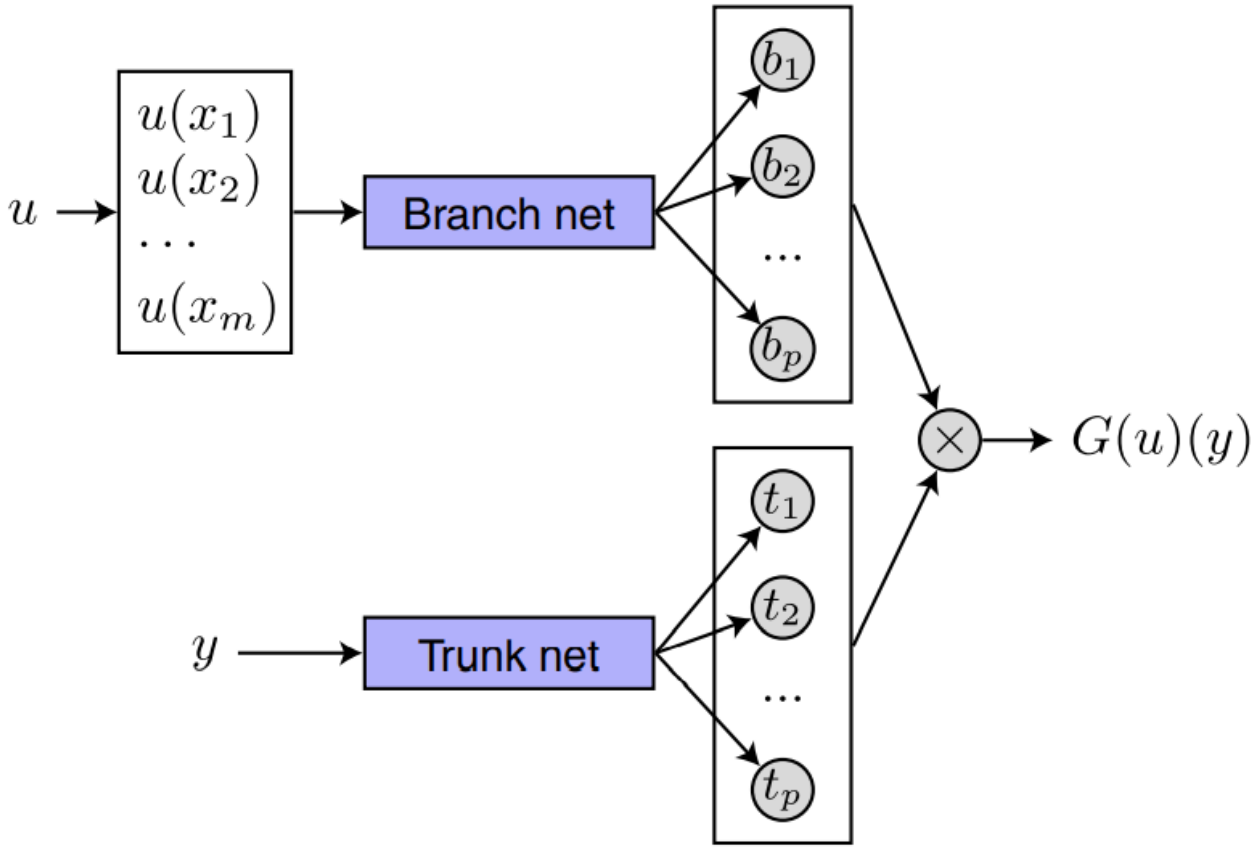
FNN: $x \rightarrow U(x)$

Neural Operator: $u(x) \rightarrow U(x)$

DeepONet

Universal Approximation Theorem for Operator

$$| G(u)(z) - \langle \underbrace{\mathbf{g}\left(u\left(x_1\right), u\left(x_2\right), \cdots, u\left(x_m\right)\right)}_{\text{branch}}, \underbrace{\mathbf{f}(z)}_{\text{trunk}} \rangle | < \epsilon.$$



DeepONet

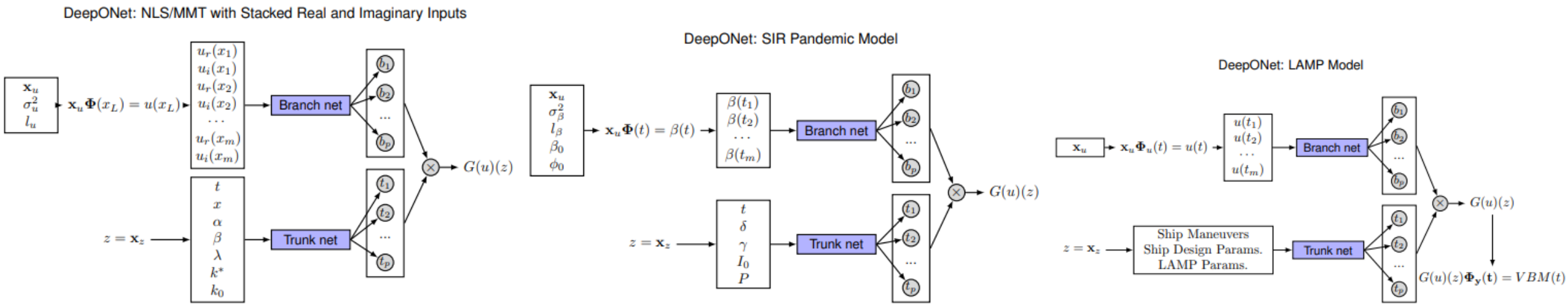
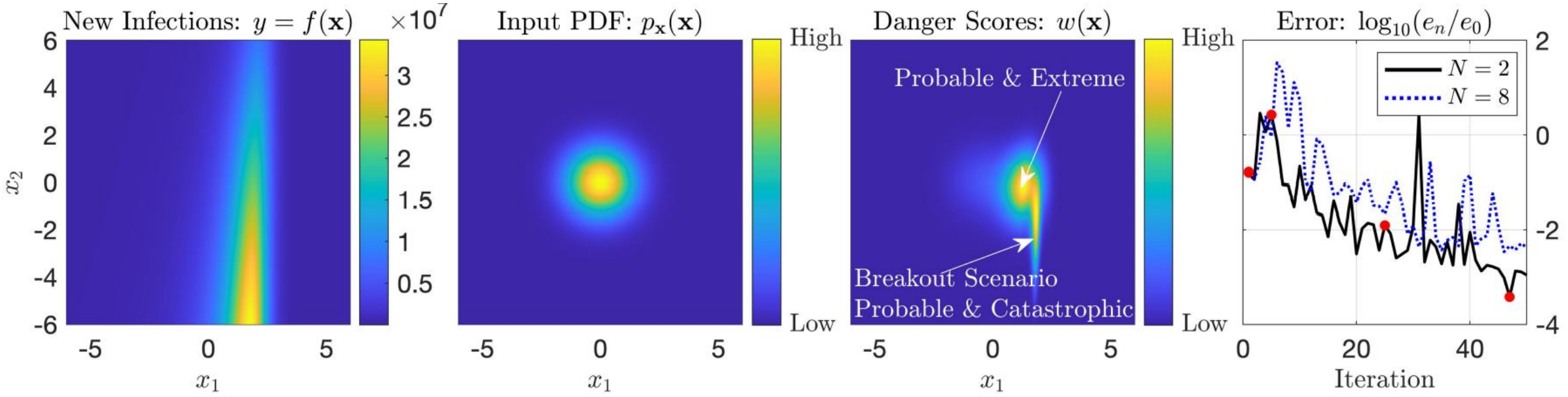


Figure 8: **The application of the MMT, SIR, and LAMP operators to DeepONet.**

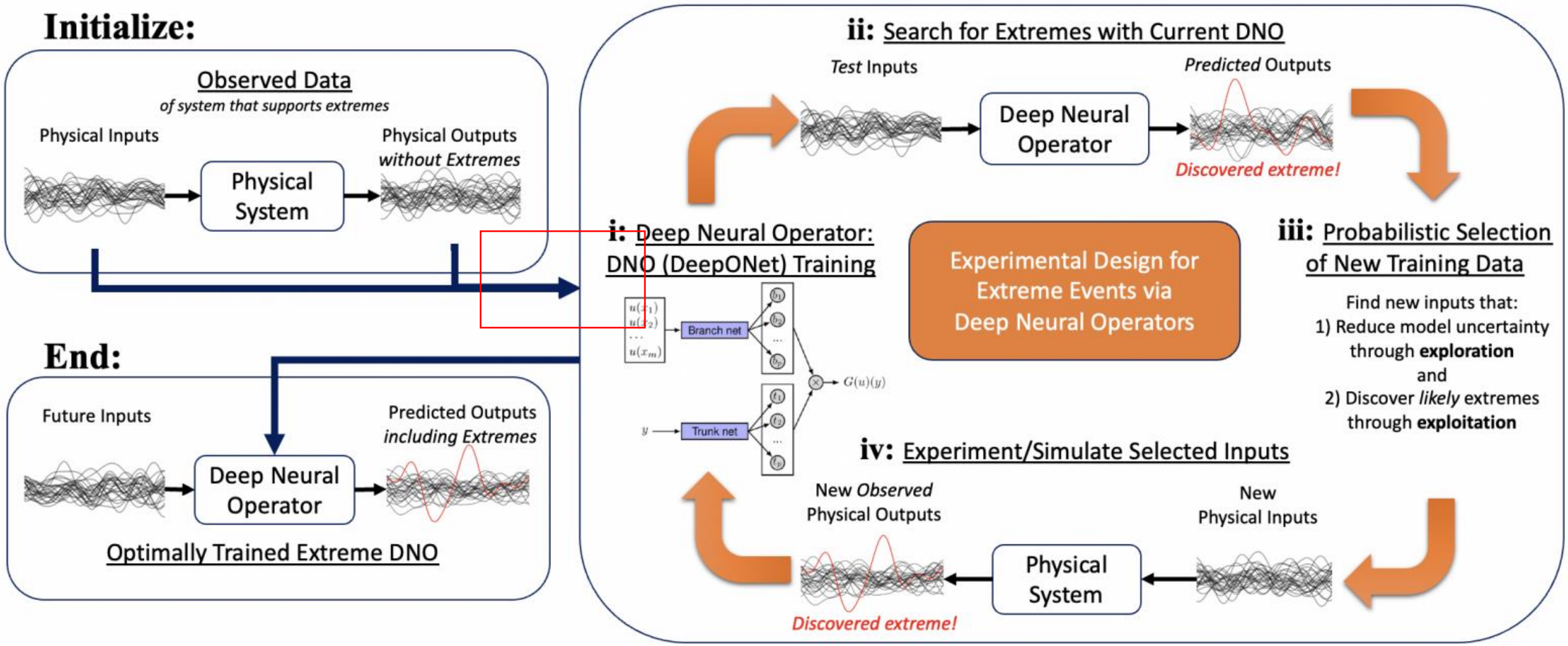
Method

Goal: accurately quantify the probability distribution function (PDF) of a stochastic quantity of interest (QoI), y

The variable y results from an observed random variable input, x, that is transformed by the underlying system or map, y = G(x).



DNO+BED Framework



DNO+BED Framework

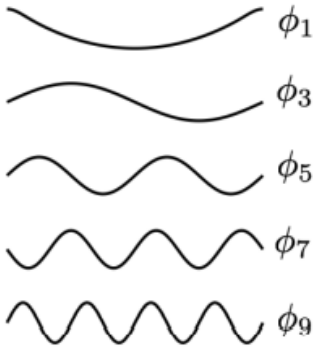
南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

I. Stochastic Process

u

II. Parameterize

$$u = \sum_{i=1}^{D} x_i \phi_i = \mathbf{x}\mathbf{\Phi}$$



III. Map

GP Approach

Parameter Mapping to QoI

$$\mathbf{x} \rightarrow \tilde{G}_{\mathrm{GP}}(\mathbf{x}) \approx y$$

DNO Approach

Functional Mapping to QoI

$$u = \mathbf{x}\mathbf{\Phi} \rightarrow \tilde{G}_{\mathrm{DNO}}(u = \mathbf{x}\mathbf{\Phi}) \approx y$$

1. Different Mapping

$$\tilde{G}_{\mathrm{GP}} \neq \tilde{G}_{\mathrm{DNO}}$$

2. Both Learn via $\mathbf{x}$

3. DNO Leverages Basis Information: $\mathbf{\Phi}$

Active Learning

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Algorithm 1 Sequential search for active learning/training of DNOs and GPs .

Input: Number of iterations: n_{iter}

Initialize: Train GP/DNO on initial dataset of input-output pairs

GP: $\mathcal{D}_0 = \{\mathbf{x}_i, y_i = G(\mathbf{x}_i)\}_{i=1}^{n_{\text{init}}}$ DNO: $\mathcal{D}_0 = \{\mathbf{x}_i, y_i = G(\mathbf{x}_{u,i}\mathbf{\Phi}(x_1, ...x_m), \mathbf{x}_{z,i})\}_{i=1}^{n_{\text{init}}}$

for $n = 1$ **to** n_{iter} **do**

Select next sample $\mathbf{x}_n$ by maximizing* acquisition function $a(\mathbf{x})$:

*Maximization using Monte Carlo, See Section 4.4 and Appendix G

$$\mathbf{x}_n = \arg\max_{\mathbf{x}\in\mathcal{X}} \ (a(\mathbf{x}; y), \mathcal{D}_{n-1})$$

Evaluate objective function at $\boldsymbol{x}_n$ and record y_n

Augment dataset: $\mathcal{D}_n = \mathcal{D}_{n-1} \cup \mathbf{x}_n, y_n$

Retrain GP/DNO.

end for

return Final GP/DNO Model

Ensemble of Neural Networks and Uncertainty Quantification

南京航空航天大學 NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

train N randomly weight-initialized DNO models, each denoted as G˜ n

$$\sigma^2(u,z) = \frac{1}{(N-1)} \sum_{n=1}^{N} (\tilde{G}_n(u)(z) - \overline{\tilde{G}(u)(z)})^2$$

Acquisiton functions

1. $$a_{US}(\mathbf{x}) = \sigma^2(\mathbf{x}).$$

2. both highly uncertain and high magnitude

likelihood ratio $$w(\mathbf{x}) = \frac{p_{\mathbf{x}}(\mathbf{x})}{p_{\mu}(\mu(\mathbf{x}))}.$$ $$a_{US-LW}(\mathbf{x}) = w(\mathbf{x})\sigma^2(\mathbf{x})$$

Experiment Batching

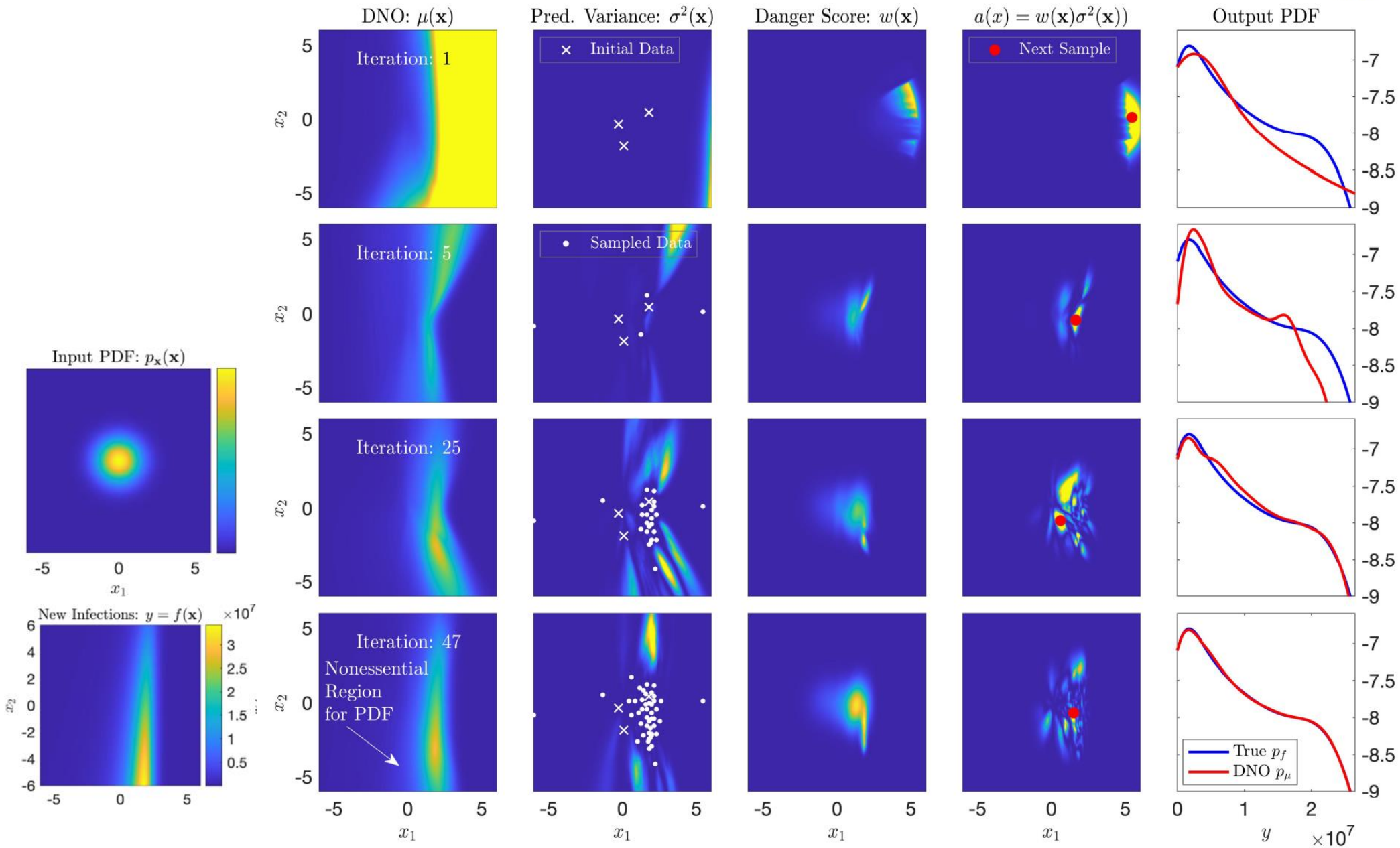
南京航空航天大學 NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

The purpose of batching is to find multiple regions of local optima of the acquisition function, rather than finding several optima in the same region.

Algorithm 2 Sequential selection of samples for batch sampling.

1: **Evaluate** acquisition function for n_q query points.
2: **for** Acquisition samples smaller than batch size $n_a < n_b$
3: **Choose** the maximum score, $\max(a) = a(\mathbf{x}_c)$, from n_q points.
4: **Augment** $\mathbf{x}_a$ with chosen point, $\mathbf{x}_c$.
5: **Compute** the distances, r, between the chosen point and the remaining query points.
6: **Eliminate** all samples from n_q where $r < r_{\min}$.
7: **end for**
8: **return** samples for the next experiment: $\mathbf{x}_a$

Danger Score Convergence in 50 Samples



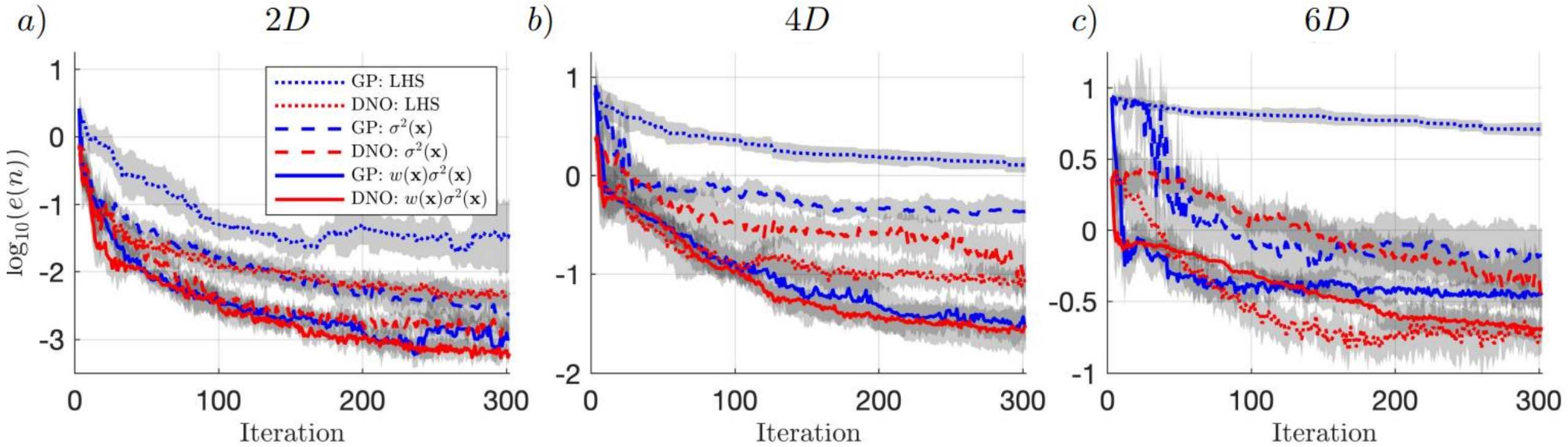
Convergence

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Accelerated convergence with DNOs plus extreme acquisition functions regardless of dimensionality

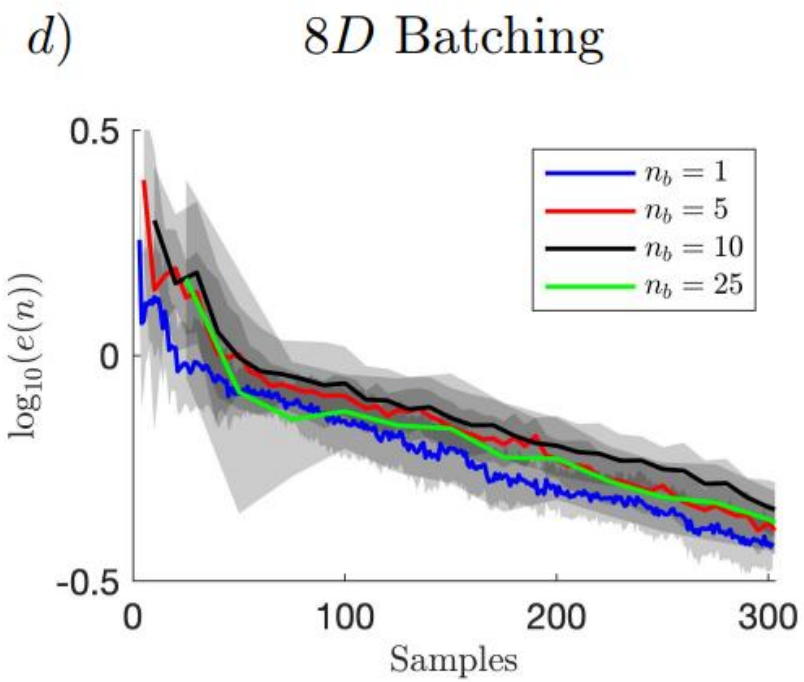
$$e(n) = \int |\log_{10} p_{\mu_n}(y) - \log_{10} p_G(y)| \mathrm{d}y.$$

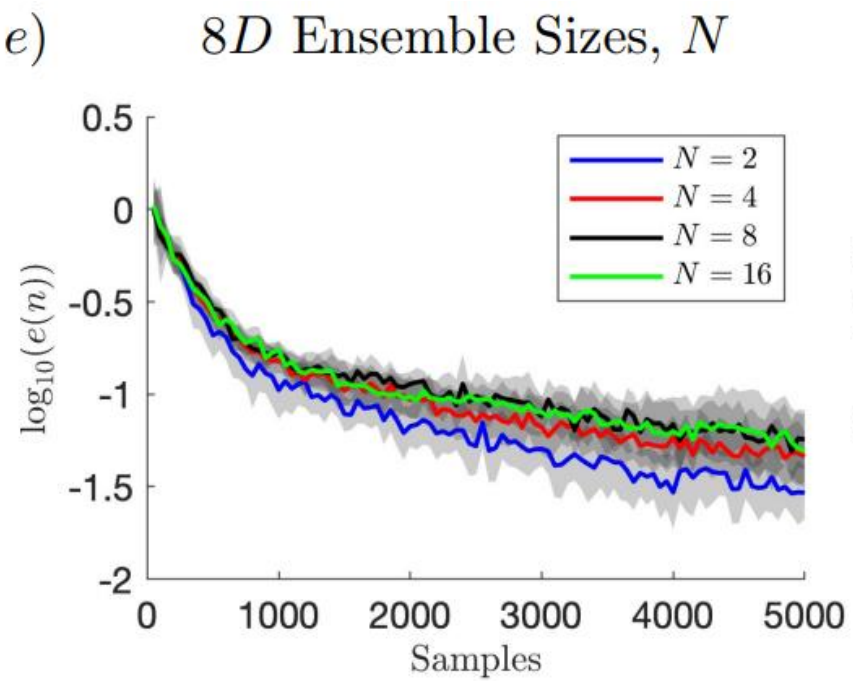
DNO ✓ w(x)σ2(x) ✓

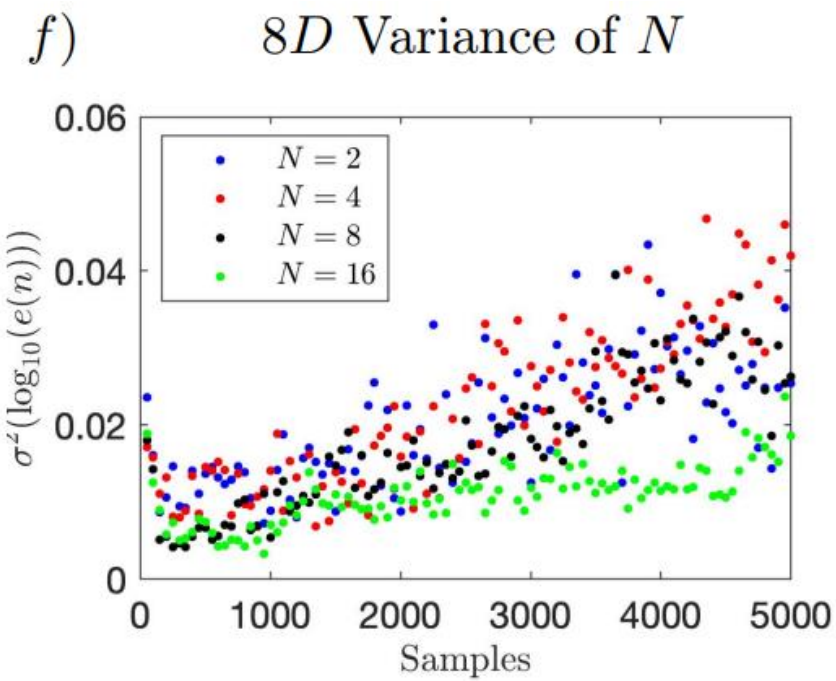


Convergence

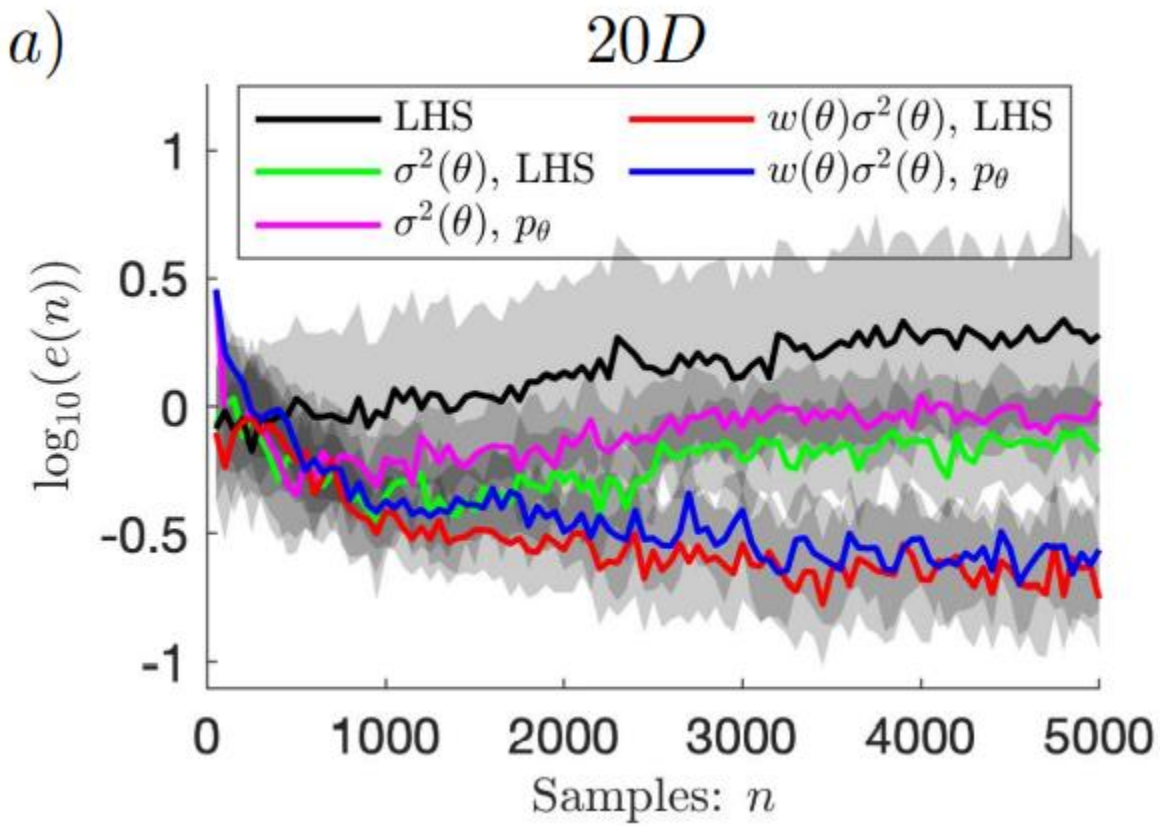
1. reducing computational cost in high dimensions
2. N=2 seems performs better, appears to disagree with the natural hypothesis (maybe use small N imposes a greedy search)
3. larger ensembles lead to a higher fidelity predictive variance

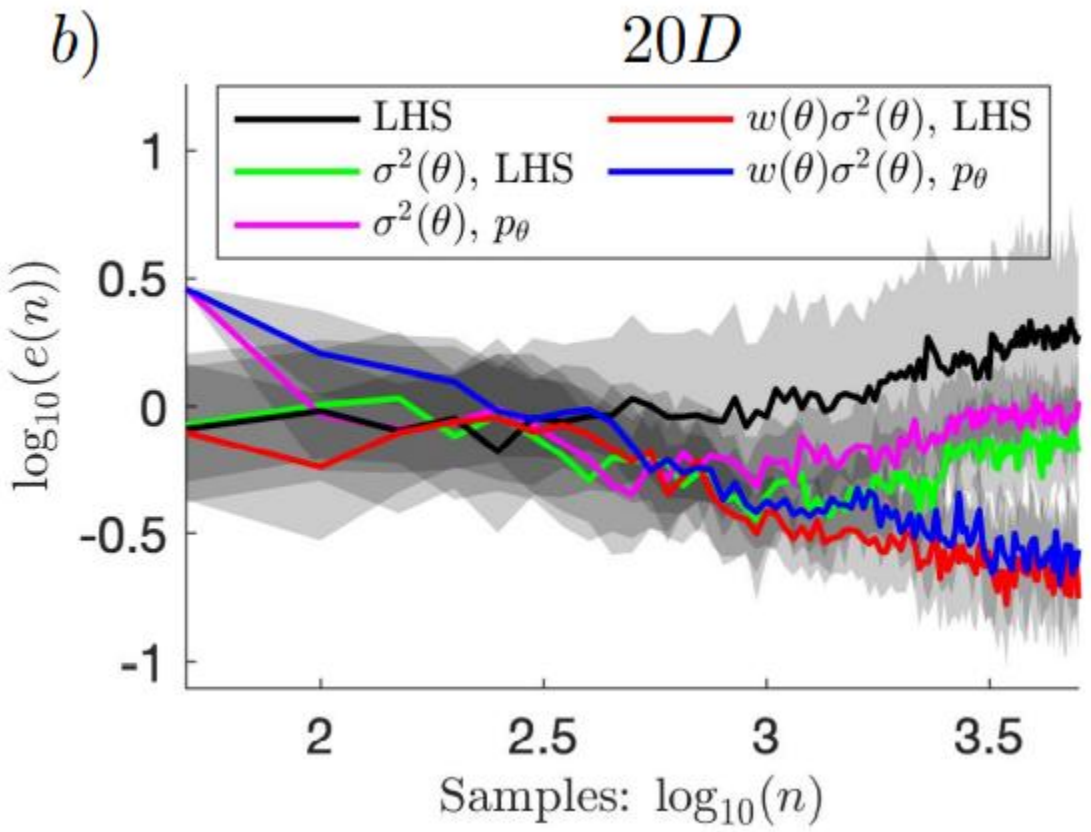






Robustness


a)
20D
LHS
$\sigma^2(\theta)$, LHS
$\sigma^2(\theta)$, p_θ
$w(\theta)\sigma^2(\theta)$, LHS
$w(\theta)\sigma^2(\theta)$, p_θ
$\log_{10}(e(n))$
1
0.5
0
-0.5
-1
0
1000
2000
3000
4000
5000
Samples: n


b)
20D
LHS
$\sigma^2(\theta)$, LHS
$\sigma^2(\theta)$, p_θ
$w(\theta)\sigma^2(\theta)$, LHS
$w(\theta)\sigma^2(\theta)$, p_θ
$\log_{10}(e(n))$
1
0.5
0
-0.5
-1
2
2.5
3
3.5
Samples: $\log_{10}(n)$

Thanks