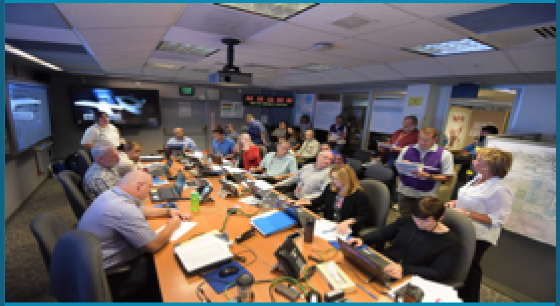
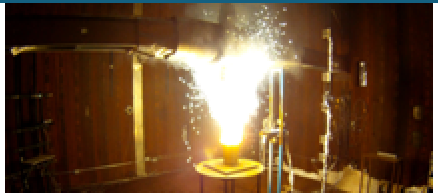
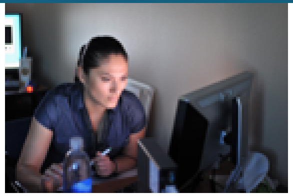
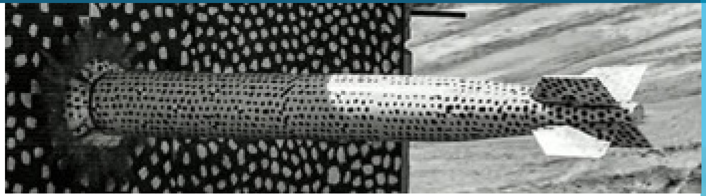
SAND2022-10972C



Error-in-variables modelling for operator learning







Mathematical and Scientific Machine Learning 2022

Ravi G. Patel[1], Indu Manickam[1], MK Lee[2], Mamikon Gulian[1],

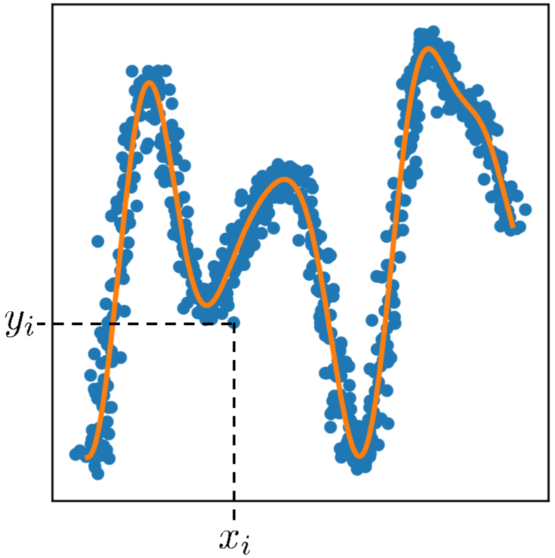
[1]Sandia National Laboratories
[2]University of Alabama

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No: SAND2022-5307 C

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

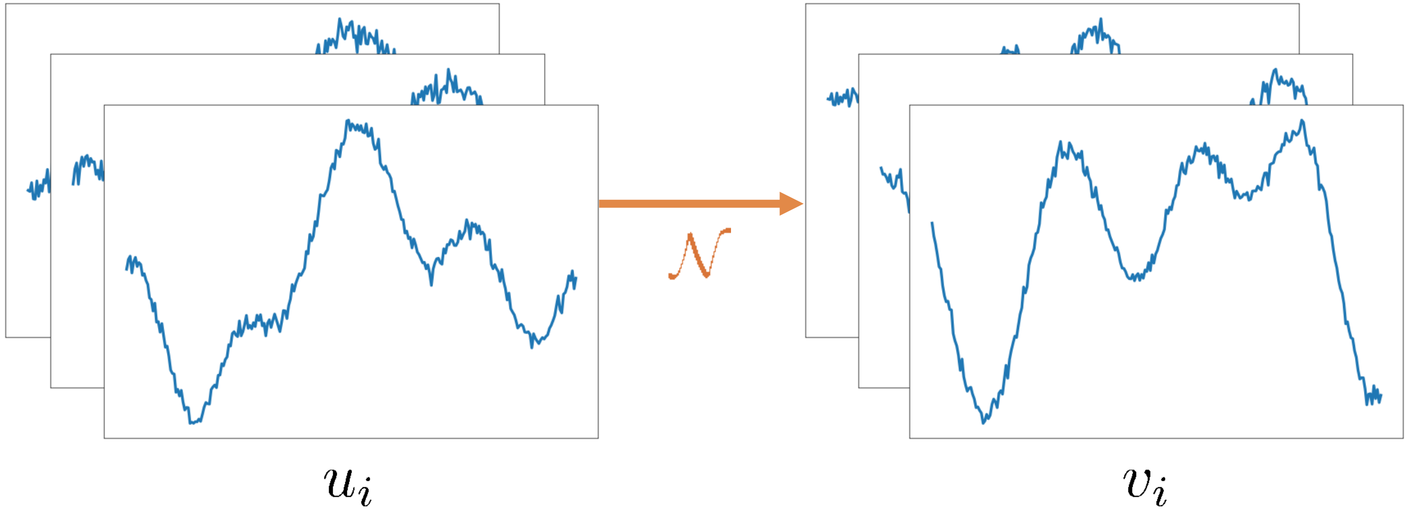
Operator learning – an ingredient for surrogate models

Fitting functions



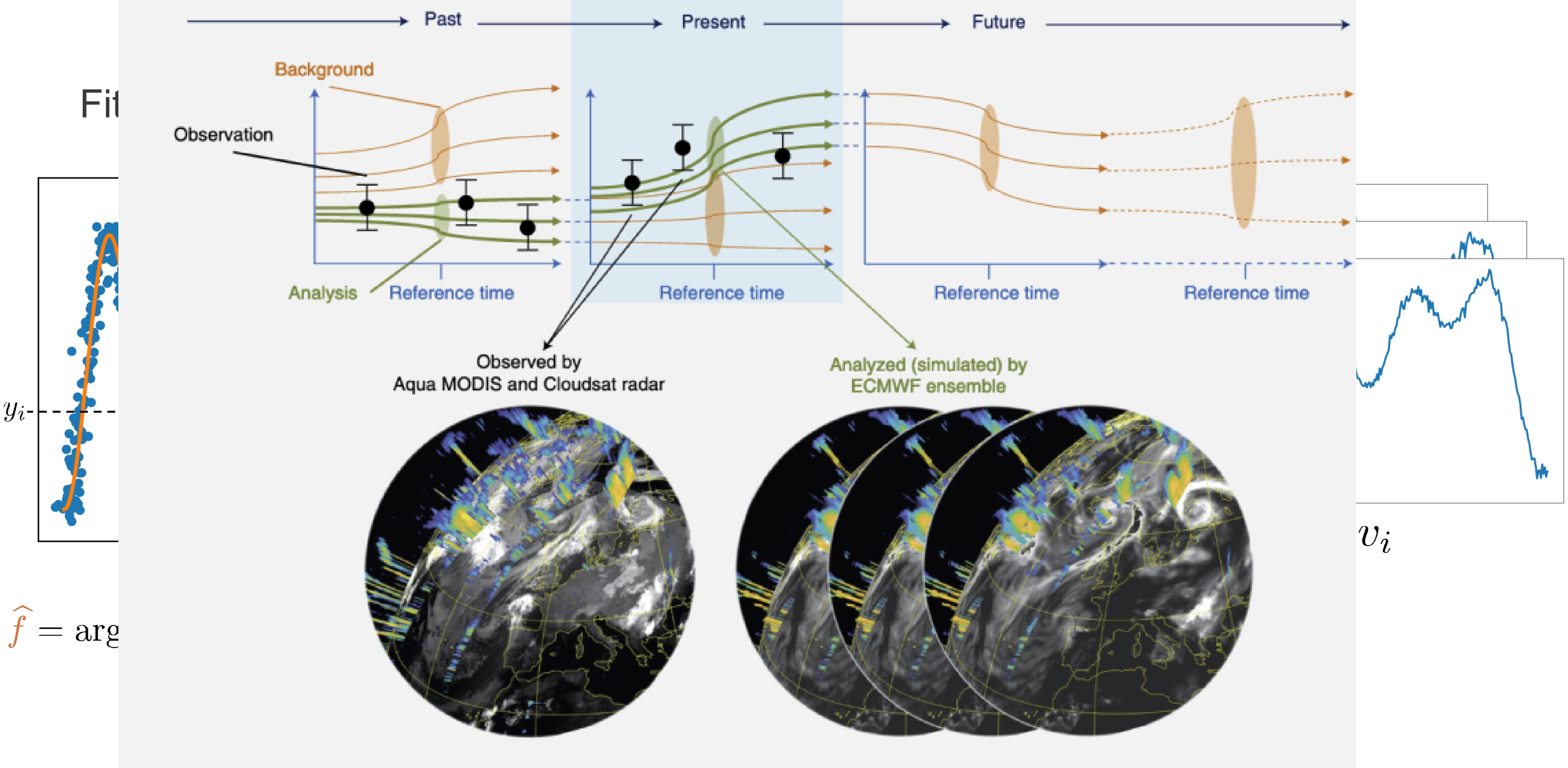
$$\widehat{f} = \underset{f}{\text{argmin}} \sum_i ||y_i - f(x_i)||$$

Fitting operators



$$\widehat{\mathcal{N}} = \underset{\mathcal{N}}{\text{argmin}} \sum_i ||v_i - \mathcal{N}[u_i]||$$

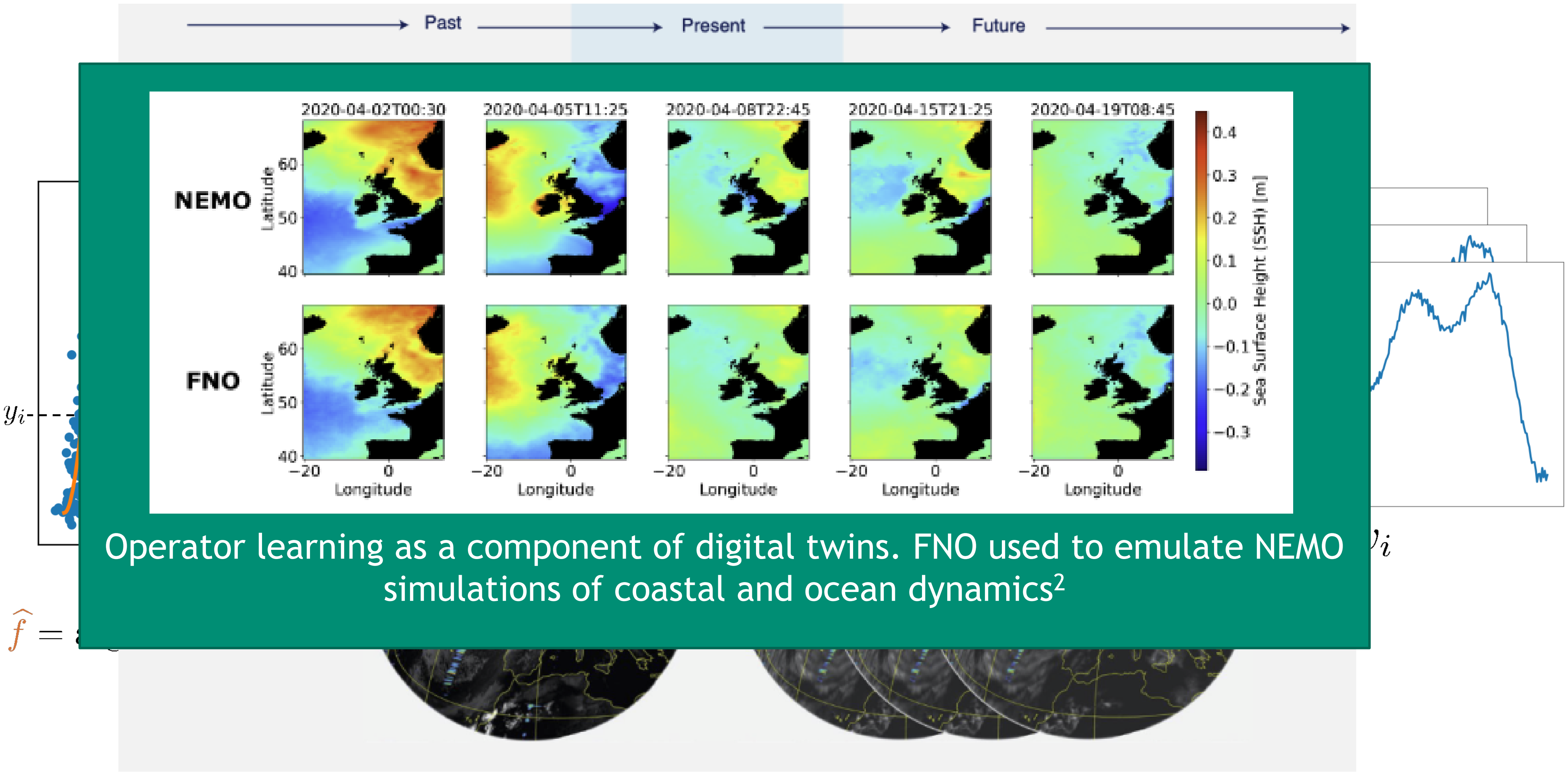
Operator learning – an ingredient for surrogate models



An Earth digital twin - combining MODIS and Cloudsat observations with ECMWF simulations[1]

[1]Bauer et al., *Nature Computational Science*, 2021

Operator learning – an ingredient for surrogate models



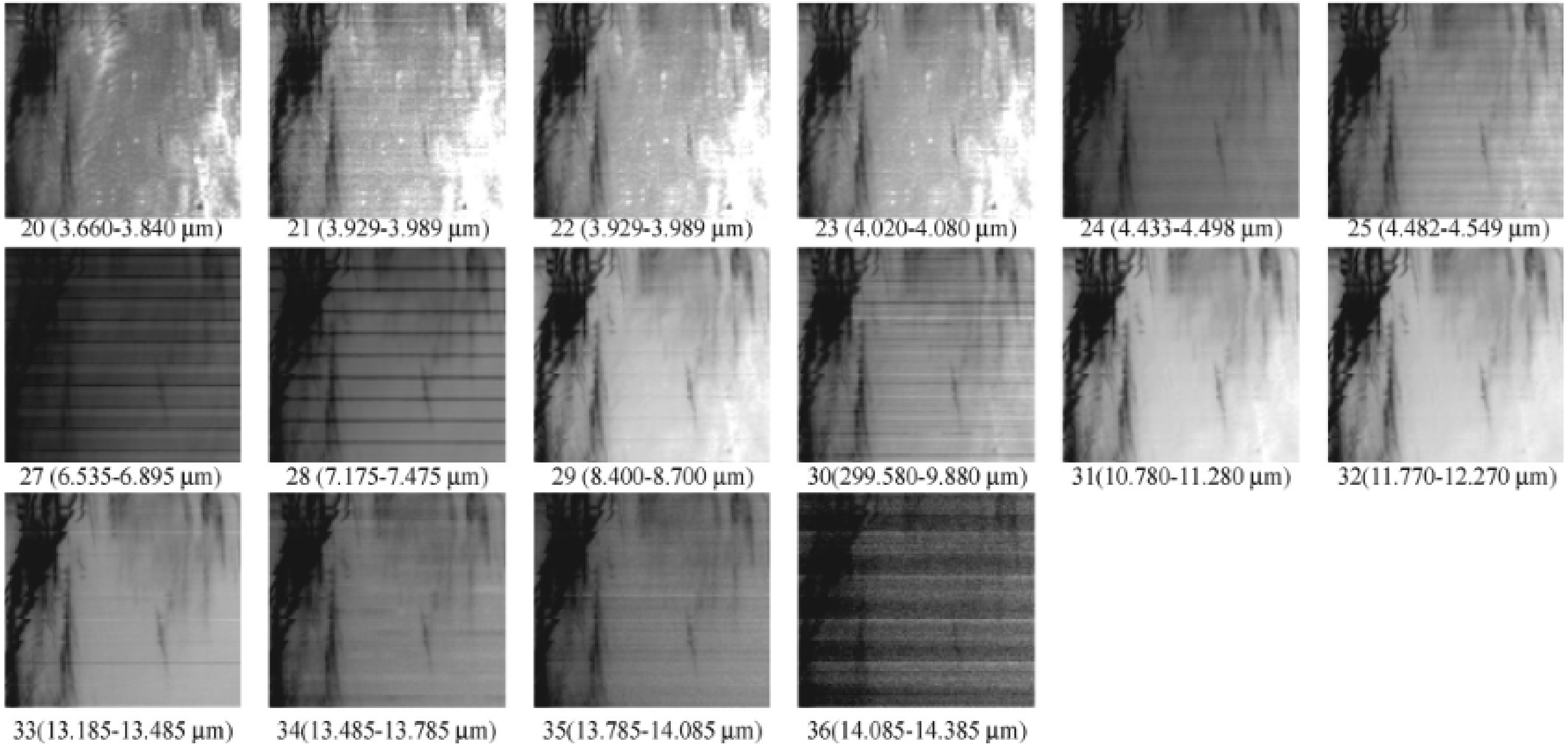
Operator learning as a component of digital twins. FNO used to emulate NEMO simulations of coastal and ocean dynamics[2]

An Earth digital twin - combining MODIS and Cloudsat observations with ECMWF simulations[1]

[1]Bauer et al., *Nature Computational Science*, 2021

[2]Jiang et al., arXiv:2110.07100, 2022

Functional data is always noisy



Terrestrial MODIS data exhibits striped noise pattern[1]

[1]Liu et al., *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*, 2006

Noisy independent variables lead to biased estimates in ordinary least-squares (OLS) operator regression

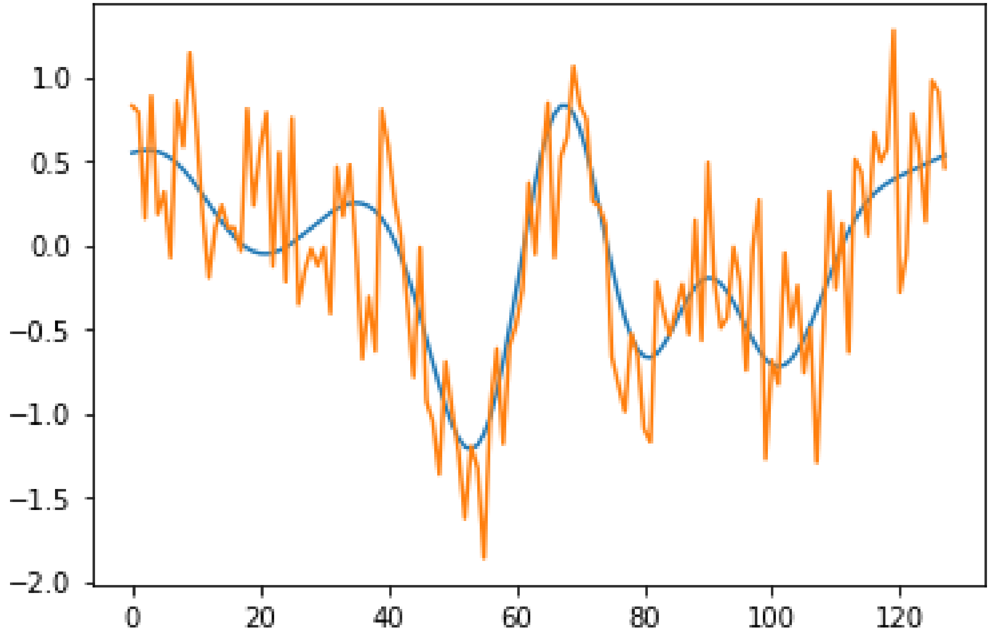
Find $L\hat{u} \approx \partial_x \hat{u}^2$

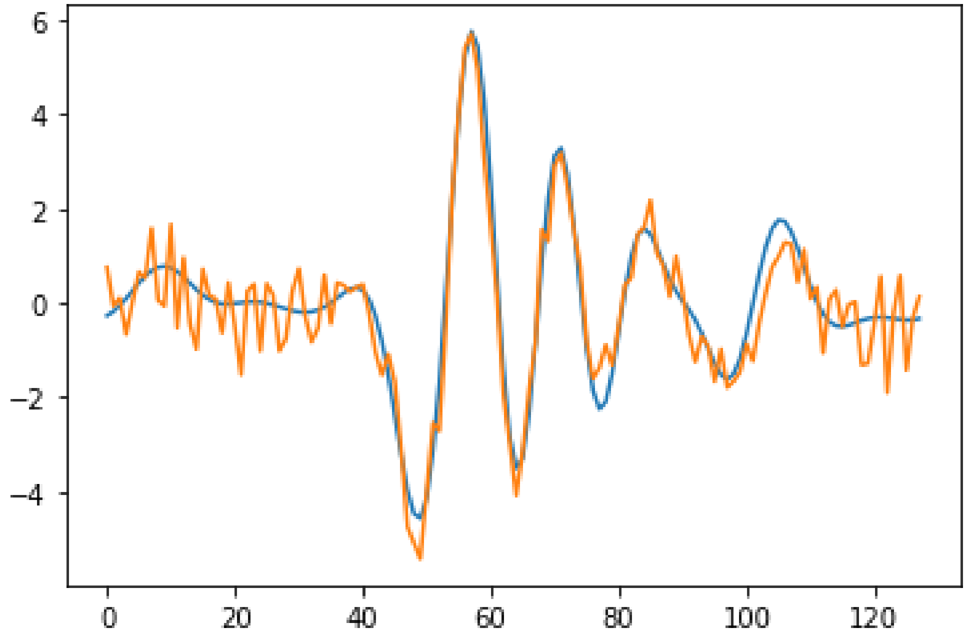
$$u = \hat{u} + \epsilon_u$$

$$\epsilon_u \sim \mathcal{GP}(0, \sigma_u \delta_{x,x'})$$

$$v = \partial_x \hat{u}^2 + \epsilon_v$$

$$\epsilon_v \sim \mathcal{GP}(0, \sigma_v \delta_{x,x'})$$





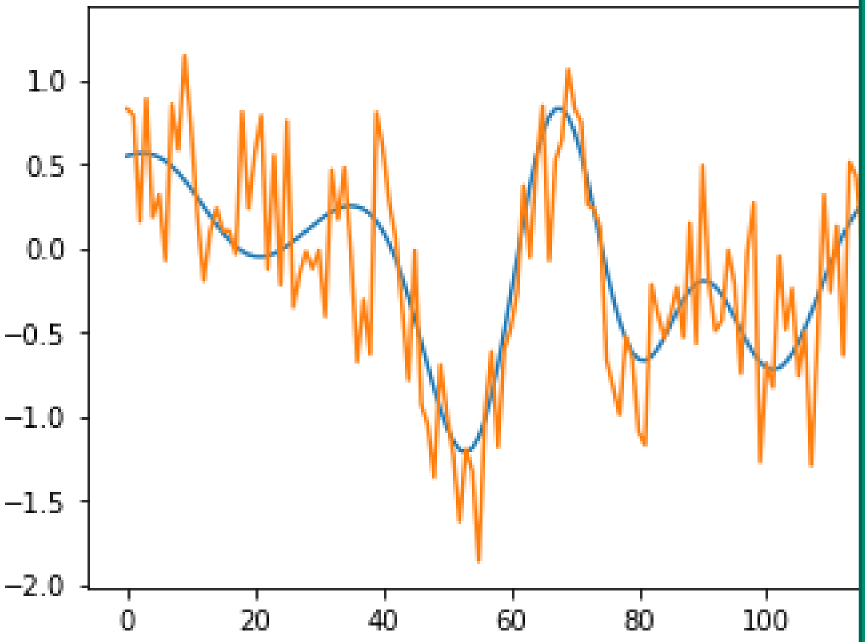
Noisy independent variables lead to biased estimates in ordinary least-squares (OLS) operator regression

OLS:

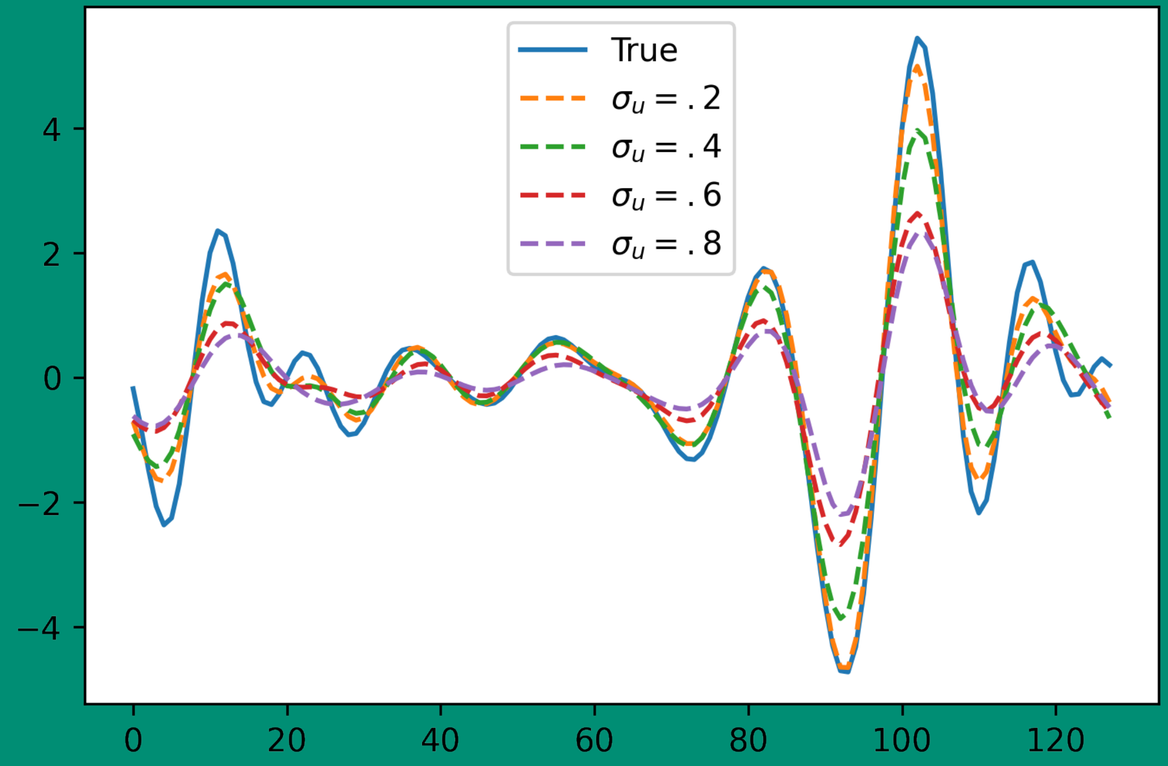
$$\min_{L} ||L(u) - v||_V^2$$

F

$$\epsilon_u \sim \mathcal{GP}(0, \sigma_u \delta_{x,x'})$$

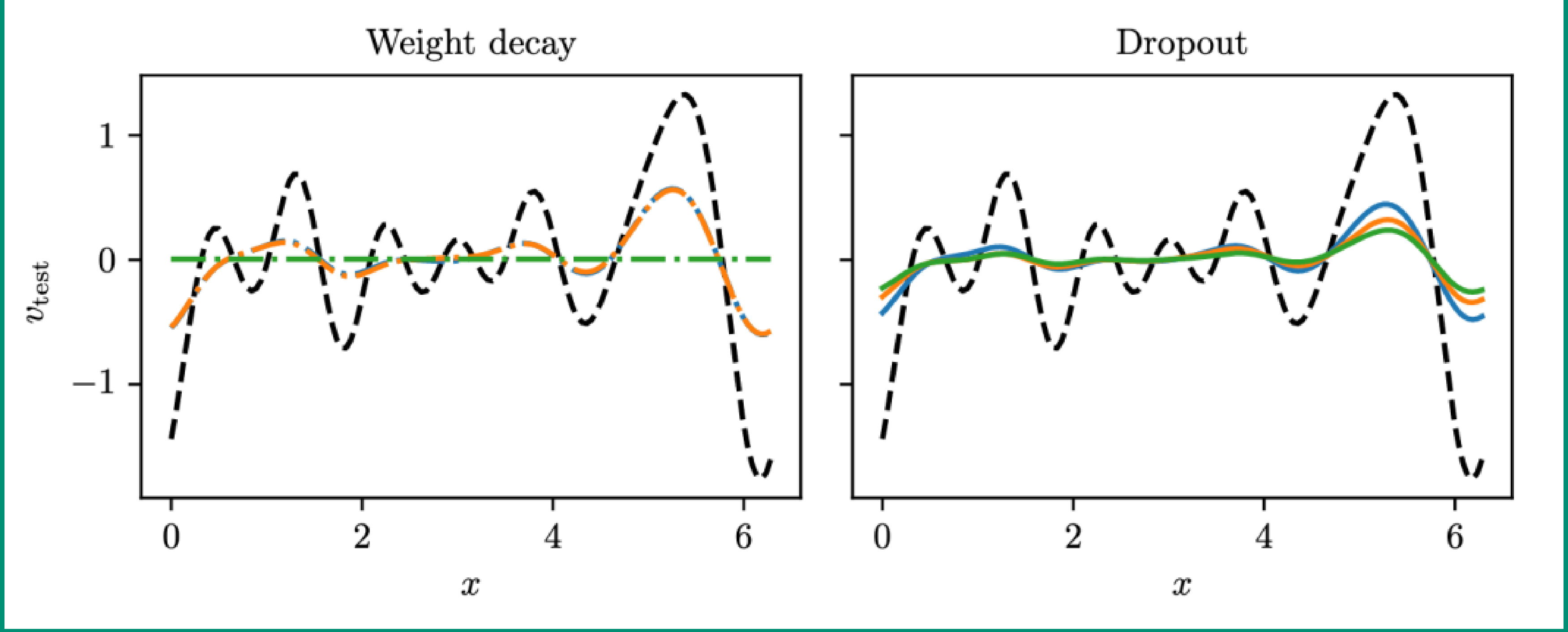


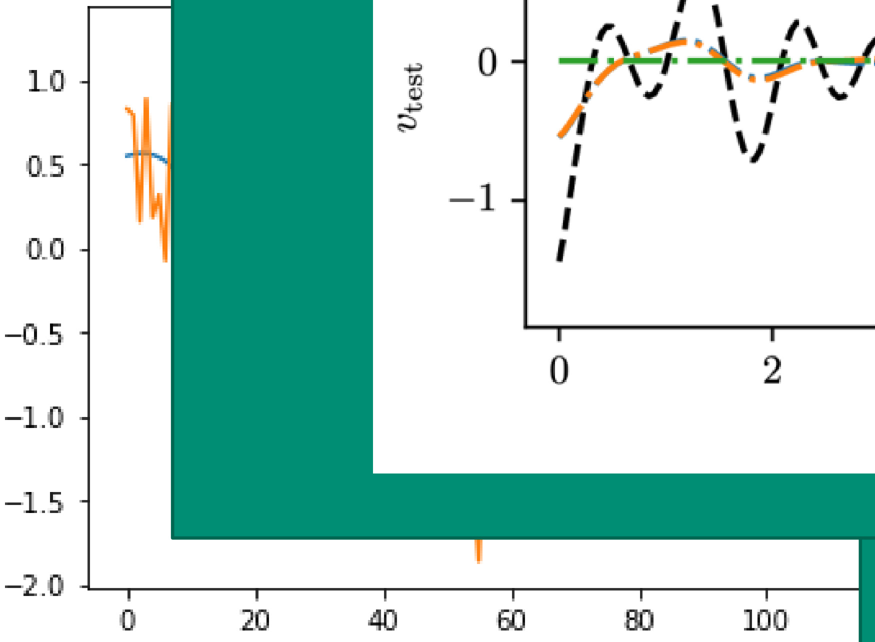
Action of learned operators on noiseless test u:



Noisy independent variables lead to biased estimates in ordinary least-squares (OLS) operator regression

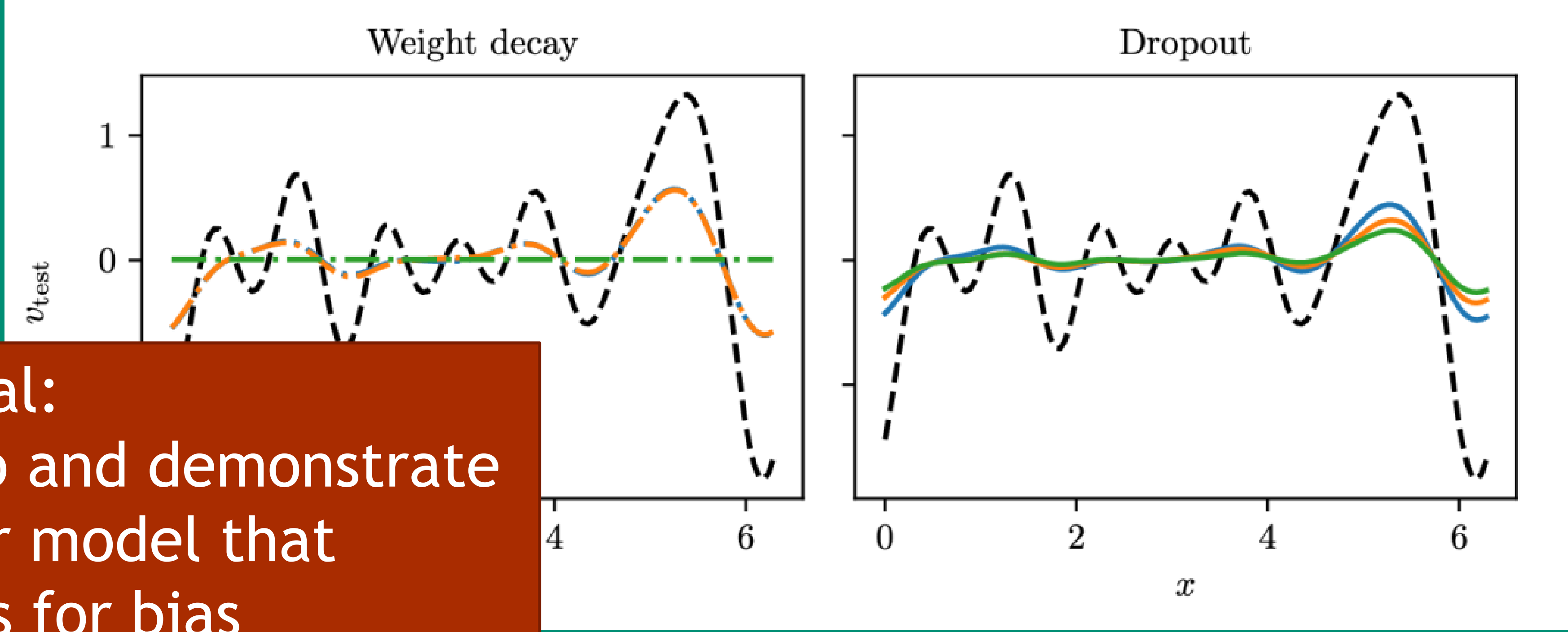
Standard neural network regularization does not remove bias





Noisy independent variables lead to biased estimates in ordinary least-squares (OLS) operator regression

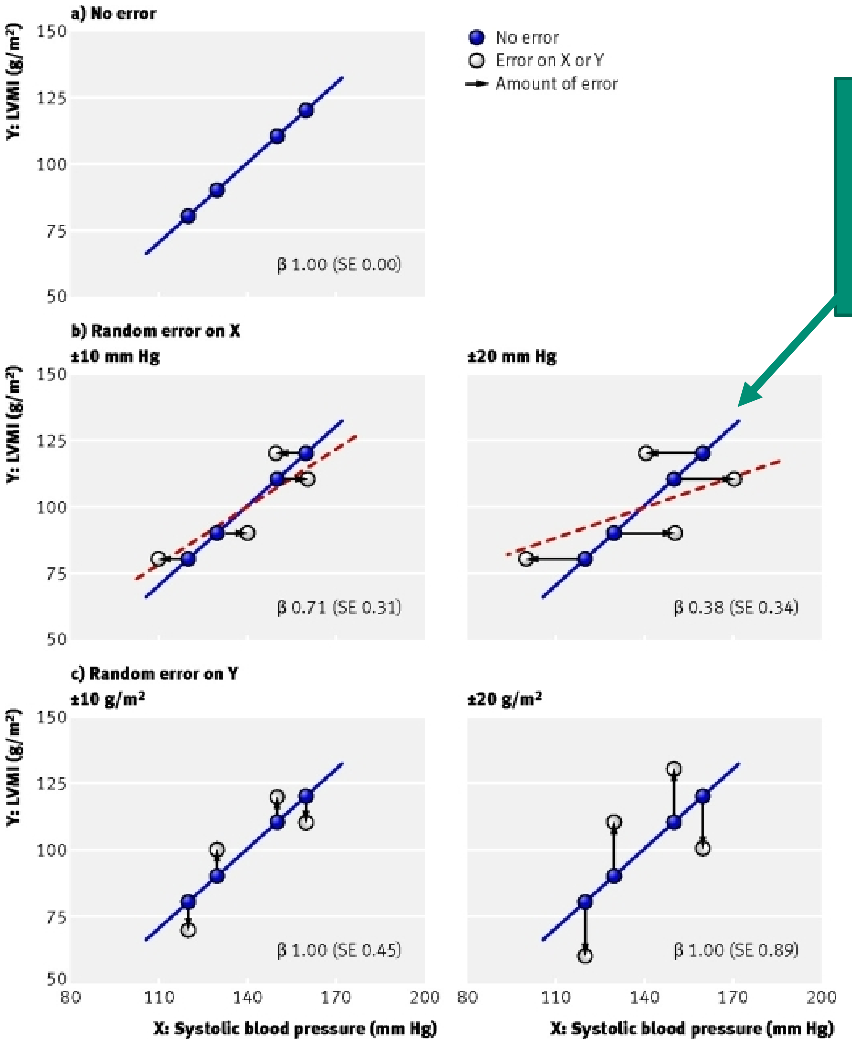
Standard neural network regularization does not remove bias



Project goal:

- Develop and demonstrate an error model that corrects for bias
- Assuming smooth functions and white noise

Attenuation bias for scalar linear regression



Slope is underpredicted by OLS with error in x

Hutcheon, Chiolero, and Hanley, *BMJ*, 2010

True slope vs. predicted slope via OLS:

$$m^* = m \frac{\mathrm{var}(x)}{\mathrm{var}(x) + \mathrm{var}(\epsilon_x)}$$

Error-in-variable (EiV) models for standard regression

- Given,

  (x, y) where $x = \hat{x} + \epsilon_x$ and $y = f(\hat{x}) + \epsilon_y$

- Find f
- Tools are narrowly tailored
- Deming regression and total least-squares – variance/covariance must be known
- Thesis with review of EiV models: Zwanzig, *Estimation in nonlinear functional error-in-variables models,* 1997

Generalization of attenuation bias to discrete linear operators

Given (u, v) where $u = \hat{u} + \epsilon_u$ and $v = L(\hat{u}) + \epsilon_v$

Let U, V be finite dimensional and L be linear

Assume enough data such that the sample statistics converge

The optimum of the OLS problem,

$$\min_L ||L(u) - v||_V^2$$

is $L = E[vu^T](E[uu^T] + \sigma_u I)^{-1}$

With norm upperbound,

$$||L|| \leq \frac{||E[vu^T]||}{||E[uu^T] + \sigma_u I||}$$

EiV model for operator regression

Error model,

$$\begin{bmatrix} \tilde{u}^i - u^i \\ L\tilde{u}^i - v^i \end{bmatrix} \sim \mathcal{GP}\left(0, \begin{bmatrix} \sigma_u \delta_{x,x'} & 0 \\ 0 & \sigma_v \delta_{x,x'} \end{bmatrix}\right)$$

Use maximum likelihood estimation (MLE)

$$\max_{L,\tilde{u},\sigma_u,\sigma_v} \prod_i P\left(\begin{bmatrix} \tilde{u}^i - u^i \\ L\tilde{u}^i - v^i \end{bmatrix}\right)$$

Assume $\hat{u}$ is a smooth function and introduce a filter

$$\max_{L,\mathcal{G},\sigma_u,\sigma_v} \prod_i P\left(\begin{bmatrix} \mathcal{G}u^i - u^i \\ L\tilde{u}^i - v^i \end{bmatrix}\right)$$

EiV model for operator regression

Error model,

$$\begin{bmatrix} \tilde{u}^i - u^i \\ L\tilde{u}^i - v^i \end{bmatrix} \sim \mathcal{GP}\left(0, \begin{bmatrix} \sigma_u \delta_{x,x'} & 0 \\ 0 & \sigma_v \delta_{x,x'} \end{bmatrix}\right)$$

Use maximum likelihood estimation (MLE)

$$\max_{L,\tilde{u},\sigma_u,\sigma_v} \prod_i P\left(\begin{bmatrix} \tilde{u}^i - u^i \\ L\tilde{u}^i - v^i \end{bmatrix}\right)$$

Methods of parameterizing the operator

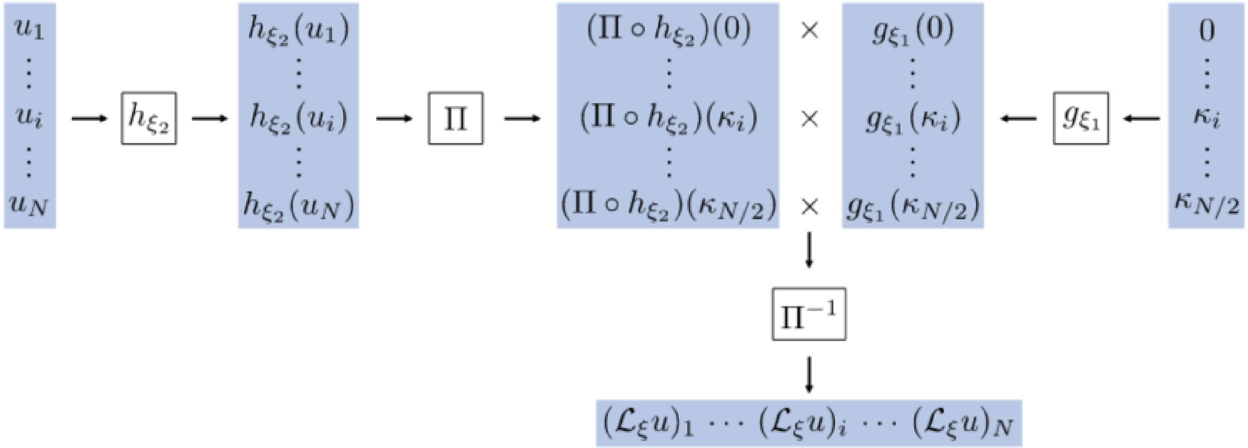
1. MOR-Physics
2. DeepONet

Assume $\hat{u}$ is a smooth function and introduce a filter

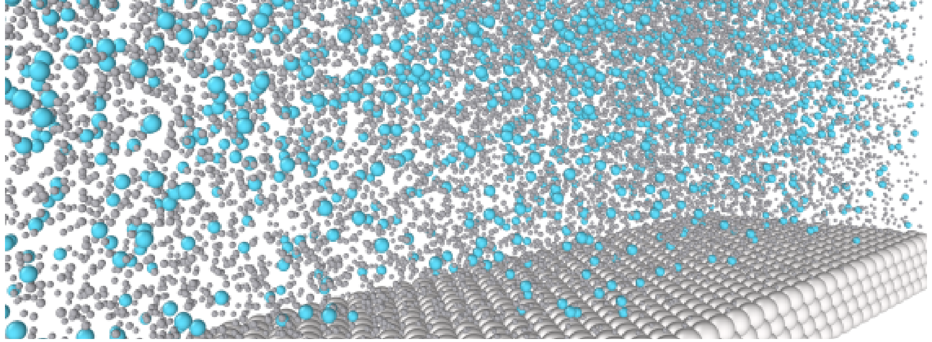
$$\max_{L,\mathcal{G},\sigma_u,\sigma_v} \prod_i P\left(\begin{bmatrix} \mathcal{G}u^i - u^i \\ L\tilde{u}^i - v^i \end{bmatrix}\right)$$

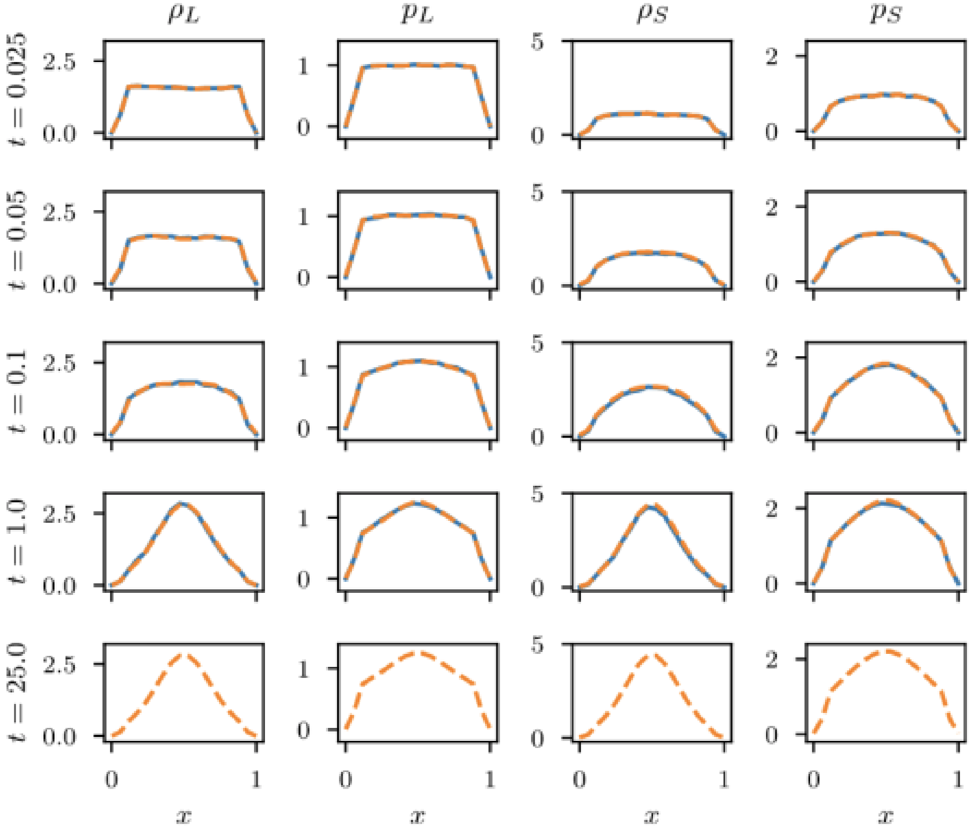
Operator learning methods

MOR-Physics[1,2]



MOR-Physics parameterization



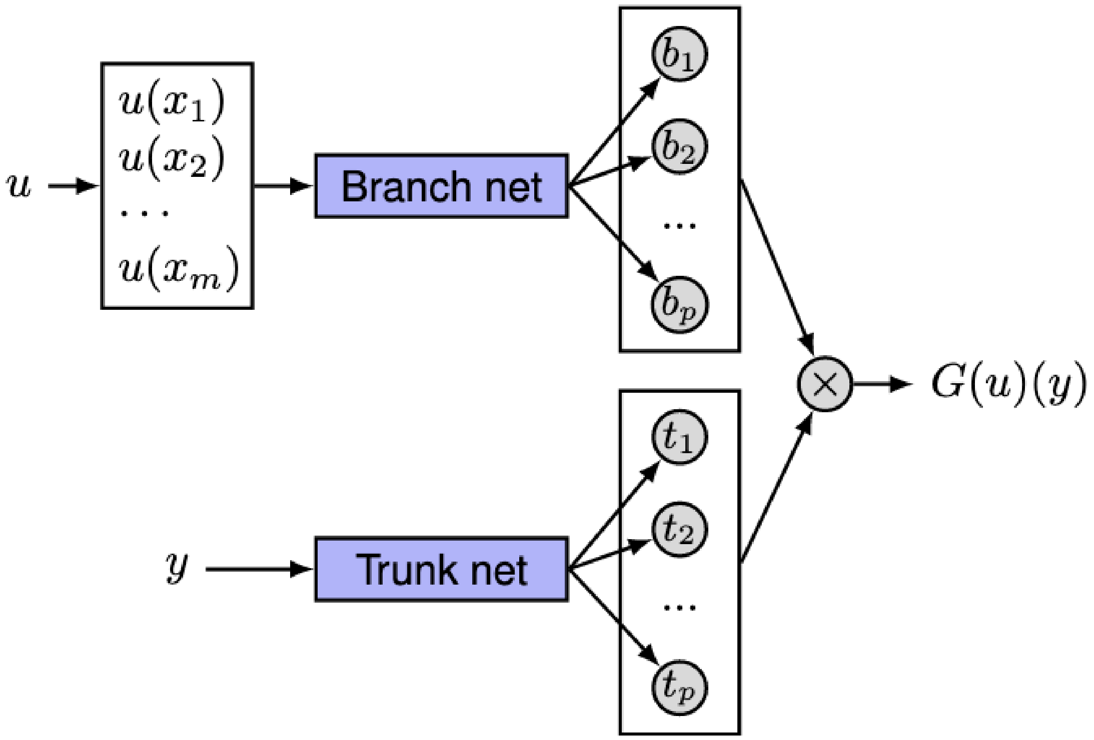


MOR-physics learns dynamics of colloidal system from molecular dynamics simulations. Generalizes to unseen concentration and colloid diameter

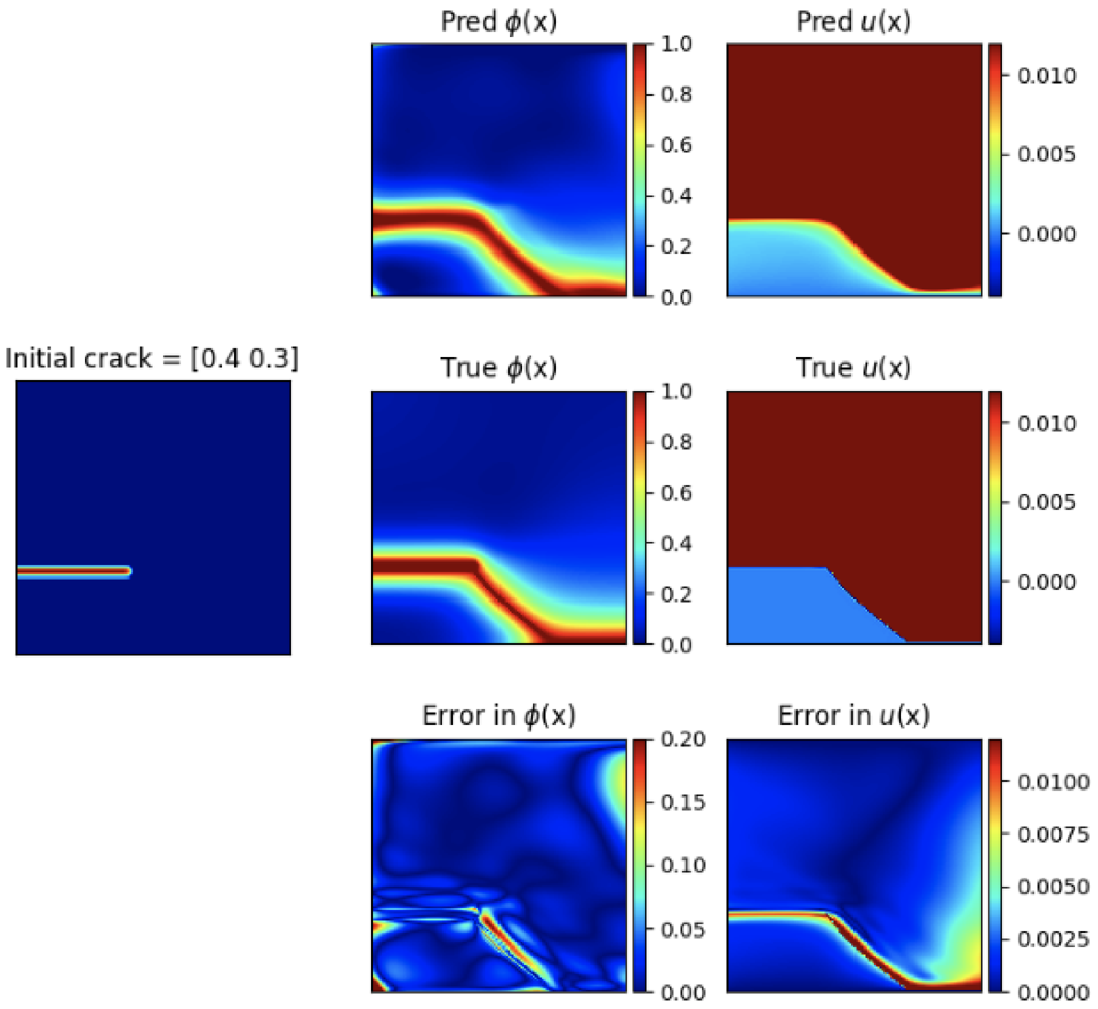
[1]Patel and Desjardins, arXiv:1810.08552, 2018
[2]Patel et al., *CMAME*, 2021

Operator learning methods

DeepONet[1]



Unstacked DeepONet parameterization



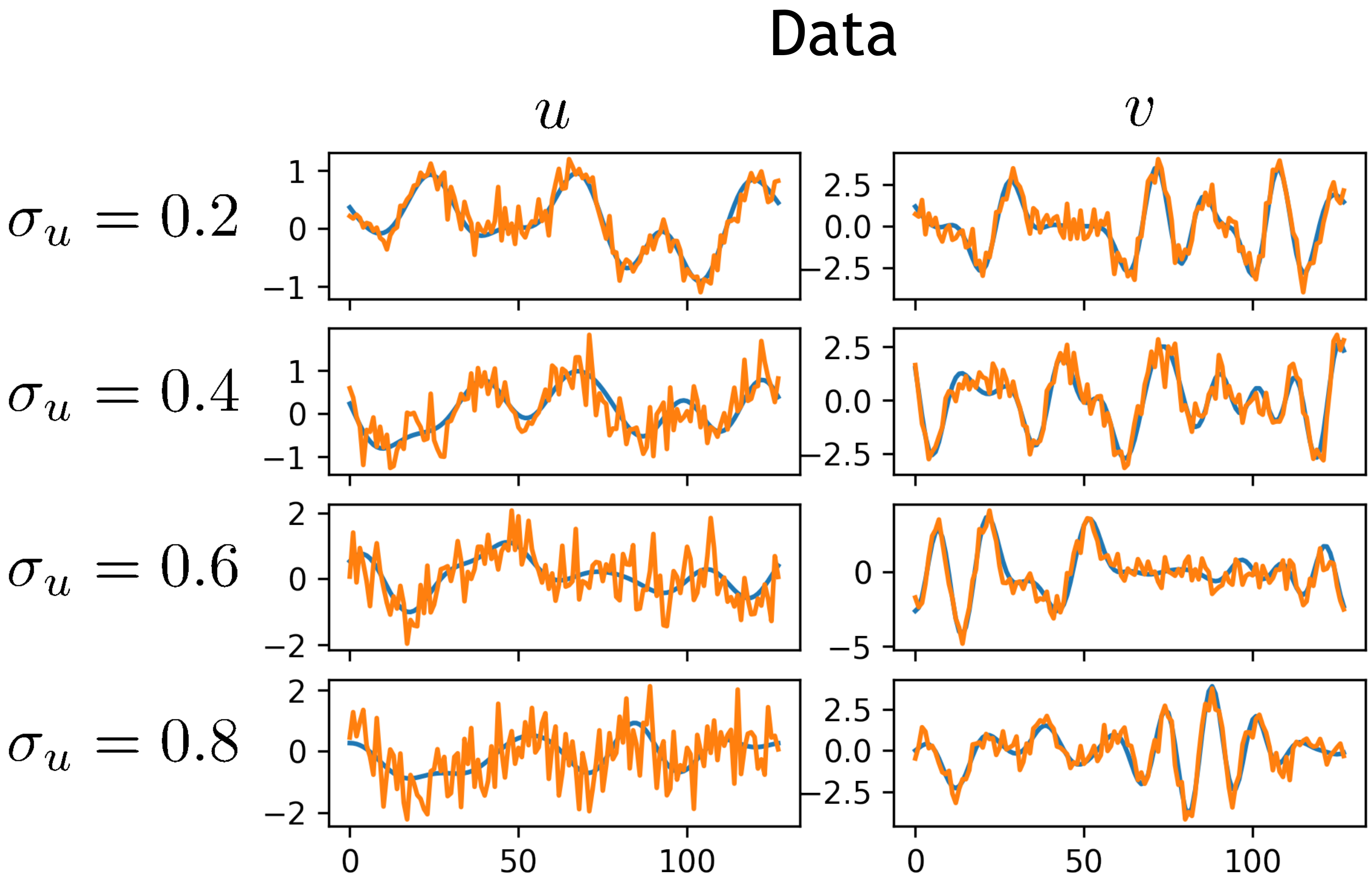
Variational DeepONet learns crack path under shear loading. Generalizes to unseen crack tip locations.[2]

[1]Lu, Jin and Karniadakis, *Nature*, 2021
[1]Goswami et al., *CMAME*, 2022

Learning the Burgers operator with noisy input

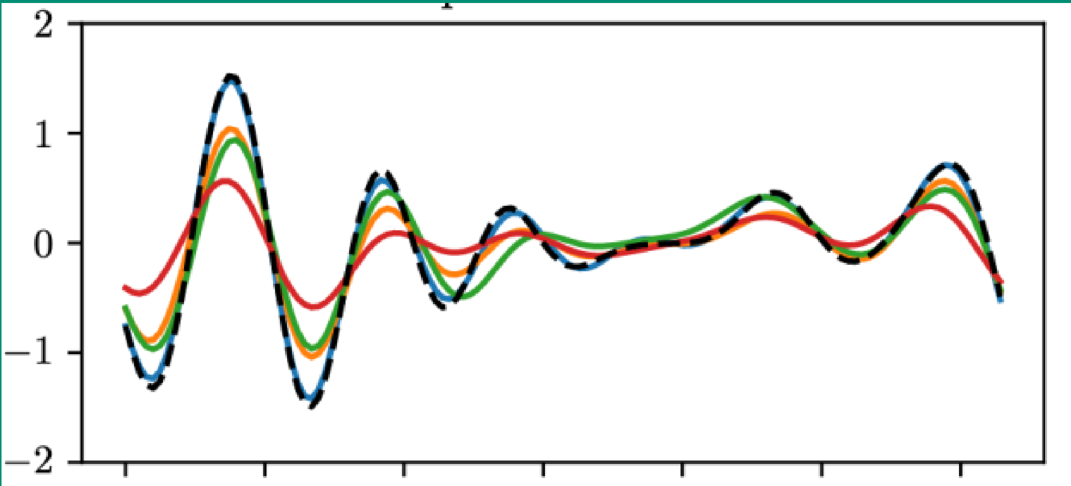
Recover Burgers operator,

$$Lu = \partial_x u^2$$

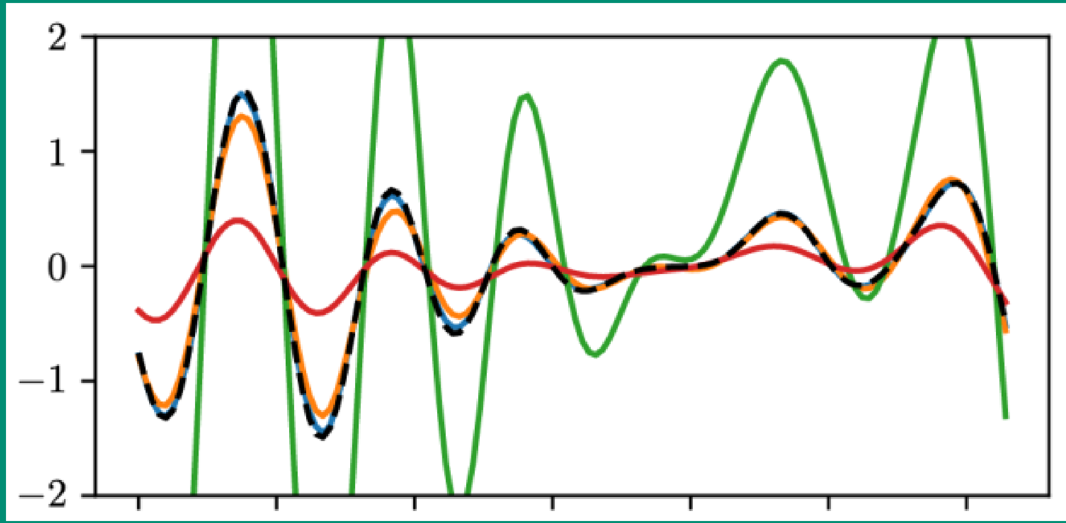


EiV model reduces attenuation bias in learning the Burgers operator – MOR-Physics

Action of OLS operator on clean u



Action of EiV operator on clean u



EiV model improves recovery of true Burgers operator in the presence of noisy independent variables. *(Left)* Underlying smooth function $\hat{u}$ (----) and training u for SNR $= 8$ (——), SNR $= 4$ (——), SNR $= 0$ (——), and SNR $= -4$ (——). *(Right)* Action of true Burgers operator (----) on noiseless test $\mathbf{u}_{\text{test}}$ and action of learned operators from data with decreasing SNR for OLS *(Top right)* and EiV *(Bottom right)*.

Smoothness prior

Use a smooth spectral filter, $\mathcal{G}u = \mathcal{F}^{-1}\mathrm{erfc}(a(\kappa - \kappa_c))\mathcal{F}u$

Use a Beta distribution for prior (approximation to $\mathcal{U}(0, 1)$),

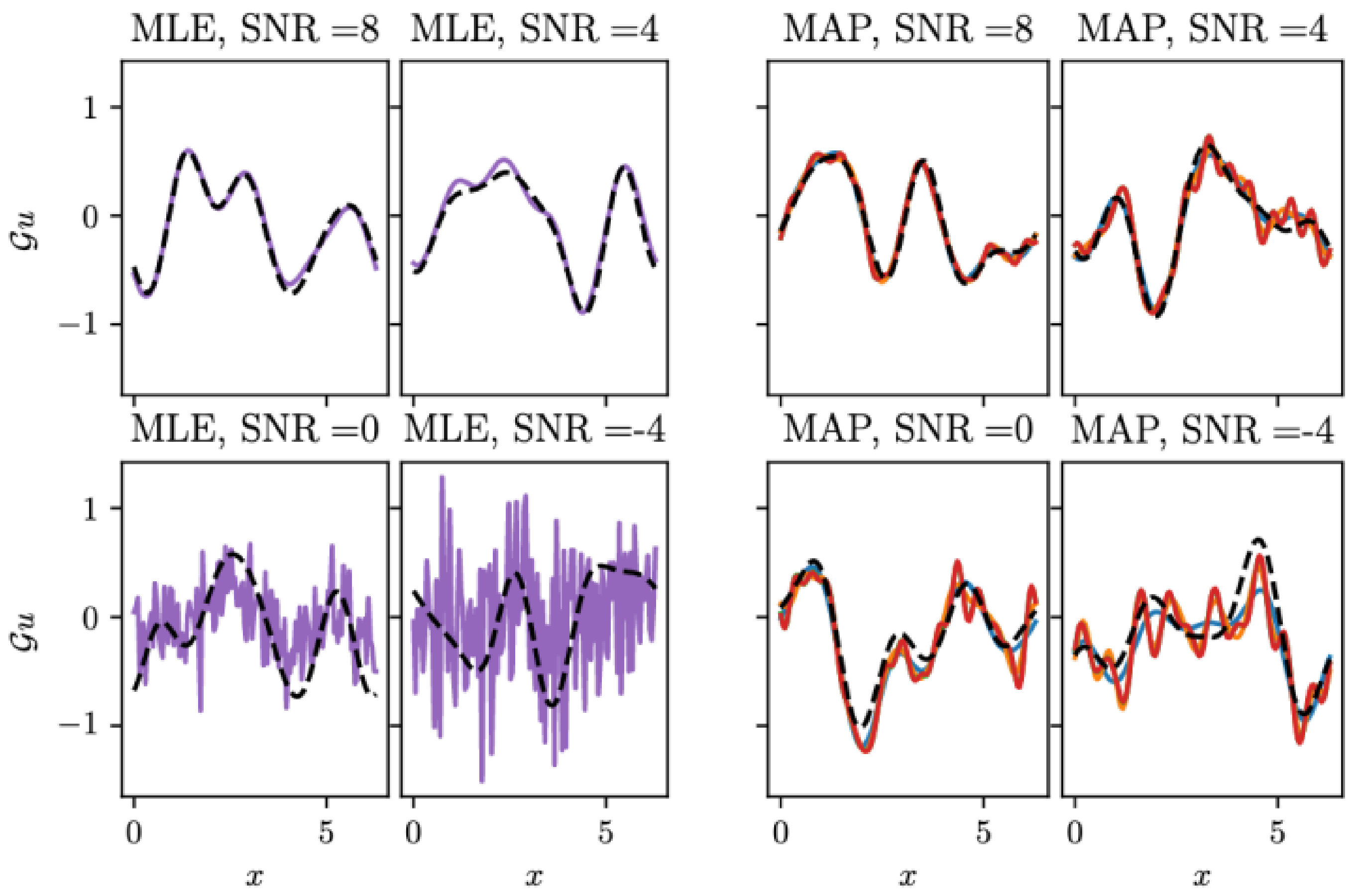
$$\kappa_c/\beta \sim \mathrm{Beta}(1+\delta, 1+\delta)$$

Maximum a posteriori estimation (MAP),

$$\max_{L,\mathcal{G},\sigma_u,\sigma_v,\kappa_c} \prod_i P\left(\begin{bmatrix}\mathcal{G}u^i - u^i \\ L\tilde{u}^i - v^i\end{bmatrix}\right) P(\kappa_c/\beta)$$

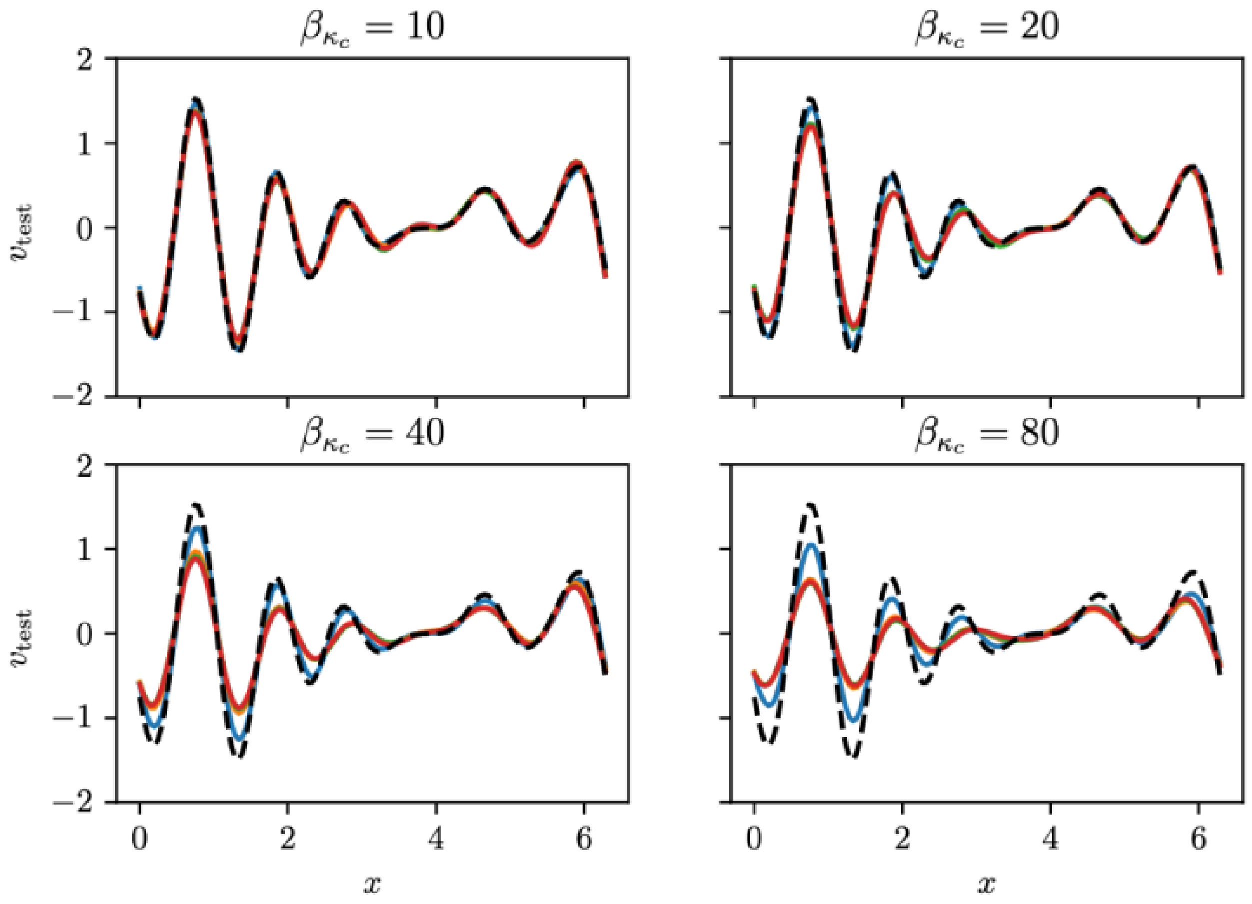
a, δ, β are hyperparameters

Effect of smoothness prior on filter



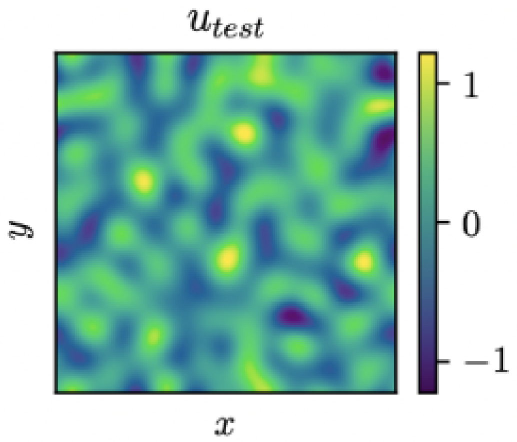
Effect of cutoff wavenumber prior on filter for EiV model. *(Left)* Action of MLE estimate of filters on noisy u^i (——) for decreasing SNR and corresponding noiseless $\hat{u}^i$ (- - - -). *(Right)* Action of MAP estimate of filters (κ_c prior) on u^i with hyperparameters, $\beta_{\kappa_c} = 10$ (——), $\beta_{\kappa_c} = 20$(——), $\beta_{\kappa_c} = 40$(——), and $\beta_{\kappa_c} = 80$(——).

Smoothness prior robustly recovers operator with EiV model and is insensitive to hyperparameters

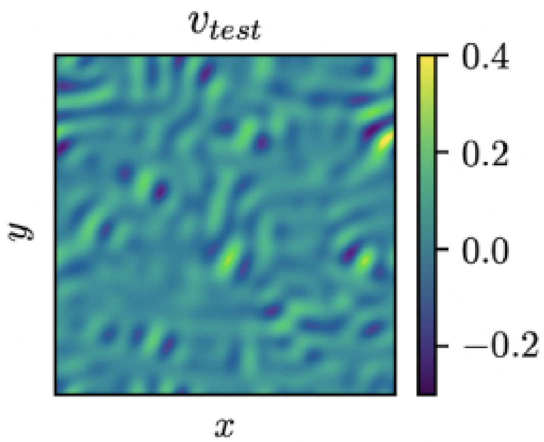


Cutoff wavenumber prior improves EiV model. Action of EiV operator on u_{test} learned from SNR = 8 (——), SNR = 4 (——), SNR = 0 (——), and SNR = −4 (——) for various β_{κ_c}. Action of true operator (----).

EiV model reduces attenuation bias in learning the 2D Burgers operator – MOR-Physics

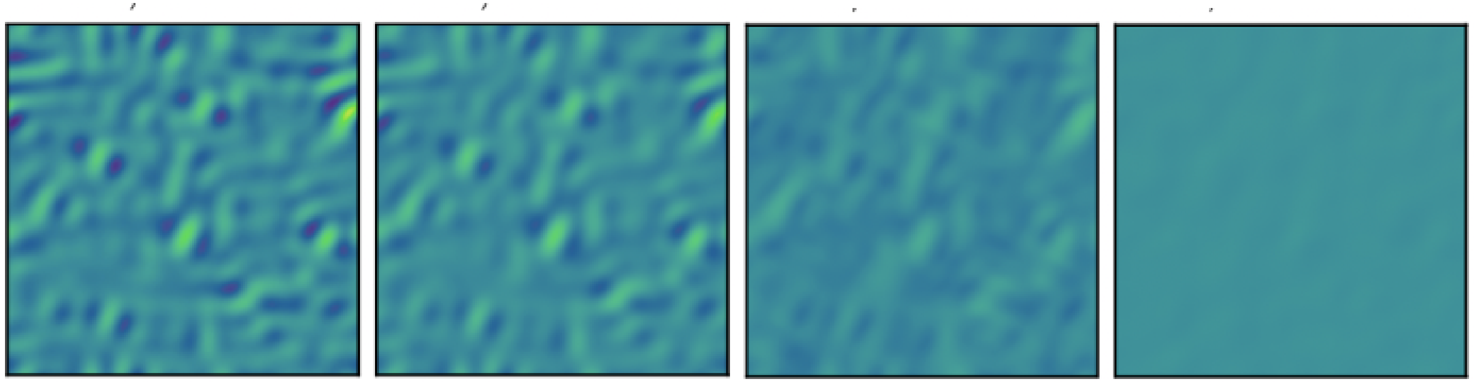


Noise free test function

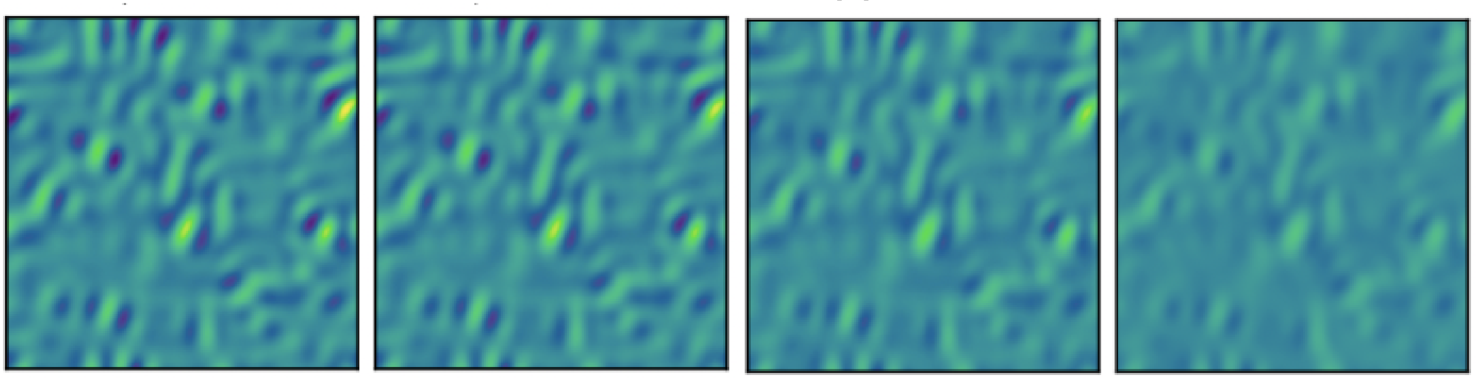


True operator's action

Increasing Signal to noise (SNR) in training data



OLS learned model applied to u_{test}



EiV learned model applied to u_{test}

Extension of operator inference and EiV model to time-dependent PDEs

For PDEs of the form,

$$\partial_t \hat{u} = \mathcal{L}\hat{u}$$

We seek to infer $\mathcal{L}$ given time independent 2white noise corrupted solutions,

$$u = \hat{u} + \epsilon_u$$

The OLS loss is computed as,

$$\min_{\mathcal{L}} ||\mathcal{P}(u(t=0), t_f) - u(t_f)||_U^2$$

where $\mathcal{P}$ is the evolution operator for the PDE (approximated via forward Euler)

The EiV model is

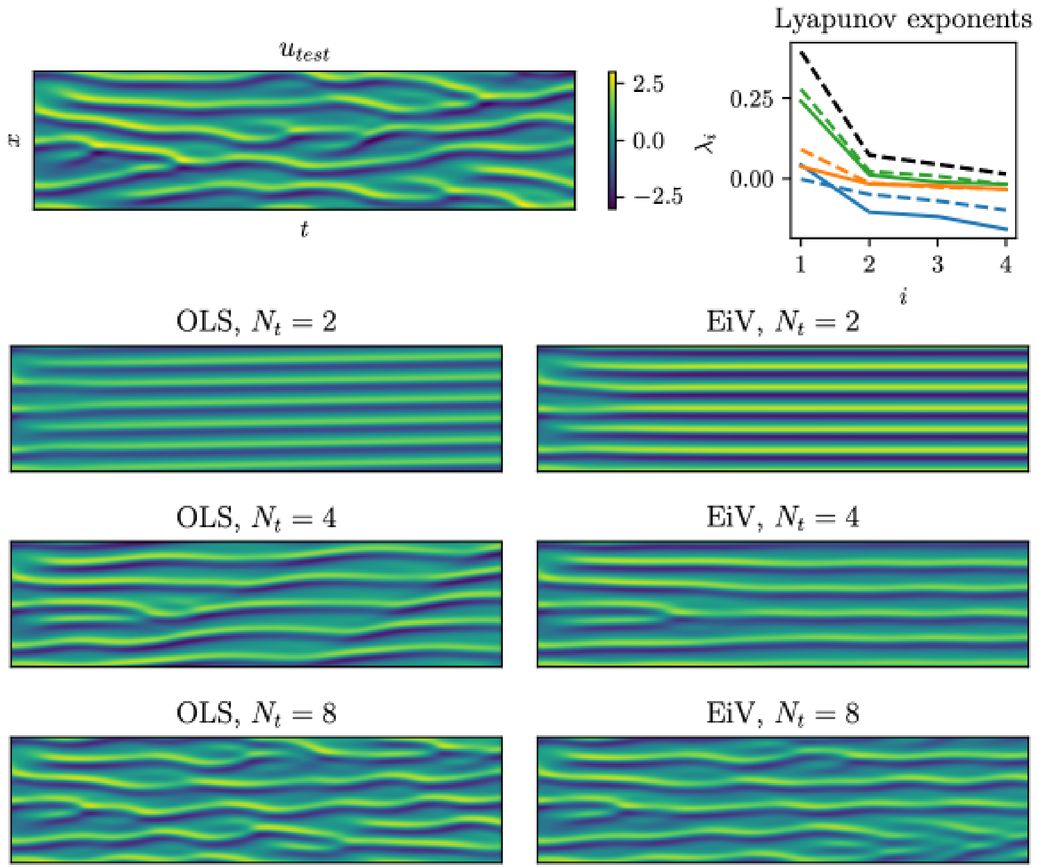
$$\begin{bmatrix} \mathcal{G}u(\cdot,0)^i - u(\cdot,0)^i \\ L\mathcal{G}u(\cdot,t_f)^i - u(\cdot,t_f)^i \end{bmatrix} \sim \mathcal{GP}\left(0, \begin{bmatrix} \sigma_u \delta_{x,x'} & 0 \\ 0 & \sigma_u \delta_{x,x'} \end{bmatrix}\right)$$

Where MLE and MAP estimation is computed as shown previously

Inferring the Kuramoto–Sivashinsky Equation with EiV vs. OLS – MOR-Physics

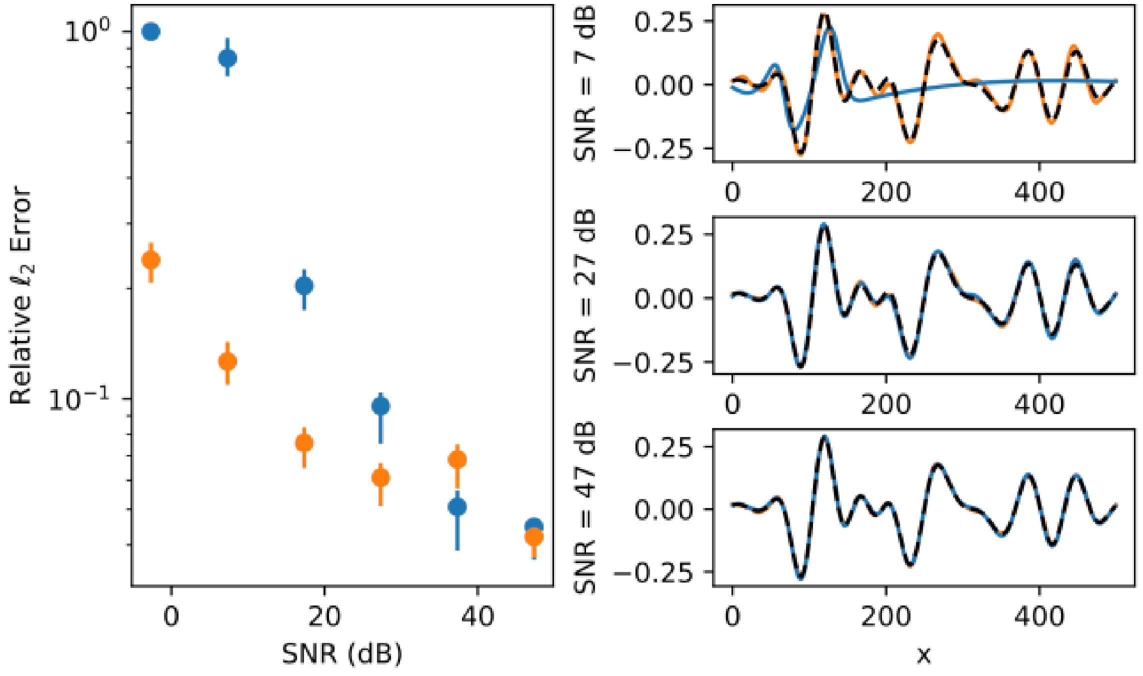
Kuramoto-Sivashinsky Equation:

$$\partial_t u + 0.5\partial_x u^2 + \partial_x^2 u + \partial_x^4 u = 0,$$

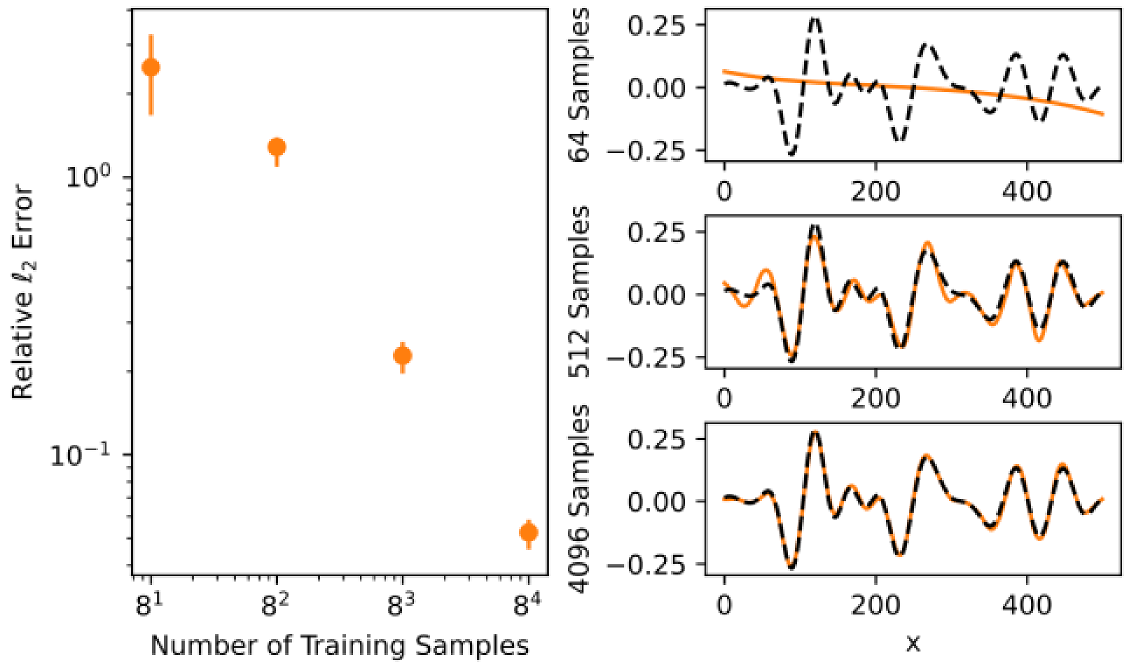


OLS and EiV models perform similarly for KS equation inference. *(Top left)* Noiseless test data, u_{test}. *(Bottom left)* OLS and *(Bottom right)* EiV inferred operators for increasing hyperparameter, N_t. *(Top right)* Lyapunov exponents for true equation (----); OLS equation with $N_t = 2$ (----), $N_t = 4$ (----), $N_t = 8$ (----); and EiV equation with $N_t = 2$ (——), $N_t = 4$ (——), $N_t = 8$ (——).

Statistics on learning the Burgers operator with EiV vs. OLS – DeepONets

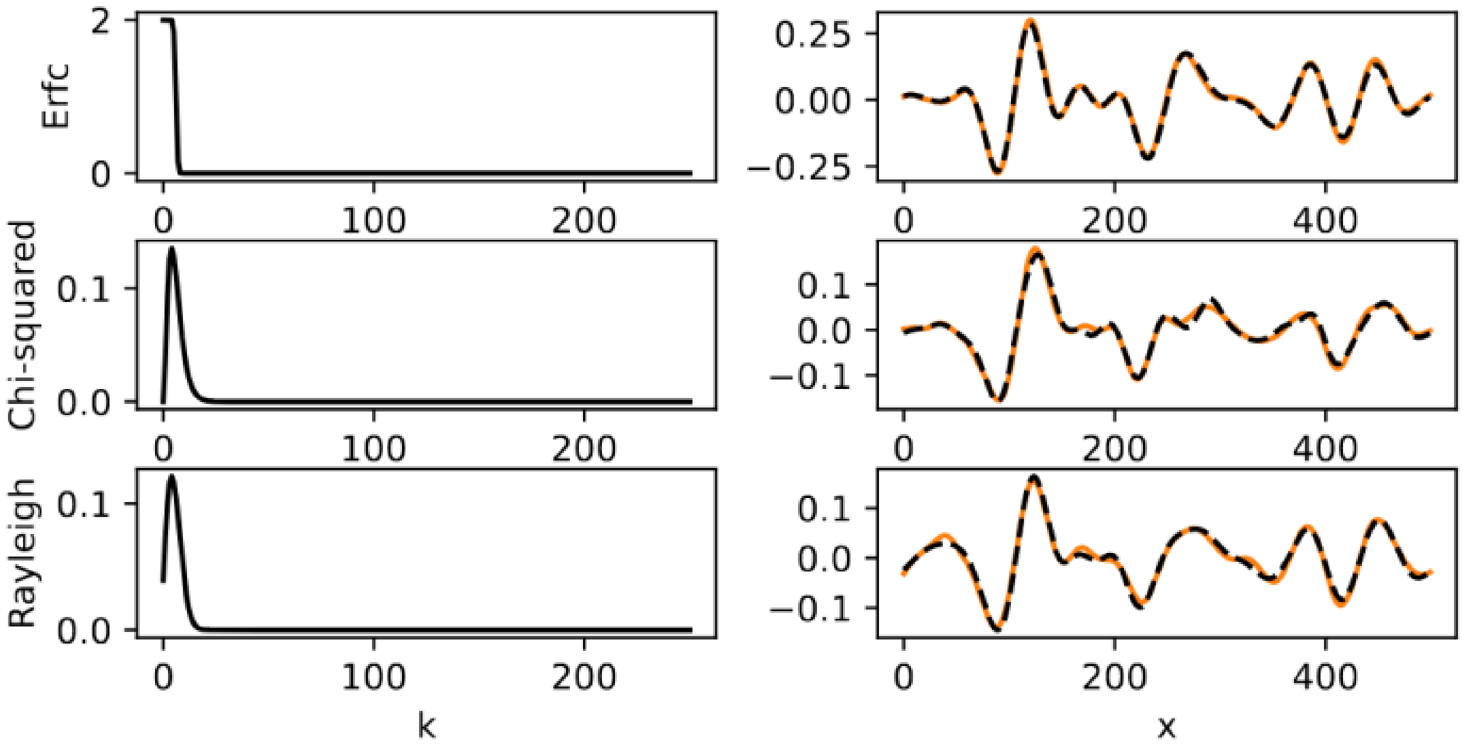


OLS (●) and EiV (●) training error vs. SNR



EiV (●) training error vs. number of samples

EiV model is robust to various distributions of the smooth underlying input functions – DeepONet



Effect of spectral filter used to generate input signals. *(Left)* Frequency content of u_{test}. *(Right)* Action of true (----) and DeepONet EiV (——) operators on u_{test}.

Conclusion

Failure to account for error in the independent variables leads to biased estimates for operator regression

Developed an error-in-variables model to correct for bias

Demonstrated this error model with MOR-Physics and DeepONet

Future work

- Explore the full posterior distribution of operators
  - Besides the MAP, how do other plausible operators behave?
  - UQ - The action of operators sampled from the posterior will give error bars
- Other error models, e.g. multiplicative noise
- Relax smoothness assumption

Manuscript,

- Patel et al., arXiv:2204.10909, 2022