Sandia National Laboratories

Exceptional service in the national interest

SCIENTIFIC MACHINE LEARNING AND TENSORFLOW TUTORIAL

Additional topics and Reproducibility

Ravi G Patel

Scientific Machine Learning Department

February 1 – 2, 2024

Numerical PDEs: Analysis, Algorithms, and Data Challenges

ICERM

Brown University

SAND2024-00936O

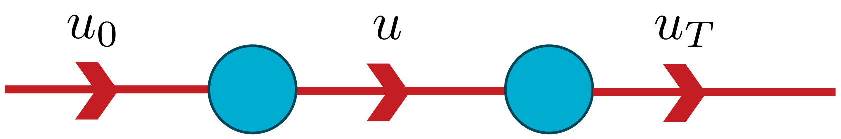
OVERVIEW

- Neural ODE
- Sparse regression
- Attention and transformers
- Physics informed constraints
- Approximate Bayesian methods
- Generative modeling
- Optimal experimental design
- Digital twins
- Reproducibility

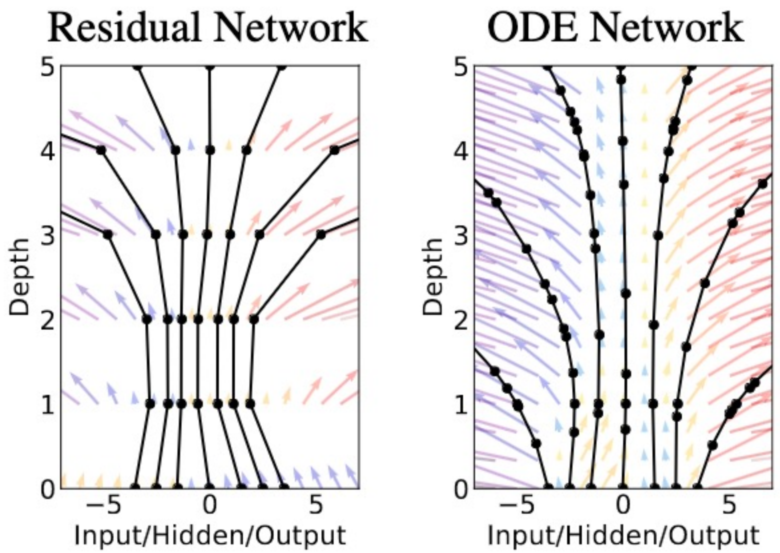
NEURAL ODE (NODE)[1]

- Limit of a neural network to infinite depth
- Backprop can be derived using adjoint sensitivity[2]

$$J(u(t=T))$$
$$\dot{u} = f(u,\xi)$$
$$u(t=0) = u_0$$
$$\mathcal{L} = J(u_T) + (\lambda, \dot{u} - f(u,\xi)) + (\mu, \int \delta(t)u(t)dt - u_0) + (\zeta, \int \delta(t-T)u(t)dt - u_T)$$



Computation graph for a NODE

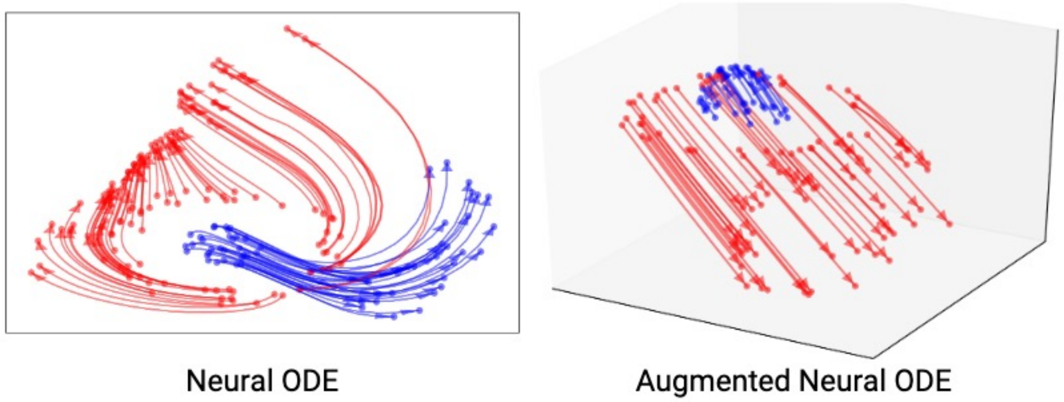


Comparison between ResNet and NODE[1]

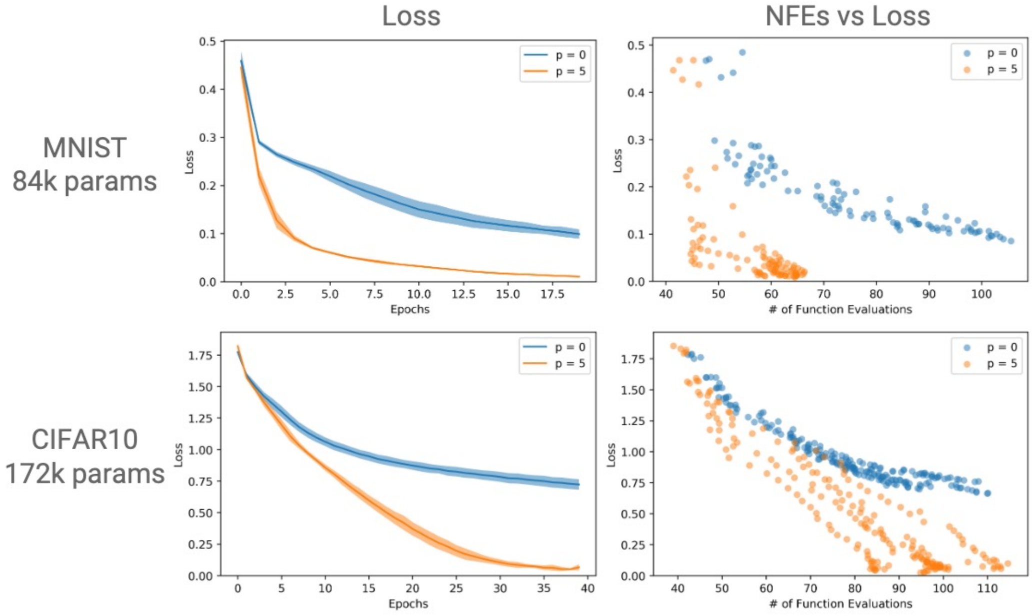
[1]R. Chen et al., *NeurIPS* (2018)
[2]A. Bradley, PDE-constrained optimization and the adjoint method (2019)

AUGMENTED NODE[1]

- Augmenting a NODE with auxiliary variables can simplify learned model



Learned flows for NODE and Augmented NODE[1]

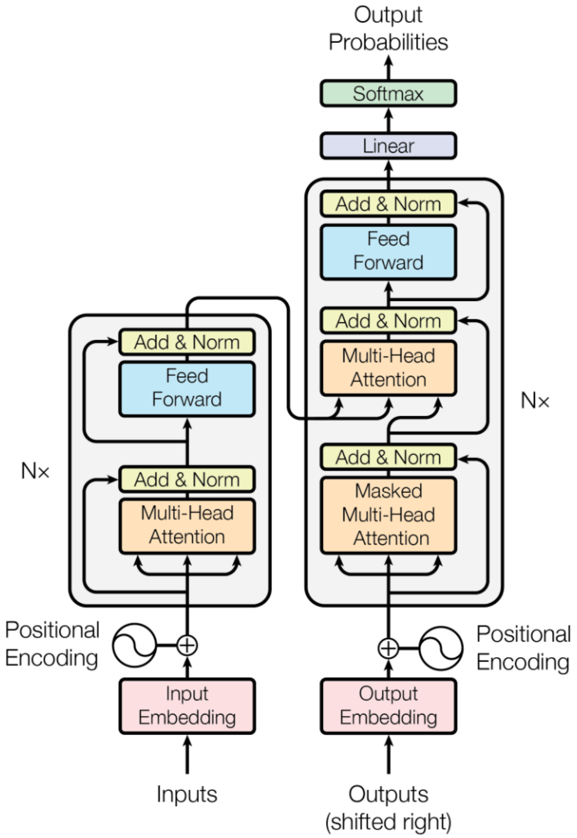


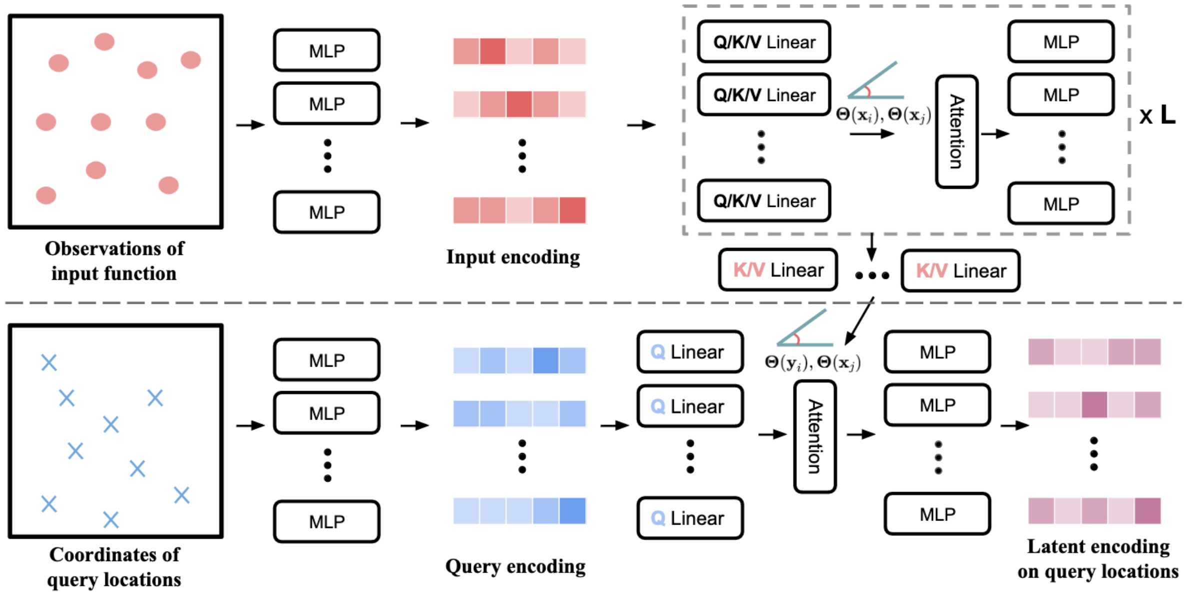
NODE vs. Augmented NODE for Classification[1]

[1]E. Dupont et al., *NeurIPS* (2019)

TRANSFORMERS

- Alternative to RNN's
- From sequences of data, correlates parts of it together. Given a new sequence of data, predicts the next piece





Original Transformer architecture for NLP
Vaswani et al., *NeurIPS* (2017)

Transformer for Operator learning
Zi et al., *TMLR* (2023)

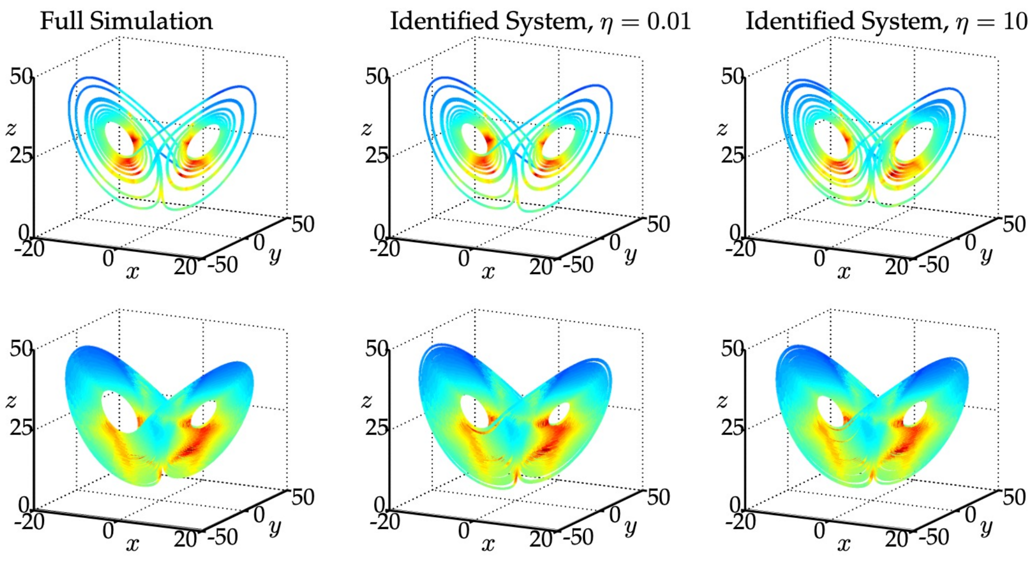
SPARSE REGRESSION - SPARSE IDENTIFICATION OF NONLINEAR DYNAMICAL SYSTEMS (SINDY)

Brunton et al., *PNAS*, 2016,

$$\mathcal{N}\{u\} = a_1 u + a_2 u^2 + \ldots$$
$$+ b_1 \partial_x u + b_2 \partial_x u^2 + \ldots$$
$$+ c_1 \sin u + c_2 \sin 2u + \ldots$$

Find coefficients using sparse regression

$$\min ||\mathcal{N}(u) - v_d||^2 + \eta||a_i||_1$$



SINDy finds the Lorenz system

PHYSICS INFORMED CONSTRAINTS

- Many SciML methods can incorporate *a priori* known information
- Hyperbolicity preservation for learning the equations of state for Euler
  - Patel et al., *CMAME* (2021)
  - Entropy is a convex function of volume and internal energy
- Other examples,
  - Conservation laws
  - Mass density is non-negative
  - Stress is positive semi-definite
  - Current vs. voltage is monotonic through a passive circuit component

PHYSICS INFORMED CONSTRAINTS - EQUIVARIANCE

- Often we'll have models with equivariance,
  - If we have a model,

$$v = \mathcal{N}(u)$$

$$v \in \mathcal{V}, u \in \mathcal{U}$$

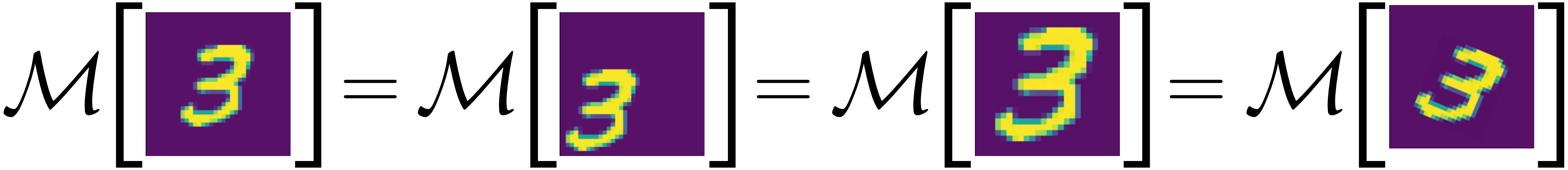
  - It is equivariant w.r.t. a group, G , if for representations, $\rho_{\mathcal{U}}, \rho_{\mathcal{V}}$,

$$\rho_{\mathcal{V}}(g)v = \mathcal{N}(\rho_{\mathcal{U}}(g)u) \quad \forall g \in G$$

  - E.g., SO(3) equivariance. If we rotate the input data and push it through the model, we'll get the same answer as we would if we left the input alone, but rotated the output
  - SO(3) is a subgroup of the Galilean group
    - Frame indifference in mechanics

EQUIVARIANCE IN IMAGE CLASSIFICATION

- Translation, scaling, and rotation shouldn't affect an image's class



Data augmentation: train with transformed versions of training data

- How thoroughly should transformations be sampled?
- Increased cost of training

Choose model form to have desired invariance/equivariance

- E.g. ConvNets for approximate translational invariance[1]

[1] Lawrence et al. *IEEE Transactions on Neural Networks*, 1997

OTHER EXAMPLES OF EQUIVARIANT MODELS

- Rotation invariant model for galaxy classification
  - Dieleman et al. *Monthly Notices of the Royal Astronomical Society*, 2015
- Warp invariant model
  - Wong et al. *DICTA*, 2016
- Permutation invariant model
  - Meltzer et al. *arXiv:1905.03046*
- Rotation and translation equivariant model for 3d point cloud data
  - Thomas et al. *arXiv:1802.08219*

PHYSICS INFORMED CONSTRAINTS FOR OPERATOR LEARNING[1,2]

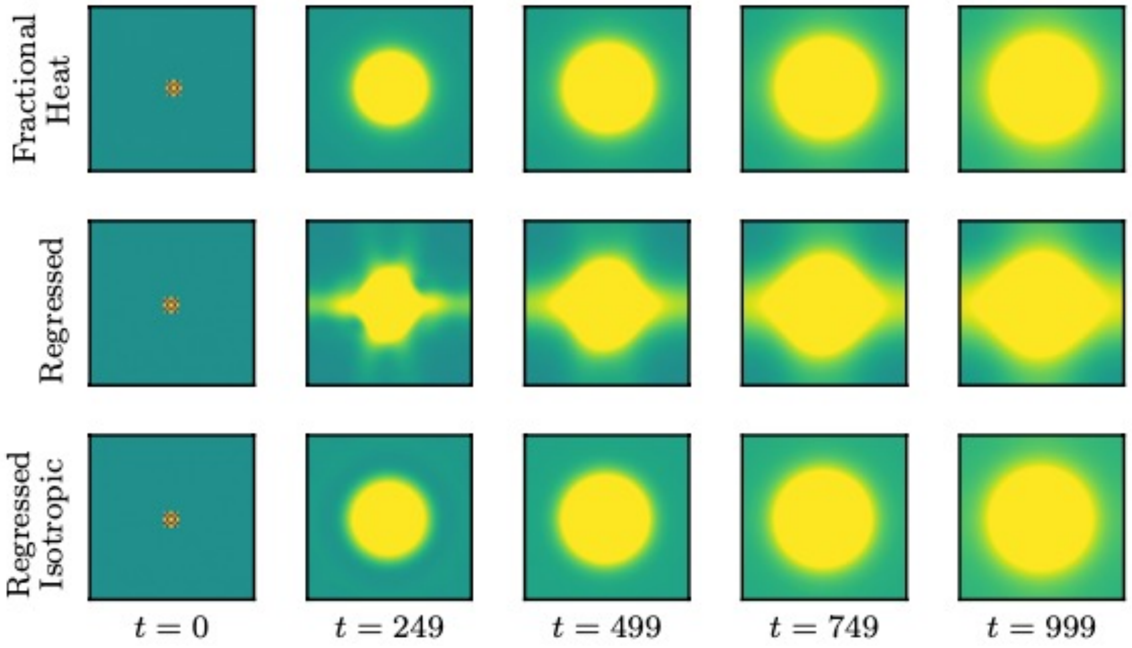
- We can enforce constraints in MOR-Physics by modifying the parameterization,

- Translational equivariance:
  - apply h point-wise, $(h \circ u)(x) = h(u(x))$

- Reflective symmetry:
  - if u solves the PDE, so does $-u$
  - let $h(u) = \mathrm{sign}(u)\tilde{h}(|u|)$

- Rotational equivariance:
- let $g(\kappa) = \tilde{g}(||\kappa||_2^2)$

- Global conservation:
- let $g(\kappa) = \tilde{g}(\kappa)(1 - \delta_{\kappa,0})$



Preserving rotational equivariance leads to better extrapolation

[1]R.G. Patel and Desjardins, *arxiv: 1810.08552* (2018)
[2]R.G. Patel et al., *CMAME* (2021)

APPROXIMATE BAYESIAN METHODS

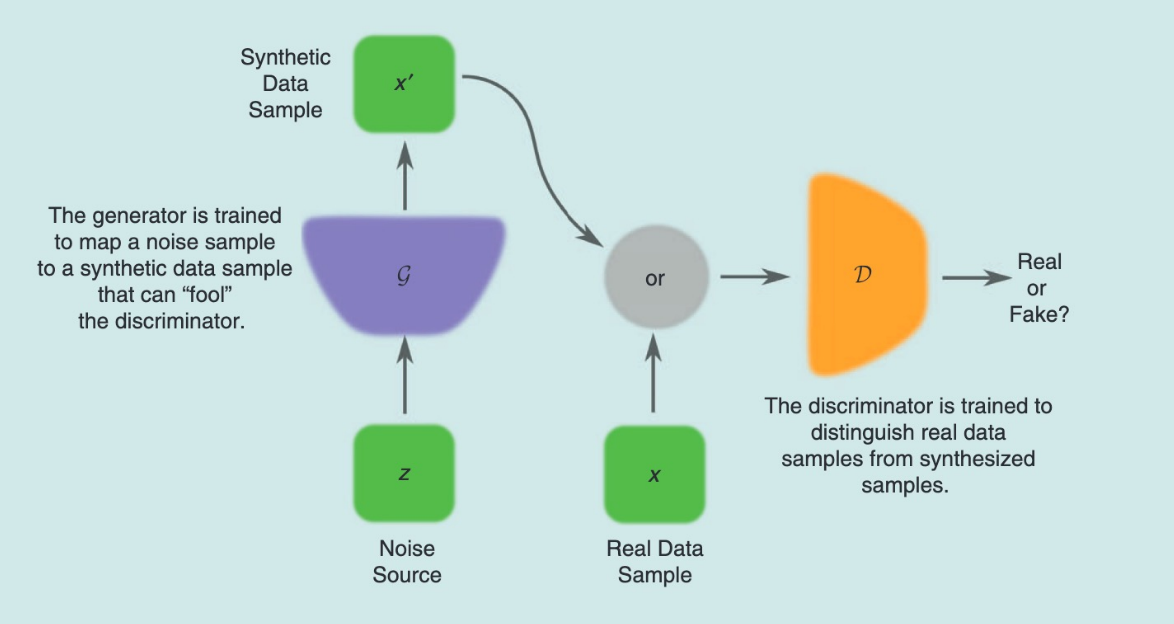
- Rarely, do we have conjugacy in Bayesian inference
  - E.g., the nonlinearity in neural networks
- But if we have

$$p(\xi|x_d, y_d) \propto p(y_d|x_d, \xi)p(\xi)$$

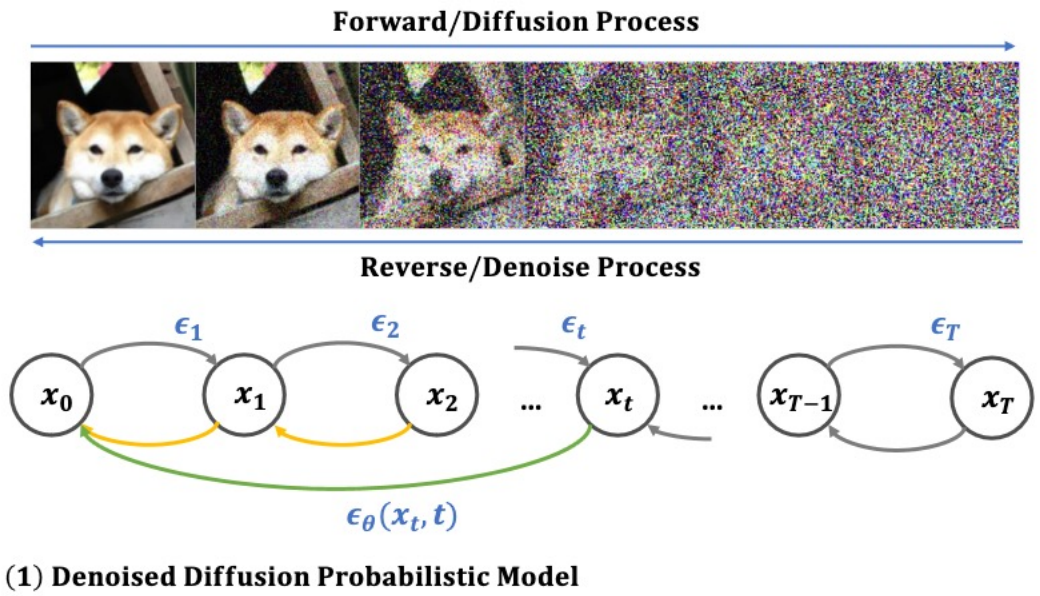
- We can still approximately get the posterior
  - Markov Chain Monte Carlo (MCMC)
    - Evolve a Markov Chain with stationary distribution converging to the posterior
    - S. Sahu, http://www.southampton.ac.uk/~sks/utrecht/mcmc.pdf (2000)
  - Variational inference (VI)
    - Introduce a parameterized (variational) distribution. Minimize the divergence between it and the posterior
    - D.M. Blei, *JASA* (2017)
- Both methods are available in tensorflow_probability

GENERATIVE MODELING

- Often, we just want samples of a distribution and don't want to model posteriors
  - E.g., $p(x), p(y|x)$



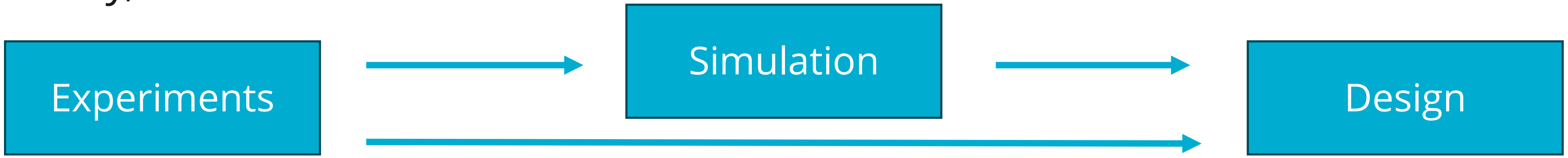
Generative adversarial networks
A. Creswell et al., *IEEE Signal Process. Mag.* (2018)



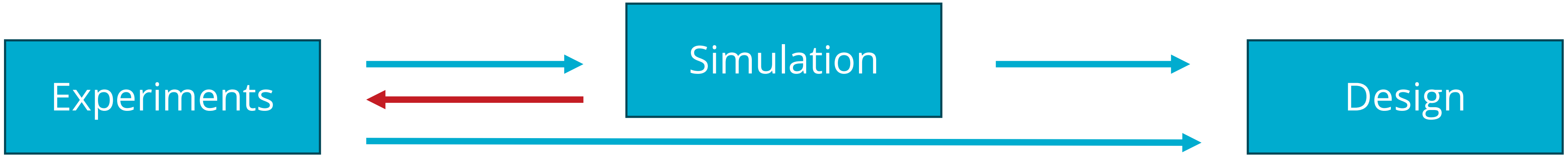
Diffusion model
H. Cao et al., *axXiv:2209.02646* (2018)

OPTIMAL EXPERIMENTAL DESIGN

- Traditionally,



- Optimal experimental design (OED),

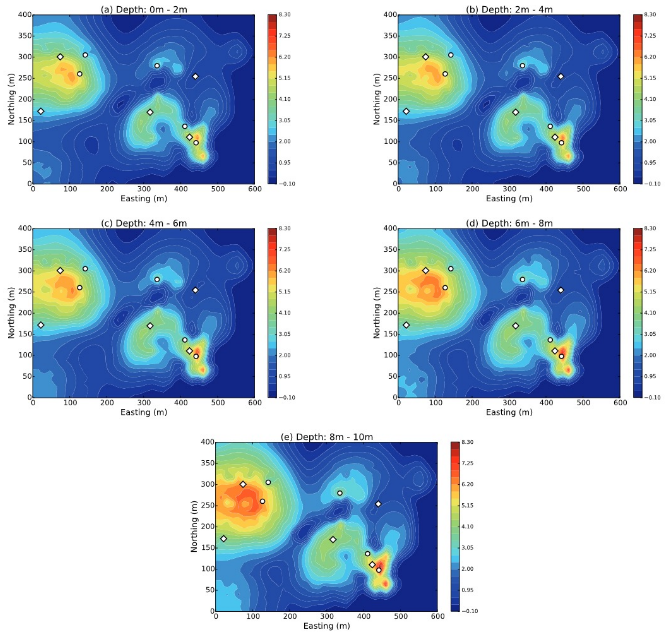


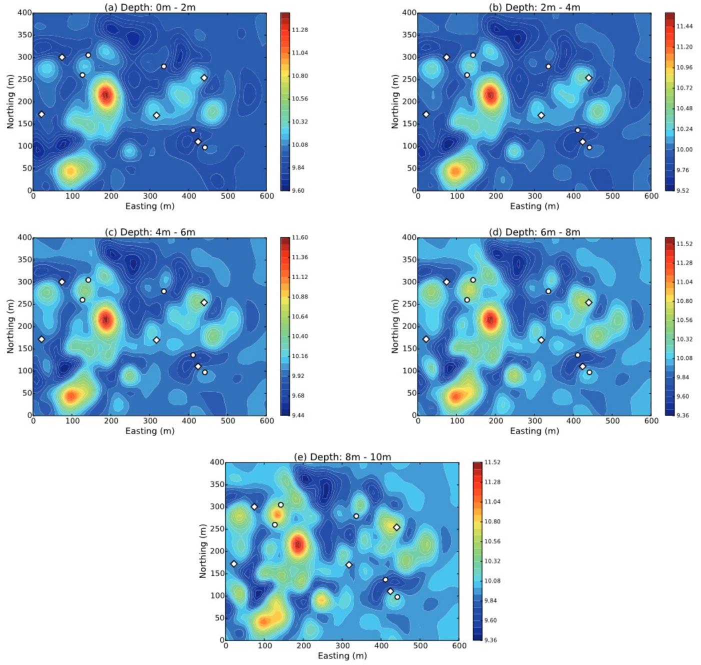
  - Given the data we have, where should the next experiment be performed?

OPTIMAL EXPERIMENTAL DESIGN

- We can optimize the next experiment against a metric
  - E.g., Expected information gain

$$\mathrm{EIG}(d) \triangleq \mathbb{E}_{p(y|d)}\left[H[p(\theta)] - H[p(\theta|y,d)]\right]$$

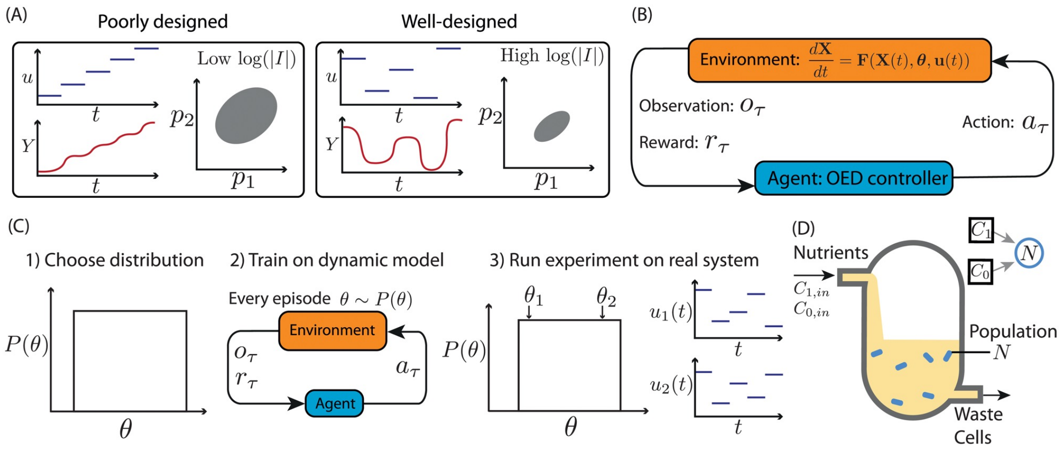




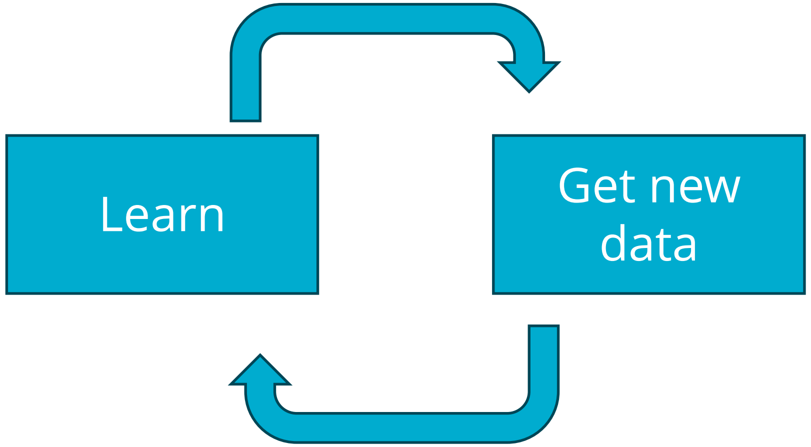
OED for permeability in contaminated soils
P. Tsilifis, *SIAM UQ* (2017)

ACTIVE LEARNING

- Active learning is an iterative process
- Often agent based
  - Reinforcement learning
    - Introduce actions and rewards for actions
    - Introduce an agent that maximizes the reward w.r.t. actions

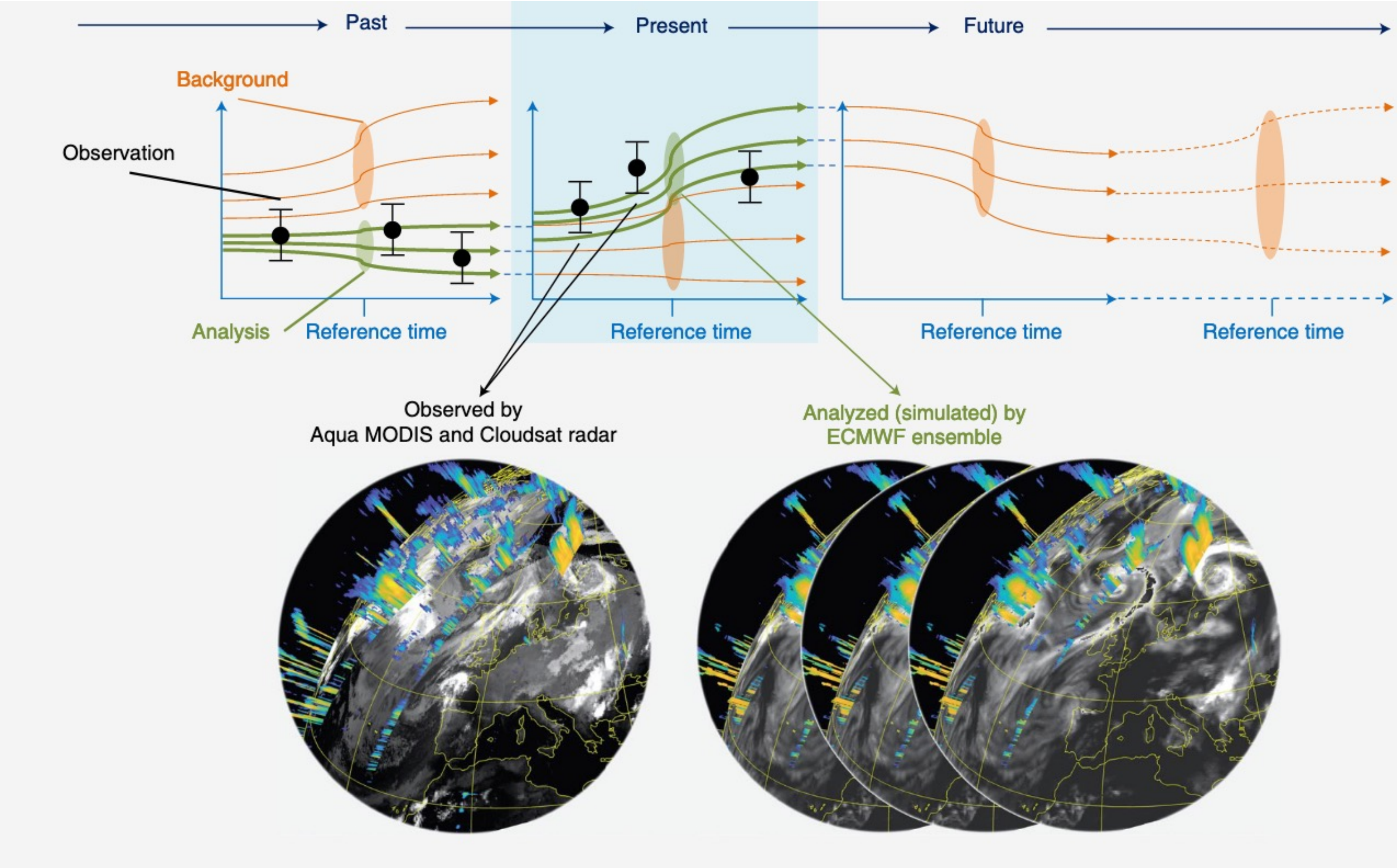


Reinforcement learning for OED of bacteria growth in chemostat
N.J. Treloar, *PLoS Comput. Biol.* (2022)



DIGITAL TWINS

- Combining active learning from multiple sources, we can construct digital representations of systems,



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

REPRODUCIBILITY IN SCIML

- Workflows for sharing code
  - Users should be able to run our code and get the same results
  - Control for dependencies and random seeds
- Containers
  - E.g., Docker
  - C. Boettiger, *Oper. Syst. Rev. (ACM)* (2015)
- Package managers
  - E.g., conda
  - J. Shenouda et al., *IEEE Signal Process. Mag.* (2023)

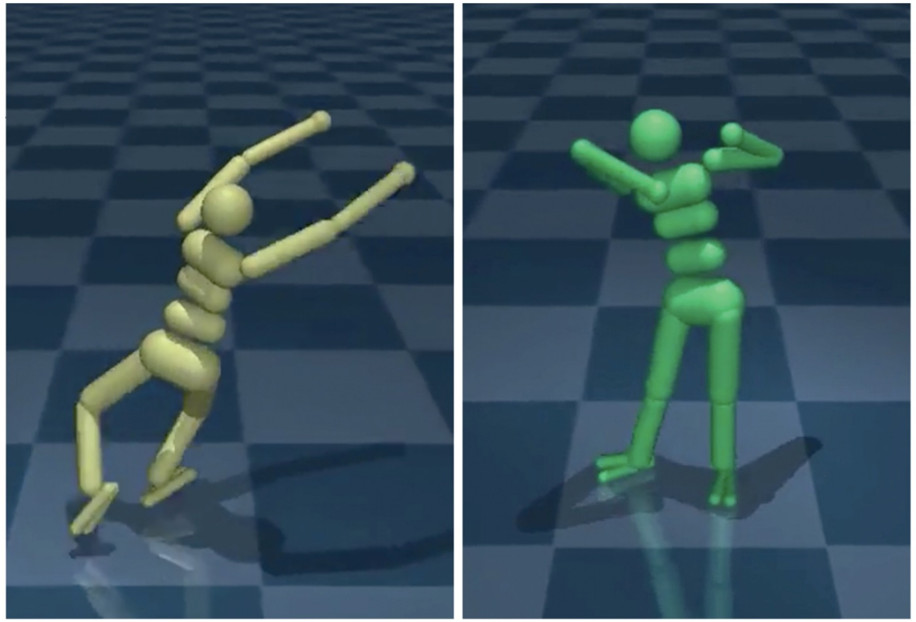
Science   Current Issue   First release papers   Archive   About

Artificial intelligence faces reproducibility crisis

Unpublished code and sensitivity to training conditions make many claims hard to verify.

MATTHEW HUTSON Authors Info & Affiliations

SCIENCE • 16 Feb 2018 • Vol 359, Issue 6377 • pp. 725-726 • DOI: 10.1126/science.359.6377.725



The same algorithm can learn to walk in wildly different ways.

IMAGES: YUVAL TASSA

Last year, computer scientists at the University of Montreal (U of M) in Canada were eager to show off a new speech recognition algorithm, and they wanted to compare it to a benchmark, an algorithm from a well-known scientist. The only problem: The benchmark's source code wasn't published. The researchers had to recreate it from the published description. But they couldn't get their version to match the benchmark's claimed performance, says Nan Rosemary Ke, a Ph.D. student in the U of M lab. "We tried for 2 months and we couldn't get anywhere close."