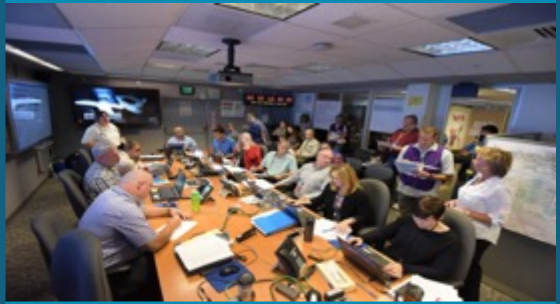
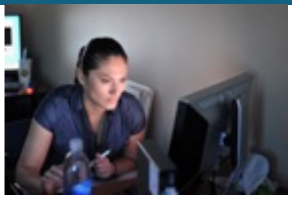
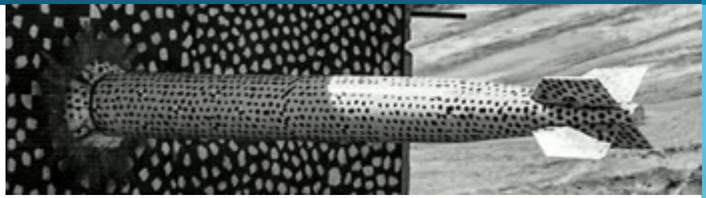
Analysis of Neural Networks as Random Dynamical Systems

Project # 21-0528







PI: Khachik Sargsyan, org. 8351
PM: Janine Bennett, org. 8739

Team: Joshua Hudson (8351),
Marta D'Elia (8734), Habib Najm (8300)

Continuation review: Dec 9, 2021

FY21-24, $500K/yr

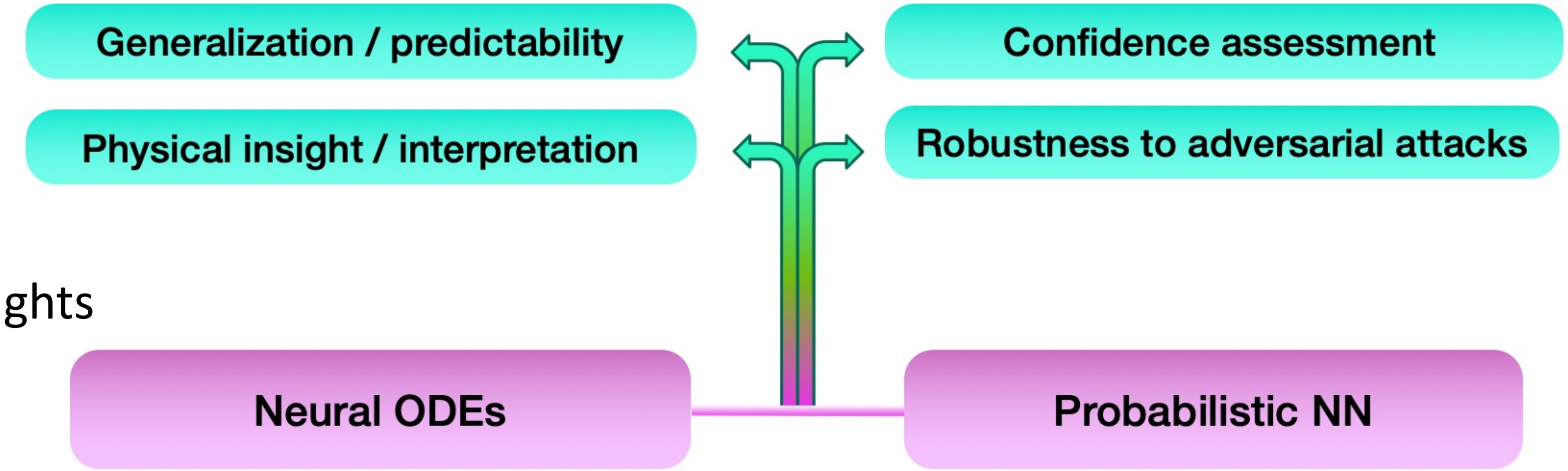
IDEA SUMMARY: Overview

No deviations from the plan

What: Analyze the performance of neural networks (NNs)
[training, generalization, predictive confidence] from dynamical and probabilistic viewpoints.

State of the art: Despite all the success, there are many recognized challenges and unknowns in NN behavior.

Why now:
There is a lot of accumulated knowledge from ODE and UQ;
prime time to build on these insights



Take advantage of ‘continuous’ limit of residual neural networks

Probabilistic viewpoint opens even more opportunities

Encouragement: a few recent papers at the intersection of Bayesian and Neural ODEs

TECHNICAL ACCOMPLISHMENTS: OVERVIEW

		Task (T) / Milestone(M)		Period
FY21	T1.1	Dynamical analysis in deterministic setting†	done	10/2020-06/2021
	T1.2	Development of deterministic, sparse weight representations	done	01/2021-09/2021
	M1	Demonstrate reduced NODE in deterministic setting	done	by 09/2021
	T1.3	Formulation of Bayesian inference of weights	done	04/2021-09/2021

Reduction by removing the fast dynamics via regularizing with stiffness.

FY22	T2.1	Formulation of stability conditions under uncertainty†	in prog.	10/2021-03/2022
	T2.2	Extension of dynamical analysis under uncertainty	in prog.	10/2021-06/2022
	T2.3	Formulation of fractional PNODE construction*	in prog., paper subm.	01/2022-09/2022
	T2.4	Regularization via weight representations in PNODEs	in prog.	01/2022-06/2022
	T2.5	Inference/training of weight functions in PNODEs	in prog.	04/2022-09/2022
	M2	Demonstrate training of regularized PNODEs		by 09/2022

From FY23: Working with E3SM land model data to hone the methodologies. in prog.

TA: We have developed in-house codes for both discrete and continuous cases

Necessary for full control

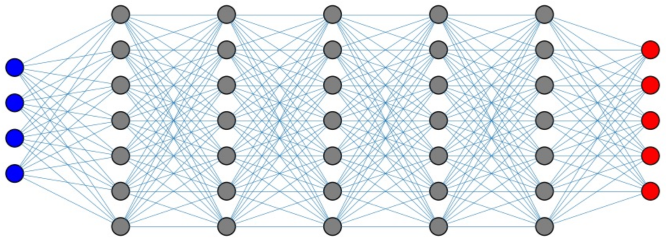
gitlab-ex.sandia.gov/ksargsy/nnrds

ResNet (discrete)

$$
\begin{cases}
\boldsymbol{x}_1 = \boldsymbol{x} + \boldsymbol{\alpha}_0 \boldsymbol{\sigma}(\boldsymbol{W}_0 \boldsymbol{x}_0 + \boldsymbol{b}_0) \\
\vdots \\
\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \boldsymbol{\alpha}_n \boldsymbol{\sigma}(\boldsymbol{W}_n \boldsymbol{x}_n + \boldsymbol{b}_n) \\
\vdots \\
\boldsymbol{y} = \boldsymbol{x}_{L-1} + \boldsymbol{\alpha}_{L-1} \boldsymbol{\sigma}(\boldsymbol{W}_{L-1} \boldsymbol{x}_{L-1} + \boldsymbol{b}_{L-1})
\end{cases}
$$

$\boldsymbol{x} = \boldsymbol{x}_0$

Input



$\boldsymbol{y} = \boldsymbol{x}_T$

Output

Neural ODE (continuous)

$$
\frac{d\boldsymbol{x}}{dt} = \boldsymbol{\sigma}(\boldsymbol{W}(t)\boldsymbol{x} + \boldsymbol{b}(t))
$$

$$
\boldsymbol{x}(0) = \boldsymbol{x} \qquad \boldsymbol{x}(T) = \boldsymbol{y}
$$

Recent Technical Advances

+ stiffness penalization
+ weight parameterization
+ Bayesian (MCMC, VI)
+ integral NODEs
+ app. to E3SM land model

TA: Stiffness penalty regularizes the training and improves generalization

(Had to) Define/quantify stiffness for ResNets

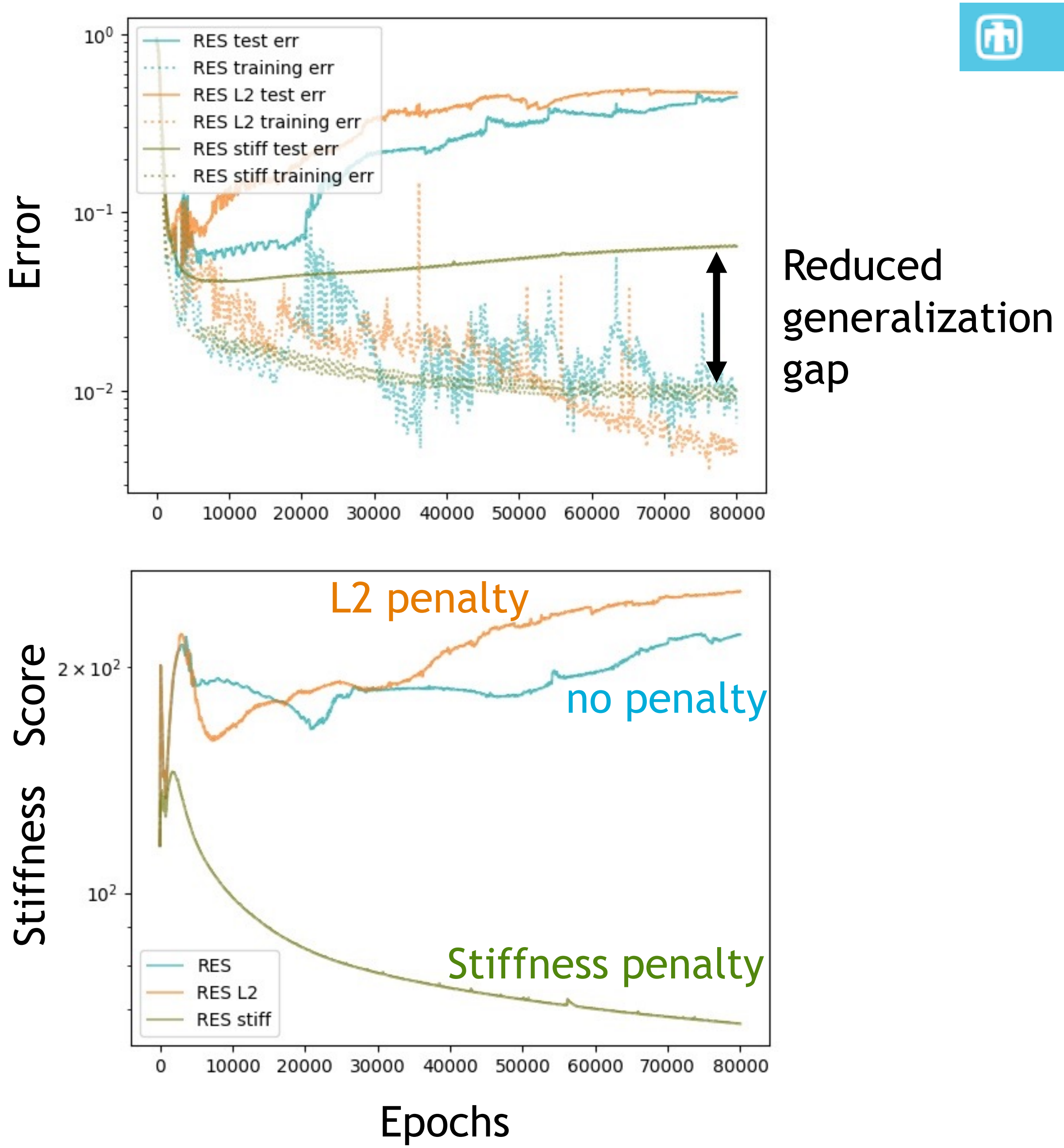
- ResNet layer n rate of change: linearize and use conventional e-value ratio definition:

$$s_{n,i} = r\left(\sigma'\left(W_n x_n^i + b_n\right) W_n\right) = |\lambda_1| / |\lambda_d|$$

- Approximate: $|\lambda_1| \approx \mathrm{Gelfand}(A) = \sqrt[k]{\|A^k\|}$
- Replace $|\lambda_d|$ with active time scale:

$$\tau_{n,i} = \left|\sigma\left(W_n x_n^i + b_n\right)\right| / \left|x_n^i\right|$$

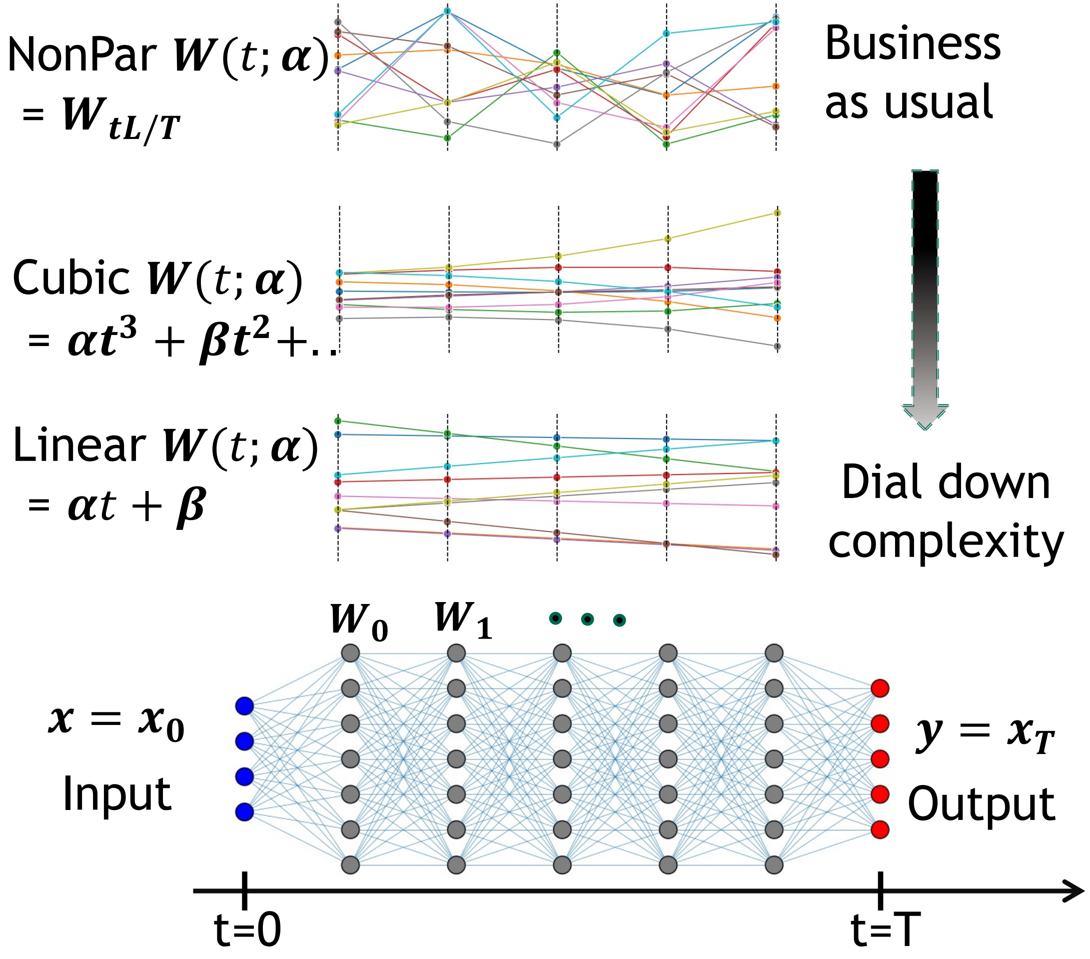
- Stiffness score is then defined as a sum over samples and layers: $\sum_i \sum_n s_{n,i}$



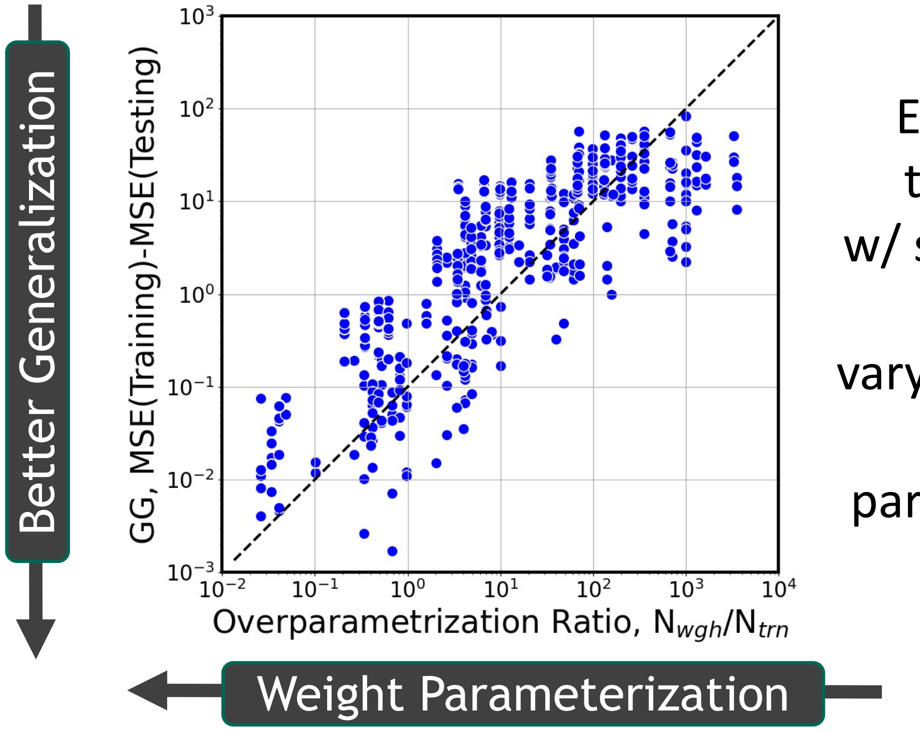
TA: Weight parameterization improves generalization

ResNet: $\boldsymbol{x_{n+1}} = \boldsymbol{x_n} + \boldsymbol{\alpha_n \sigma}(\boldsymbol{W_n x_n} + \boldsymbol{b_n})$

Instead of training for weight matrices $\boldsymbol{W_0}, \boldsymbol{W_1}, \ldots$ parameterize $\boldsymbol{W}(t;\boldsymbol{\alpha})$ and train for $\boldsymbol{\alpha}$'s.



- **Generalization Gap (GG) strongly correlates with overparameterization**
- **Weight-parameterized ResNets reduce GG**



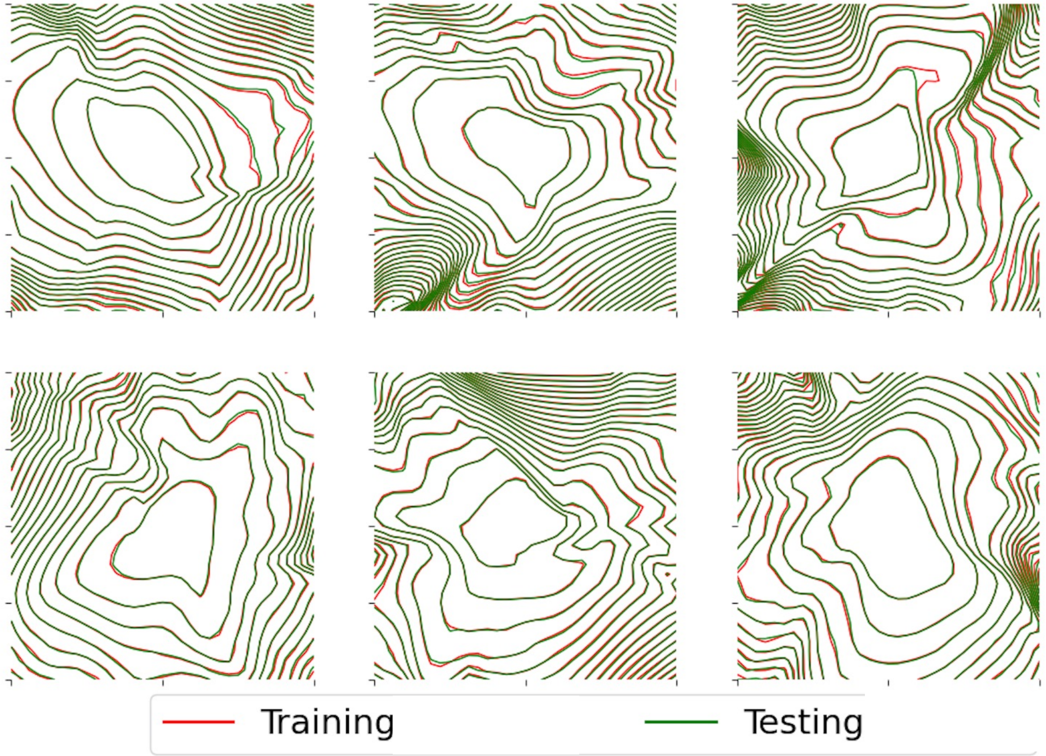
Each dot is a training run w/ synthetic data and varying options of weight parameterization

TA: With weight parameterization, ResNets regularize loss landscape compared to MLPs

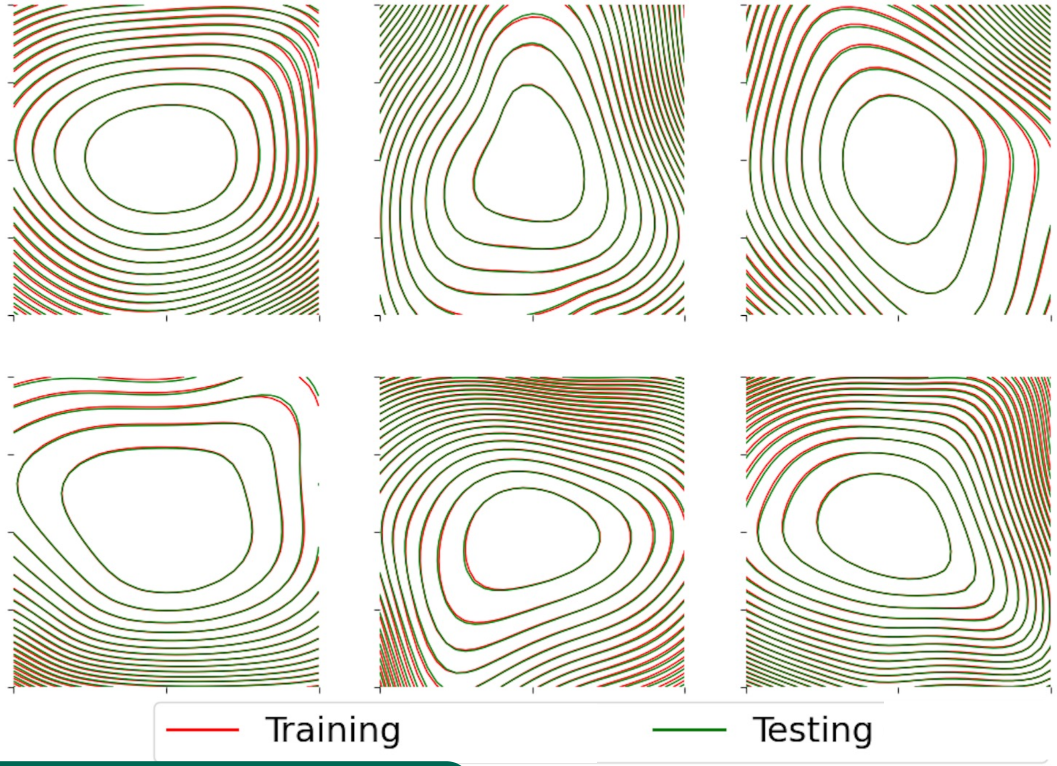
MLP NN: $x_{n+1} = \sigma(W_n x_n + b_n)$

ResNet: $x_{n+1} = x_n + \alpha_n \sigma(W_n x_n + b_n)$

Multilayer Perceptron (learning the layer)



ResNets (learning the layer diff.)



This is with E3SM model data – and this feature helps both the training and the accuracy of model surrogate

TA: We hone the methods on climate model data

E3SM Vegetation Dynamics data:

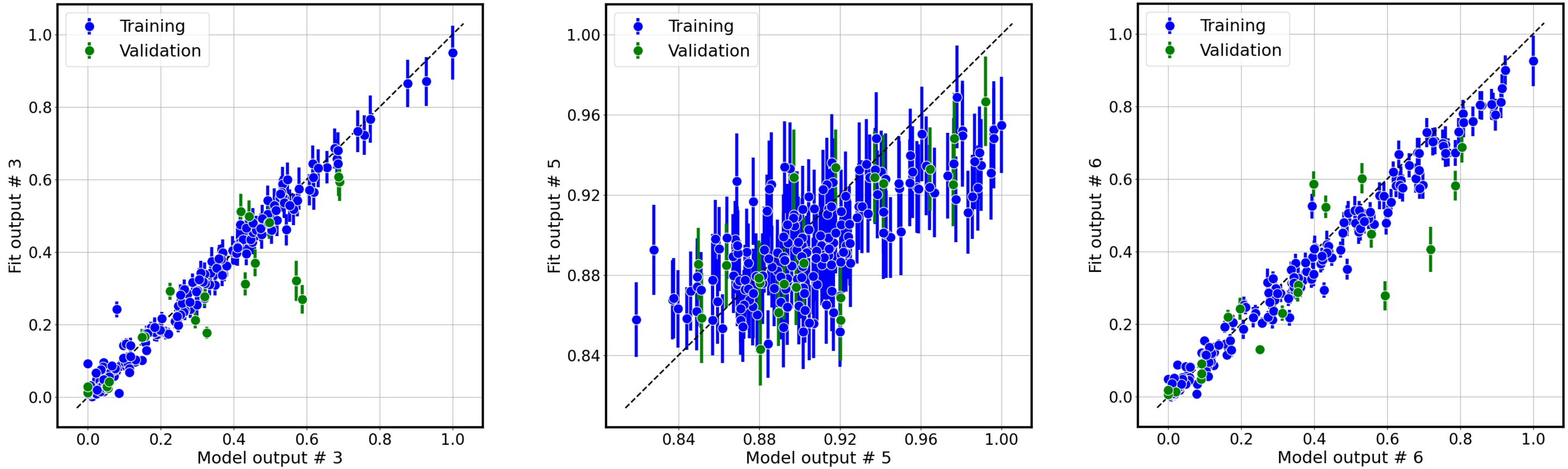
- 15 input parameters
- 10 static output QoIs
- 2000 training simulations

Augmenting NN predictions with uncertainties is important, and has not been done in E3SM

E3SM
Energy Exascale
Earth System Model

Code development: Bayesian wrapper for any PyTorch module, including weight-parameterized ResNets

Three Different E3SM Outputs vs their Resnet Approximation with Uncertainties



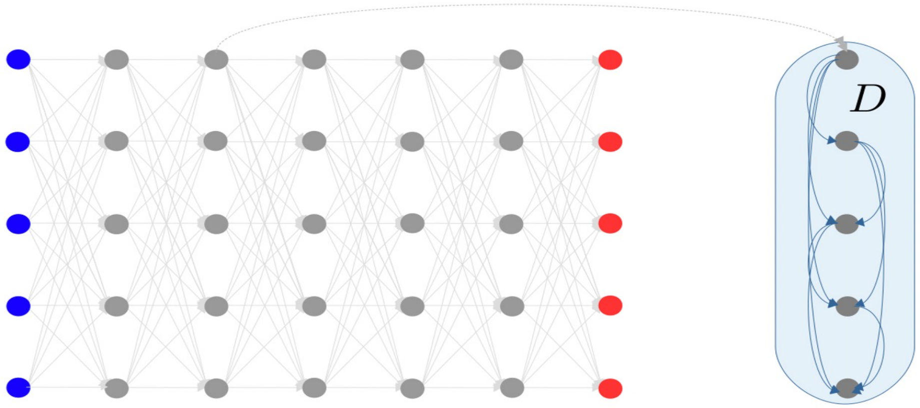
TA: Integral NODEs for PDE learning outperforms conventional operator learning methods

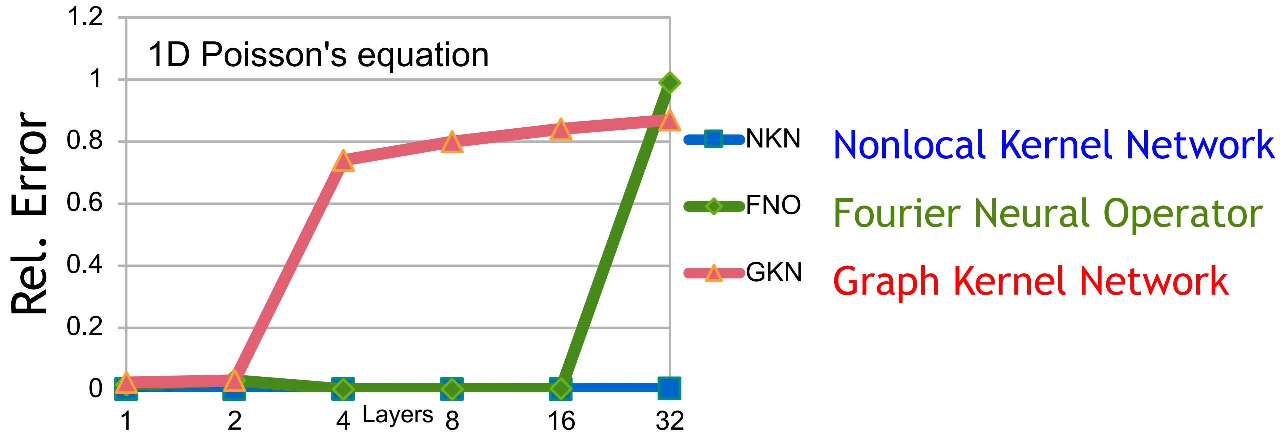
$$\frac{d\mathbf{h}}{dt} = \sigma(W_t \mathbf{h}_t + \beta_t)$$

NODE

$$\frac{\partial \mathbf{h}(\mathbf{x},t)}{\partial t} = \int_D k_{NN}(\mathbf{x},\mathbf{y};\mathbf{v})(\mathbf{h}(\mathbf{y},t) - \mathbf{h}(\mathbf{x},t))d\mathbf{y} - R_{NN}(\mathbf{x})\mathbf{h}(\mathbf{x},t)$$

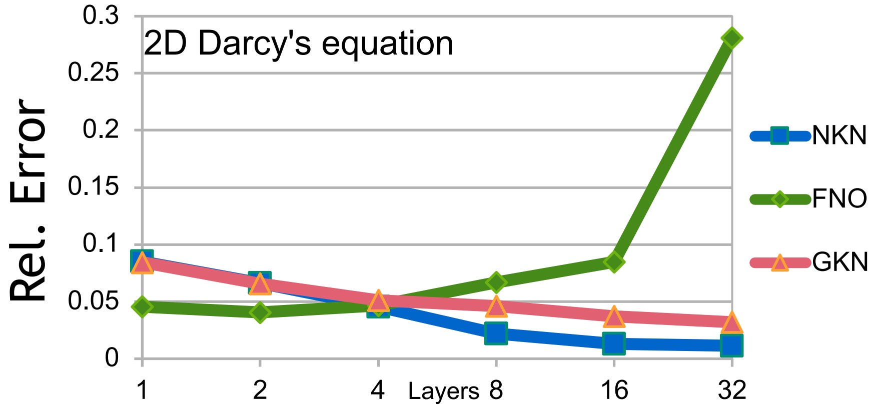
Nonlocal Kernel Network





*You, Yu, **D'Elia**, Gao, Silling, Nonlocal Kernel Network (NKN): A Stable and Resolution-Independent Deep Neural Network. To be Submitted.*

- Resolution-independent
- Stable operator learning
- **NKN beats leading competitors FNO and GKN**



PROJECT STATUS: No adjustments needed

- The Y1 technical milestone is met with a slight re-interpretation.

- Hired Joshua Hudson in November 2020 via CSRI postdoc position announcement.
- Offer pending for the second postdoc, expected start date June 2022.
- The team has increased the effort to advance the tasks while searching for a second postdoc.

- Subcontract pending with prof. Lars Ruthotto (Emory U), who is a leader in the field and has available student(s). They will primarily help with the analysis of weight parametrization.

Programmatic / Spend plan: One postdoc amount behind in spending. Mitigated by the commitment to U Emory partner.

PROJECT PLAN FOR REMAINDER OF FY22

No technical risks, only staffing

- Dynamical analysis/stiffness: *[Joshua Hudson]*
  - produce publishable results on stiffness-regularized training

- Weight parameterization: *[Emory U]*
  - demonstrate regularization with weight parameterization
  - incorporate sparsity, and/or complex parameterization
  - loss surface and NN model capacity analysis

- Probabilistic NODEs: *[Ushnish Sengupta, postdoc to start in June]*
  - demonstrate probabilistic reformulation – e.g. Bayesian NN made feasible by simplified parameterization
  - publishable results with E3SM model ensembles

ANTICIPATED OUTPUTS AND TIE TO INVESTMENT AREA CALL

How did this project contribute to the CIS IA strategic goals and objectives?

- CIS Trusted AI RC Thrusts: at least 2/3 directly relevant: Math Foundations and Usability/Trust
- Mission relevance through impact on DOE-wide climate modeling efforts
- Collab. with Emory U (Lars Ruthotto) will enhance our capabilities and lay ground for external proposals
- Targeting high-impact, peer-reviewed venues such as SIAM journals and ML conferences

What are the key results from this research that will be useful to other current and future projects?

- Evident practical benefits of ResNets/NODEs in terms of training efficiency, accuracy and generalization
- Augmenting NN predictions with uncertainty is sought after both in climate (ongoing BER work) and chemistry (ongoing BES and FES work)
- Connections to these efforts exist and the team is working with application data to hone the methods
- Early results are already having positive impact on E3SM research

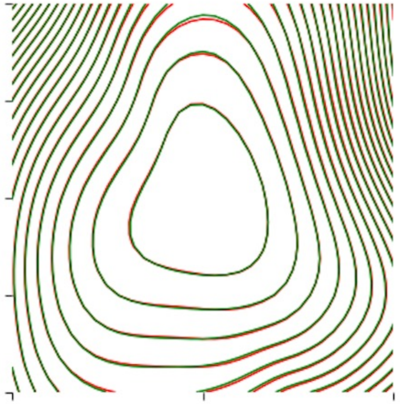
Technology insertion and follow-on funding for potential and realized ROI

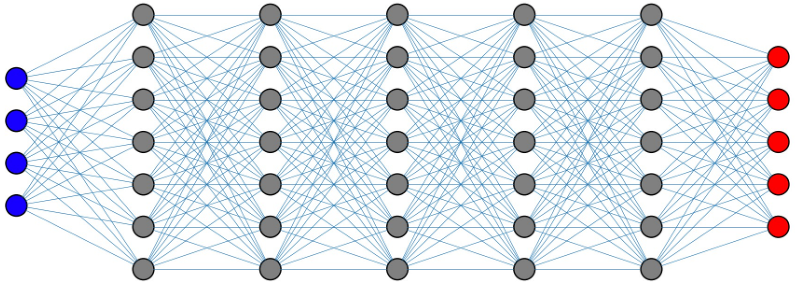
- Fundamental mathematical work, with both code development and impact on real applications
- Strong potential to help Sandia mark the territory in ML/UQ landscape
- Follow-on DOE Early Career proposal by Marta D'Elia

Analysis of Neural Networks as Random Dynamical Systems
PI: Khachik Sargsyan (8351) PM: Janine Bennett (8739)

Project Goal(s)

- Develop methods for analysis and regularization of neural networks
- Merge probabilistic and ODE viewpoints to improve NN training and accuracy
- Demonstrate on exemplar applications: climate, material science





Mission Impact

- Bring together theory, modeling, computation, and data, under potentially noisy and adversarial conditions
- Improved NN performance can be key to many mission apps
- Probabilistic NODEs will be a unique capability and will remain mission-relevant for years to come

Transition Plan

- Unique and risky capability: if successful, can lead to further ASCR funding
- Follow-on funding (BER, FES) for applications of interest highly likely
- Software not a direct target but a bi-product which will serve us well for future funding

FY22 Technical Milestones

- Extend dynamical analysis under uncertainty
- Regularize via sparse, probabilistic weight functions
- M2: Demo regularized NODE in probabilistic setting