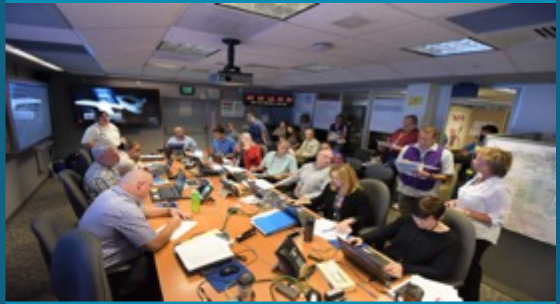
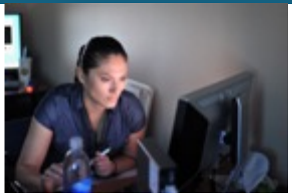
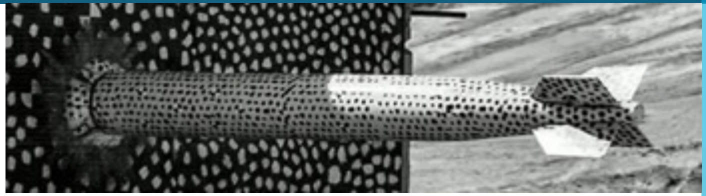
Analysis of Neural Networks as Random Dynamical Systems

Project # 21-0528







PI: Khachik Sargsyan, org. 8351
PM: Janine Bennett, org. 8730

Team: Joshua Hudson (8351),
Oscar Diaz-Ibarra (1446), Habib Najm (8300)
Lars Ruthotto, Haley Rosso (Emory U)

Continuation review: Dec 8, 2022

FY21-24, $525K/yr

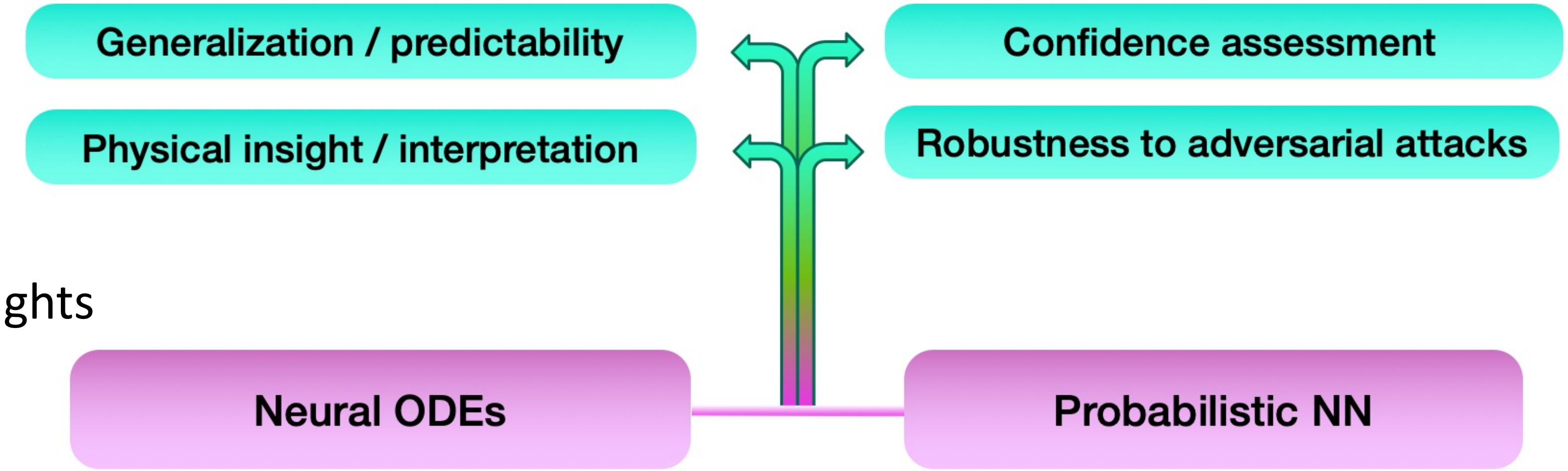
IDEA SUMMARY: Overview

No major 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

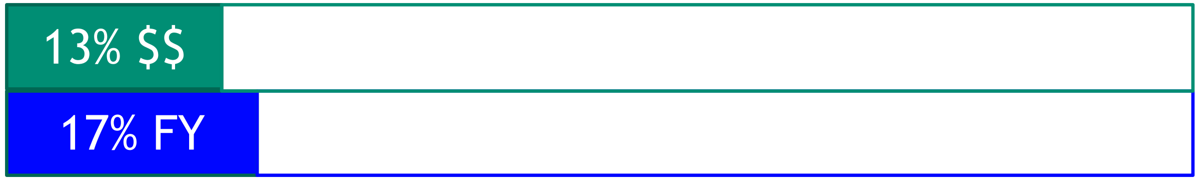
PROJECT STATUS: No adjustments needed

- Y1 (deterministic) milestone is met,
- Y2 (probabilistic) milestone is met (* with discrete ResNets),
- Y3 (application impact) milestone in progress.

- Hired Joshua Hudson in November 2020 via CSRI postdoc announcement.
- Marta D'Elia left SNL in June 2022.
- Involved Oscar Diaz-Ibarra part time, who transitioned to 1446 staff.
- Second postdoc: still in search, in conjunction with other projects.

- FY22: Subcontract with Prof. Lars Ruthotto (Emory U), who is a leader in the field.
  Graduate student Haley Rosso summer visit,
  presentation at SIAM MDS 2022, works on paper.

Programmatic / Spend plan:



TECHNICAL ACCOMPLISHMENTS: Overview

M1 (09/2021): Reduced NODE in deterministic setting done

Reduction by removing the fast dynamics via regularizing with stiffness — *Paper submitted to JMLMC, Paper submitted to RAMSES Proc.;*

Integral NODEs — *Paper published in JCP*

M2 (09/2022): Training of regularized probabilistic ResNets/NODEs done (for ResNets)

ResNet regularization via weight parameterization — *Collaboration with Emory U, paper in prog.*

Weight parameterization (on E3SM data) — *Paper in prog.*

Probabilistic augmentation for the continuous NODE — *Current work.*

M3 (09/2023): Quantifiable improvement on exemplar applications in progress

QUiNN: software for quantifying uncertainties in NNs — *About to be posted in github.com/sandialabs*

Climate modeling, E3SM Land Model — *BER-funded work*

Materials science — *FES-funded work*

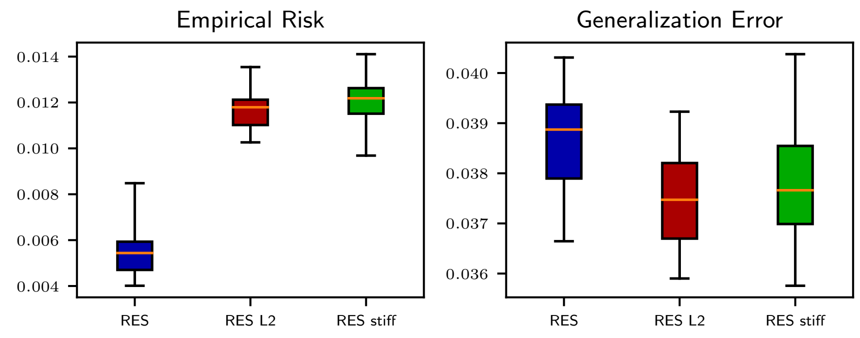
ASC interest, data from Mitch Wood, Aidan Thompson

Feedback from the last review

- **Publications:**

…nice to see the team has one publication in review and another planned journal submission in June as well as a conference presentation.

- 1 published,
- **2 submitted**,
- 2 in prep.



- **Exemplar problems:**

…happy to see the team's focus on some DOE-relevant exemplar problems: climate land modeling, materials science, and catalytic chemistry. This should help with the transition of this research to follow-on customer projects at the end of FY23.

- Climate work impacted
- Materials science datasets currently studied (both FES-DOE and ASC-SNL)

- **Software:**

During this next review, we would also like to hear about the software and UQ tools that they have created within this project.

- Quantification of Uncertainties in Neural Networks (QUiNN)
- To be published under github.com/sandialabs/quinn/

Foundational capabilities impacting multiple applications

Predictive capability of Neural Networks (NNs) hinges on **generalization** (ability to predict well outside training data).

Regularization of NNs as a tool to achieve generalization.

- ✓ Stiffness Penalization (reported prev.)
- ✓ Weight Parameterization
- ✓ Probabilistic Weights (+software)

Methods

- ✓ Climate Land Modeling
- ✓ Materials Science

Applications

ResNet and NODE in a regression setting (supervised ML)

ResNet (discrete)

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

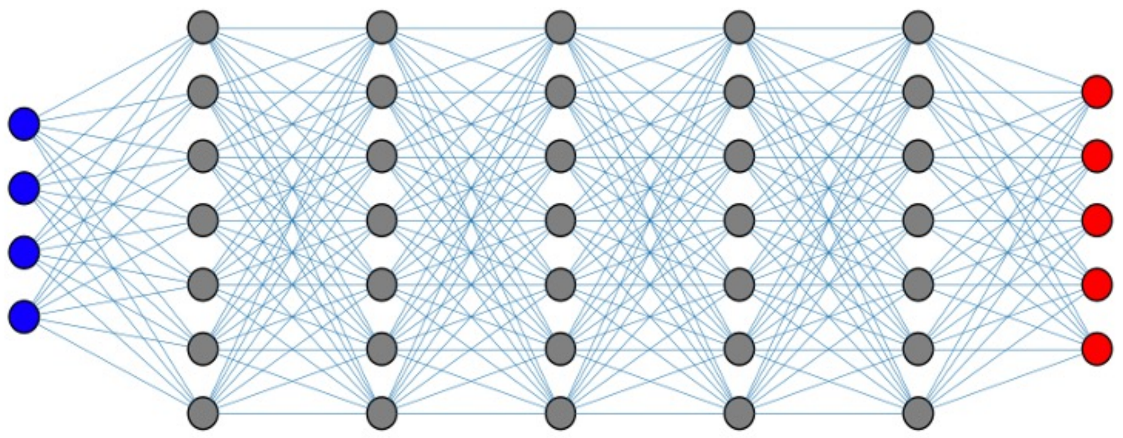
Neural ODE (continuous)

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

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

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

Input



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

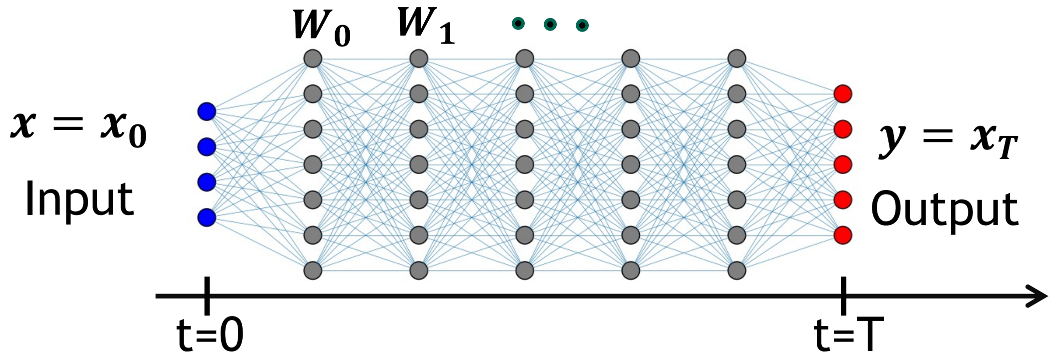
Output

gitlab-ex.sandia.gov/ksargsy/nnrds

In-house codes for both discrete and continuous cases

Weight parameterization setup

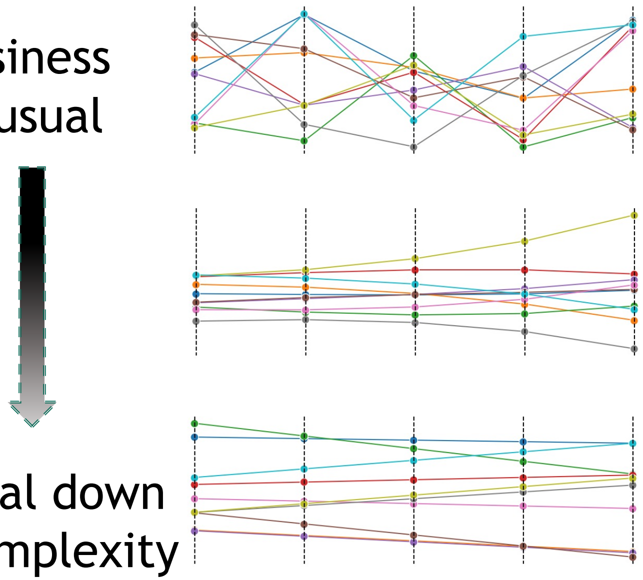
ResNet: $x_{n+1} = x_n + \alpha_n \sigma(W_n x_n + b_n)$



Training for weight matrices $W_0, W_1, \ldots$

Heavily overparameterized, does not generalize well

Business as usual



NonPar $W(t;\alpha)$ $= W_{tL/T}$

Parameterize $W(t;\alpha)$ and train for α's.

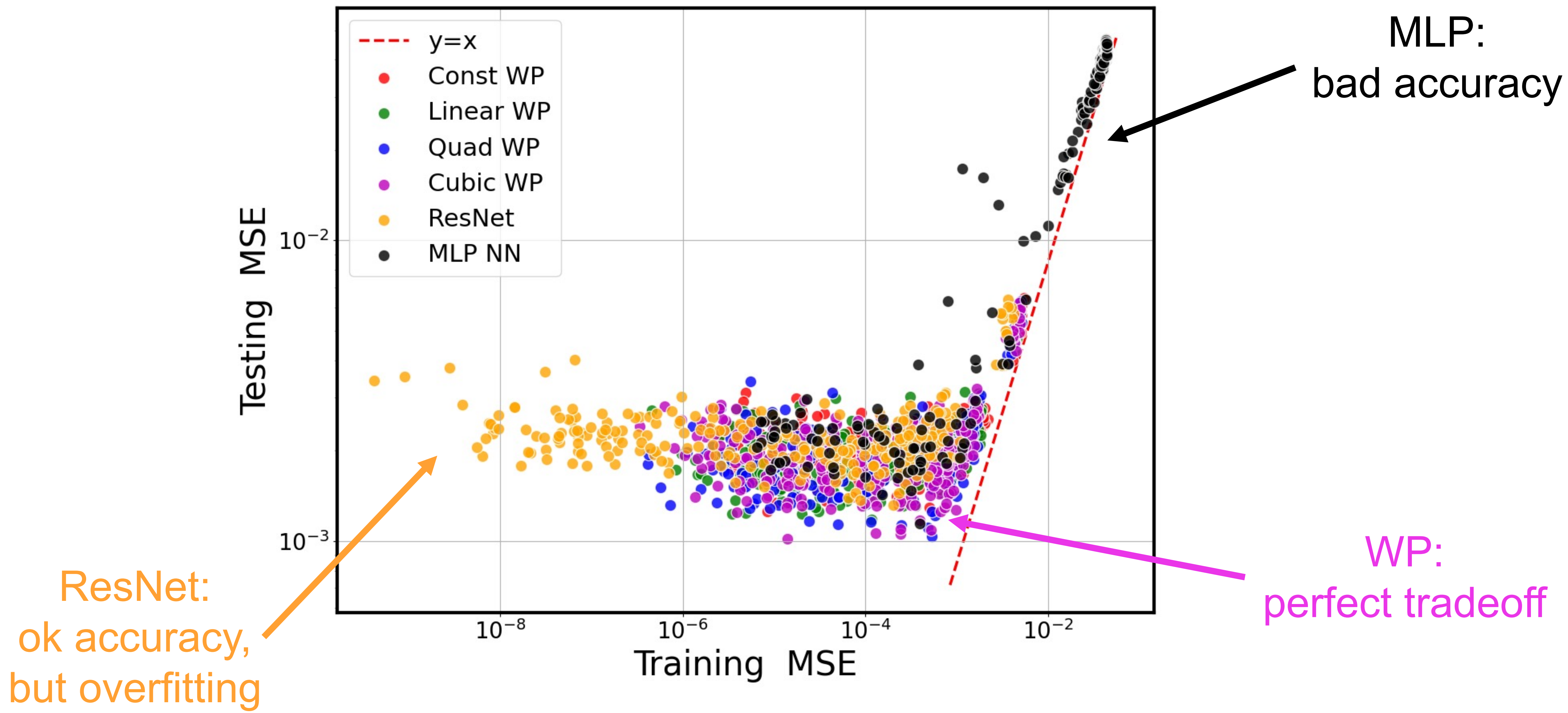
Cubic $W(t;\alpha)$ $= \alpha t^3 + \beta t^2 + ..$

Parameterization of weight functions reduces capacity and improves generalization

Linear $W(t;\alpha)$ $= \alpha t + \beta$

Dial down complexity

Weight parameterization improves generalization



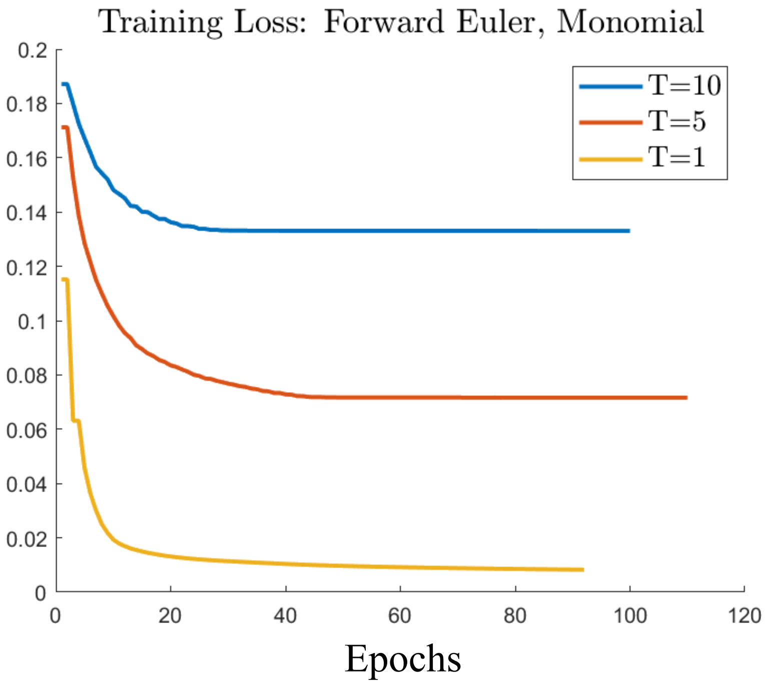
Orthogonal-basis weight parameterization

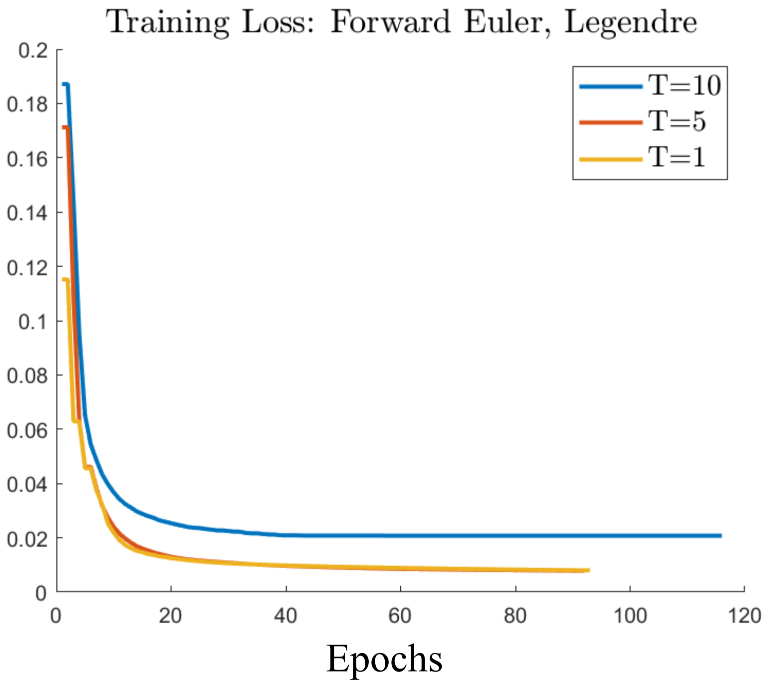
Prof. Lars Ruthotto
Ph.D. student Hailey Rosso
EMORY UNIVERSITY

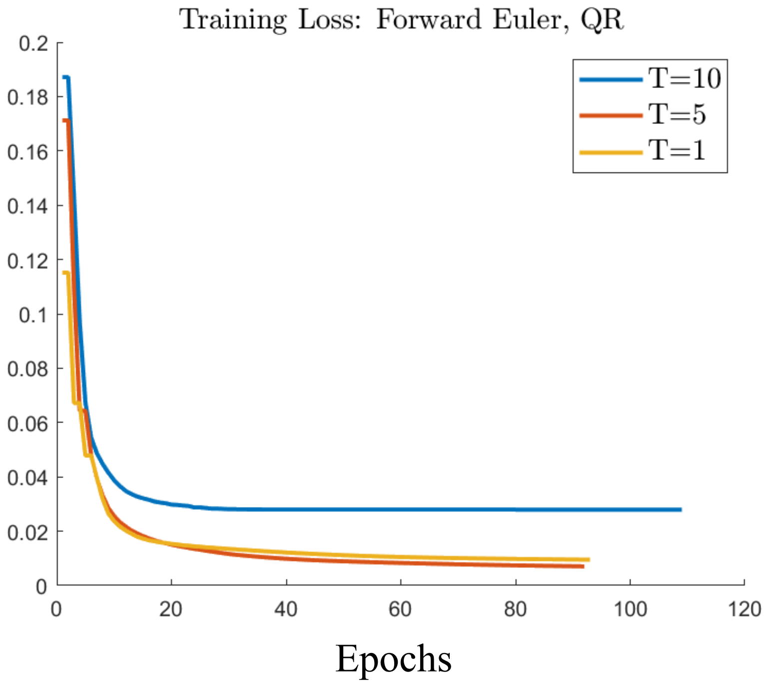
- Basis expansions as weight parameterization
- Training deteriorates as time (*i.e.* NN depth) increases
- Orthogonal basis for NN weight parameterization helps

$$W(t) = \sum_{k=0}^{K} w_k \, \varphi_k \, (t)$$

- Work was presented at SIAM MDS 2022 meeting
- Journal paper in prep.

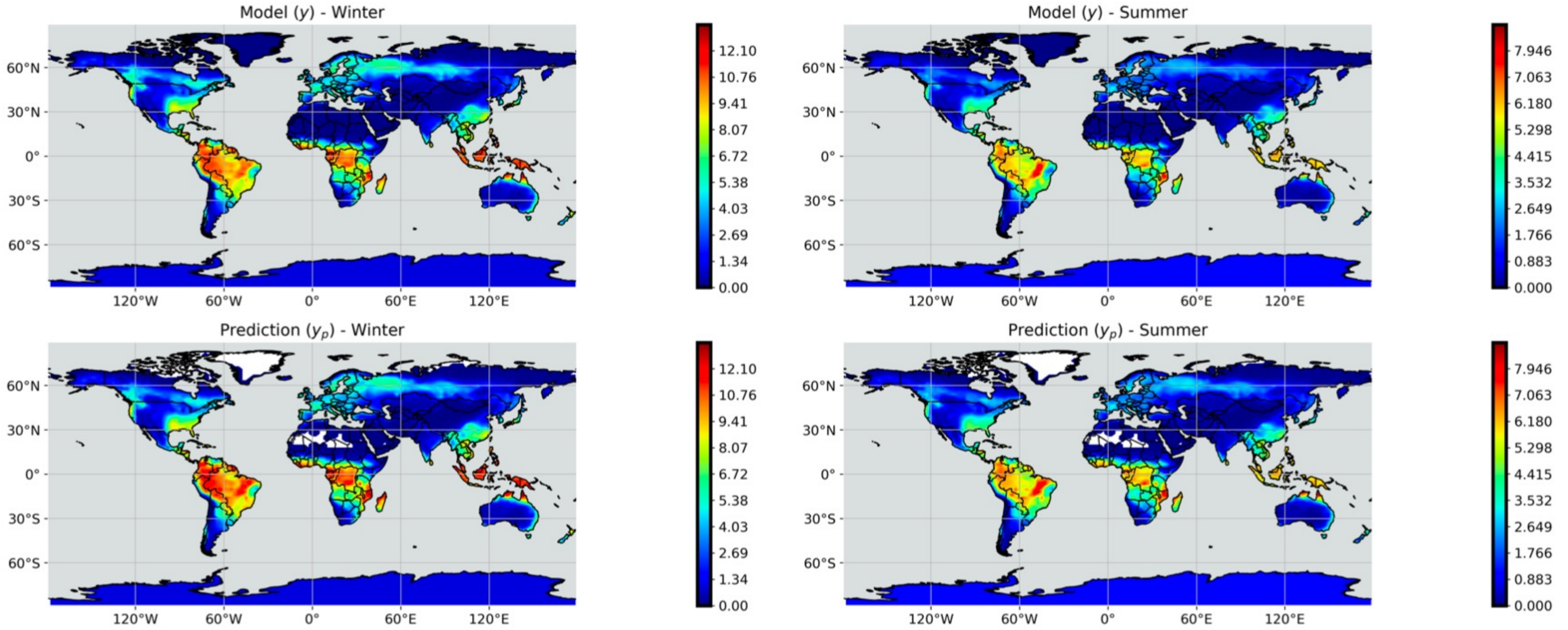






E3SM Land Model (ELM) surrogate

Weight-parameterized ResNets allowed construction of accurate spatio-temporal surrogate models for ELM



QUiNN: probabilistic wrapper for NNs

To be published under
github.com/sandialabs/quinn/

- Conventional NN: training for *deterministic* weight matrices $\boldsymbol{W_0}, \boldsymbol{W_1}, \ldots$
- *Probabilistic* approach: training for probability distributions $\boldsymbol{p(W_0)}, \boldsymbol{p(W_1)}, \ldots$
- Three classes of options:

Full Bayesian → Approximate Bayesian → Ensemble methods

Full Bayesian
- ➢ Markov chain Monte Carlo (MCMC)
- Typically infeasible for overparameterized NNs

Approximate Bayesian
- ➢ Variational methods
- Typically underestimates extrapolative predictions

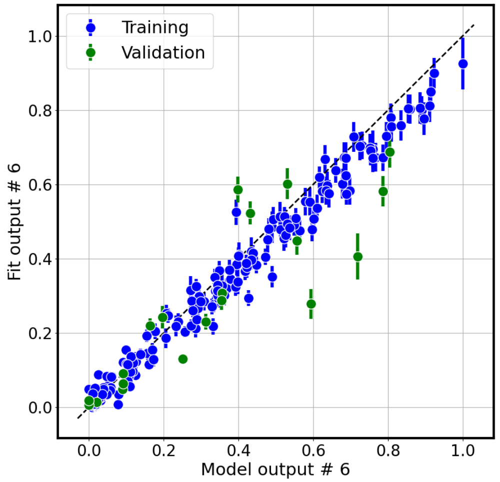
Ensemble methods
- ➢ Heuristic, but works
- Many recent papers viewing deep ensembles as Bayesian approximation

Weight-parameterized ResNets allow these methods to scale much better than before

Multiple **applications** are informing the development of foundational research and benefiting from it

Climate modeling

- E3SM land model
- 15 input parameters
- 10 static output QoIs
- 2K training simulations

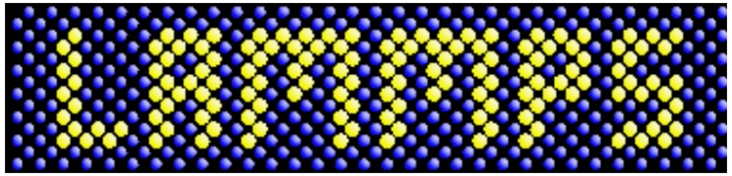


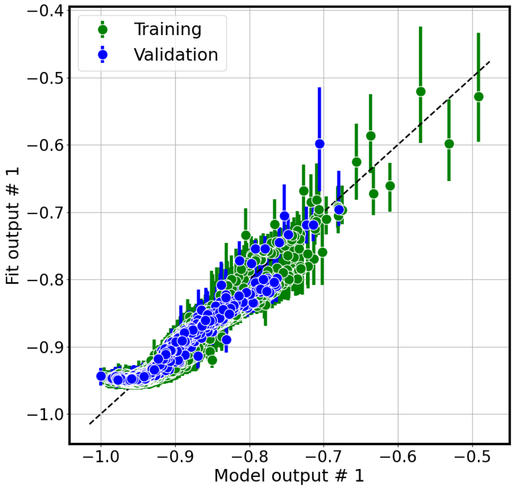
- Current: E3SM, SciDAC efforts require ELM surrogate
- Future: Promising avenues after AI4ESP, BER funding

None of these applications have been previously exposed to NN prediction uncertainties, particularly in the context of ResNets and weight parameterization

Materials science

- FitSNAP Entropy dataset
- 30 input bases
- 1 output (Energy/Force/Stress)
- 20K training DFT simulations





- Current: ASCR/FES project; and ASC data
- Future: ASC funding

PROJECT PLAN FOR THE REMAINDER OF FY

- Probabilistic regularization: *[Joshua Hudson]*
  - Implement flavors of Bayesian ResNet/NODE in QUiNN
  - Demo regularization and training improvement; plan a paper
- Weight parameterization (WP): *[Emory U, Oscar Diaz-Ibarra]*
  - Finalize the papers demonstrating WP
  - Merge with prev.: Bayesian NN made feasible by WP
- Transition/Outreach: *[Khachik Sargsyan, Habib Najm]*
  - Demo impact to climate modeling with E3SM (potential BER funding)
  - Demo results for Material Science (ASC-SNL interest)
  - SAND Report

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 efforts (BER, FES, BES-funded applications)
- Collaboration with Emory U (Lars Ruthotto): potential for external proposals

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 material science (ongoing FES work, and potential ASC work)
- Connections to these efforts exist and the team is working with application data to hone the methods

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

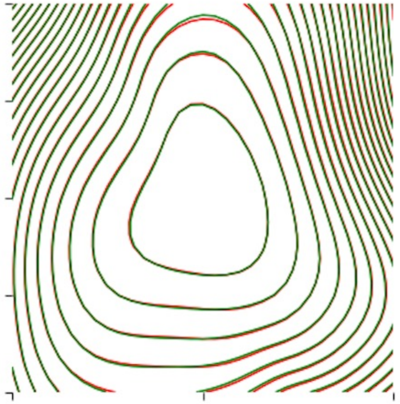
- Fundamental mathematical work, with both code development and impact on real applications
- Strong potential to help Sandia establish leadership in the ML/UQ landscape
- Weight-parameterized ResNets as well as probabilistic NN wrapper software can serve as a base for future funding, both for ASCR (theory) and for application-specific proposals DOE-wide (BER, FES, BES).

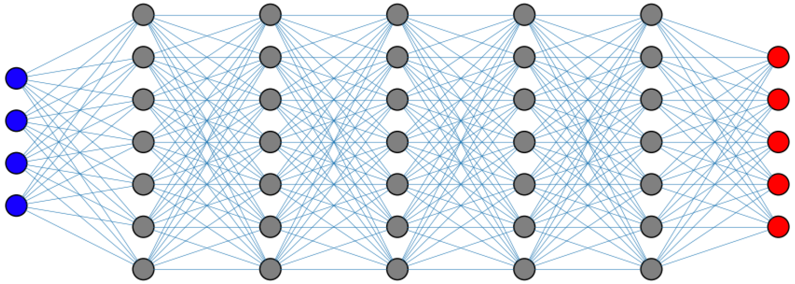
Analysis of Neural Networks as Random Dynamical Systems

PI: Khachik Sargsyan (8351) PM: Janine Bennett (8730)

Project Goal(s)

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





FY23 Technical Milestones

- Extend regularization analysis under uncertainty
- Regularize via sparse, probabilistic weight functions
- M3: Tangible impact on application exemplars

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 or ASC funding
- Follow-on funding (BER, FES, BES) for applications of interest highly likely
- QUiNN software is a bi-product which will serve us well for future funding.

Extra Materials


Conventional NN (MLP)
ResNets
Weight-parameterized (WP) ResNets
Optimal WP
(Apprx.) Bayesian

QUiNN Quantifying Uncertainties in NN

Deterministic → **Probabilistic**

torch.nn.module → **wrapper(torch.nn.module)**

Usage: →

```
uqnet = MCMC_NN(nnet)
```

```
class MCMC_NN(QUiNNBase):
    def __init__(self, nnmodule, verbose=True):
        super(MCMC_NN, self).__init__(nnmodule)
        self.verbose = verbose
```

Option 1: MCMC

```
uqnet = VI_NN(nnet)
```

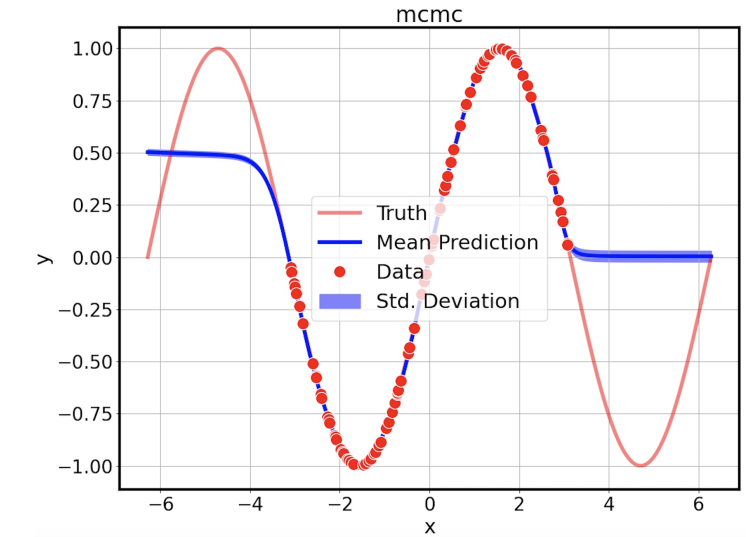
```
class VI_NN(QUiNNBase):
    def __init__(self, nnmodule, verbose=False):
        super(VI_NN, self).__init__(nnmodule)
        self.bmodel = BNet(nnmodule)
        self.verbose = verbose
```

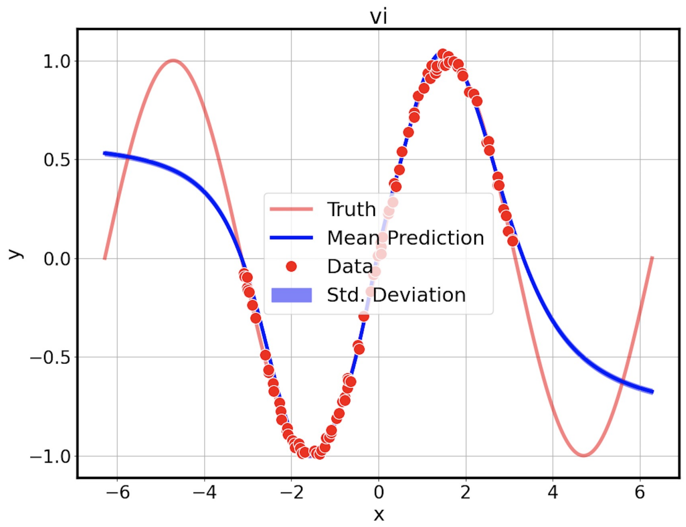
Option 2: Variational Inference

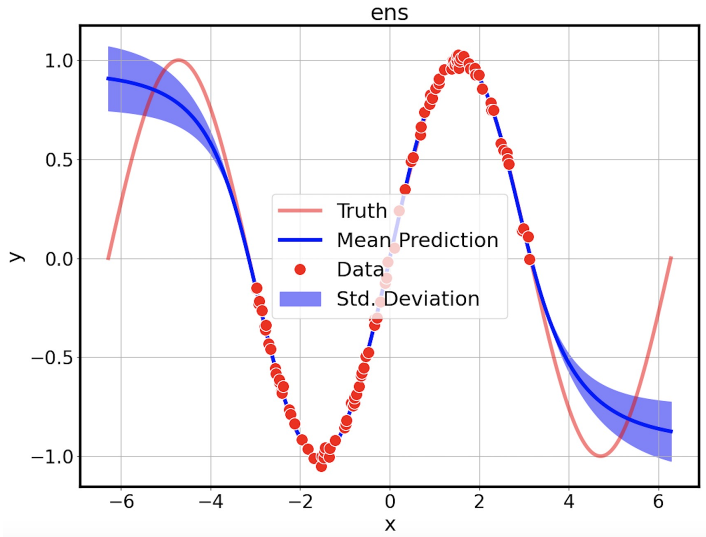
```
uqnet = Ens_NN(nnet, nens=nmc)
```

```
class Ens_NN(QUiNNBase):
    def __init__(self, nnmodule, nens=1, verbose=False):
        super(Ens_NN, self).__init__(nnmodule)
        self.verbose = verbose
        self.nens = nens
```

Option 3: Ensembling







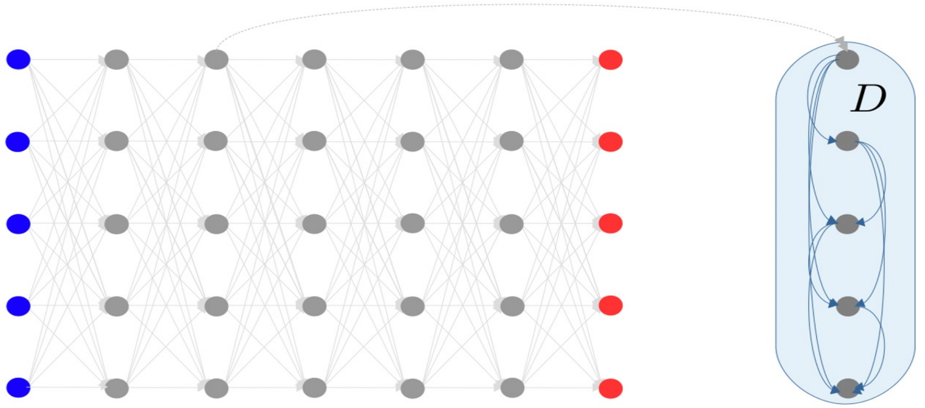
TA: Integral NODEs for PDE learning outperforms conventional operator learning methods

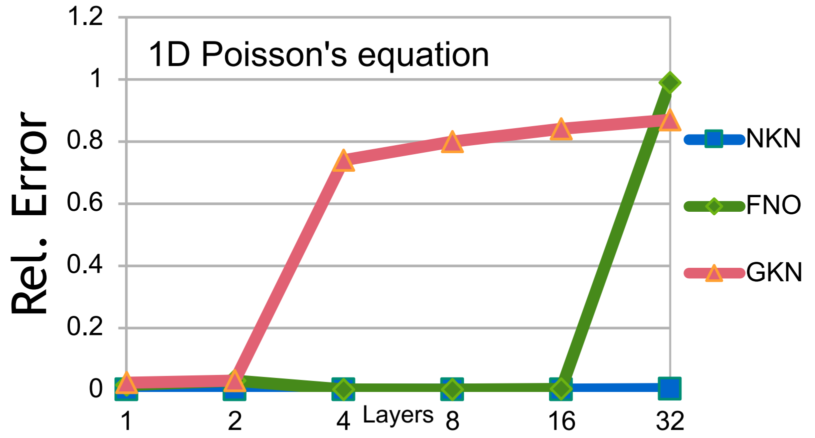
$$\frac{d\mathbf{h}}{dt} = \sigma(W_t \mathbf{h}_t + \beta_t)$$

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





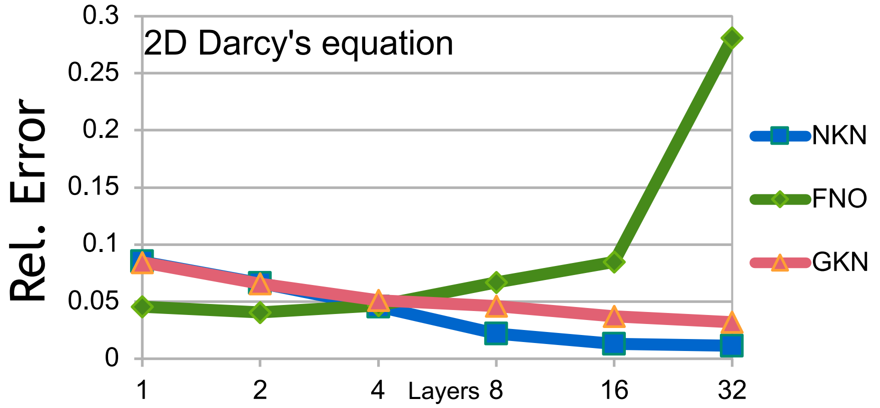
Nonlocal Kernel Network

Fourier Neural Operator

Graph Kernel Network

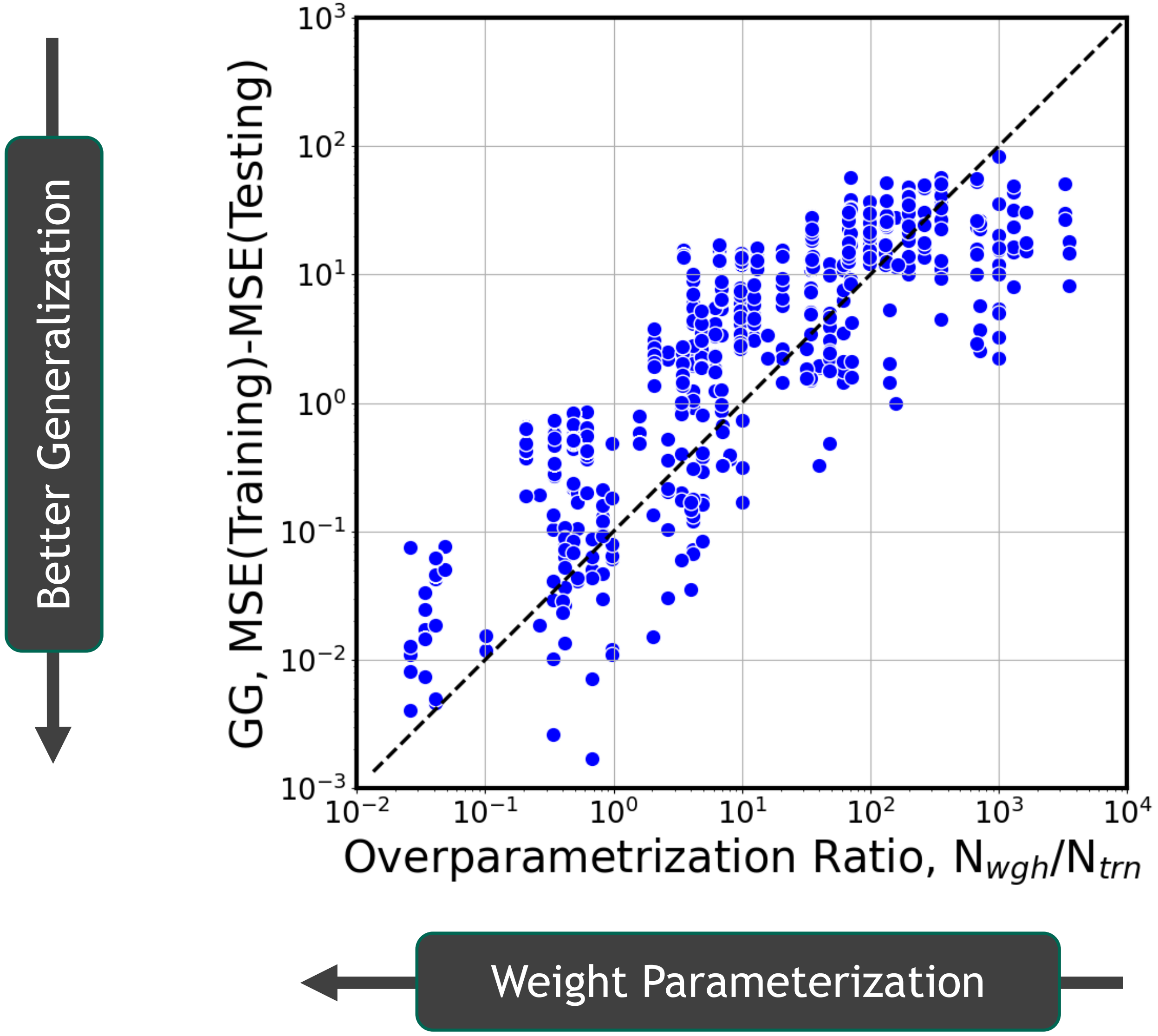
*You, Yu, **D'Elia**, Gao, Silling, Nonlocal Kernel Network (NKN): A Stable and Resolution-Independent Deep Neural Network. To be Submitted.*

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



TA: Weight parameterization improves generalization

OSCAR's work



- Generalization Gap correlates with overparameterization

- Weight-parameterized ResNets reduce Generalization Gap

Each dot is a training run with varying weight parameterization functions

		Task (T) / Milestone(M)	Period
FY21	T1.1	Dynamical analysis in deterministic setting†	10/2020-06/2021
	T1.2	Development of deterministic, sparse weight representations	01/2021-09/2021
	M1	Demonstrate reduced NODE in deterministic setting	by 09/2021
	T1.3	Formulation of Bayesian inference of weights	04/2021-09/2021
FY22	T2.1	Formulation of stability conditions under uncertainty†	10/2021-03/2022
	T2.2	Extension of dynamical analysis under uncertainty	10/2021-06/2022
	T2.3	Formulation of fractional PNODE construction*	01/2022-09/2022
	T2.4	Regularization via weight representations in PNODEs	01/2022-06/2022
	T2.5	Inference/training of weight functions in PNODEs	04/2022-09/2022
	M2	Demonstrate training of regularized PNODEs	by 09/2022
FY23	T3.1	Explore the scaling and performance of PNODEs	10/2022-03/2023
	T3.2	Construction of NNs as PNODE discretization with desired features	10/2022-09/2023
	M3	Demonstrate quantifiable improvement on exemplar applications	by 09/2023
	T3.3	Final SAND report	07/2023-09/2023

CIS LDRD Key Performance Indicators and Metrics

Enhance technology base at Sandia: New, differentiated, utilized internal tools and capabilities

R&D technologies (math, software, hardware, knowledge, skills, techniques, design, etc.) deployed resulting in mission achievement, cost savings and/or **follow-on funding**

Journal and conference publications, conference presentations, invited talks, citations, journal impact factor

Career growth: Professional awards/recognition, best papers

Intellectual Property: Technical advances, patents, commercial license, copyrights, **open source software assertions**, government use notices, royalties

Capabilities and knowledge picked up by MF IA's or sponsors to advance TRL

External collaborations: Academia, industry, other Labs and gov. agencies

Software reproducibility and reuse across Sandia