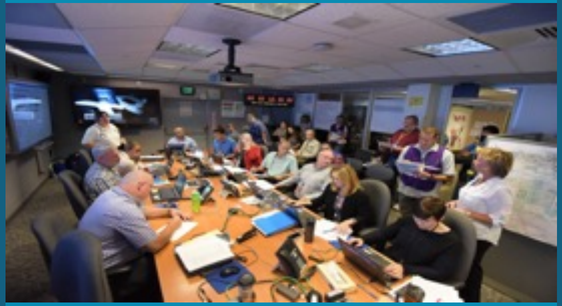
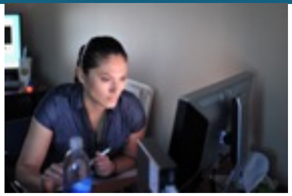
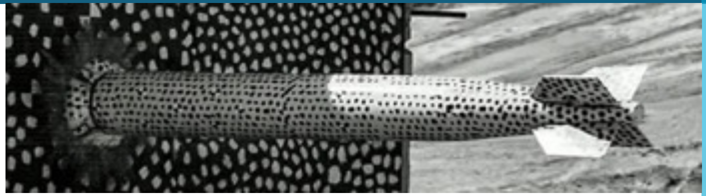
Analysis of Neural Networks as Random Dynamical Systems

Project # 21-0528







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

Team: Joshua Hudson (8351), Marta D'Elia (now Pasteur Labs)
Oscar Diaz-Ibarra (1446), Habib Najm (8300)
Lars Ruthotto, Haley Rosso (Emory U)

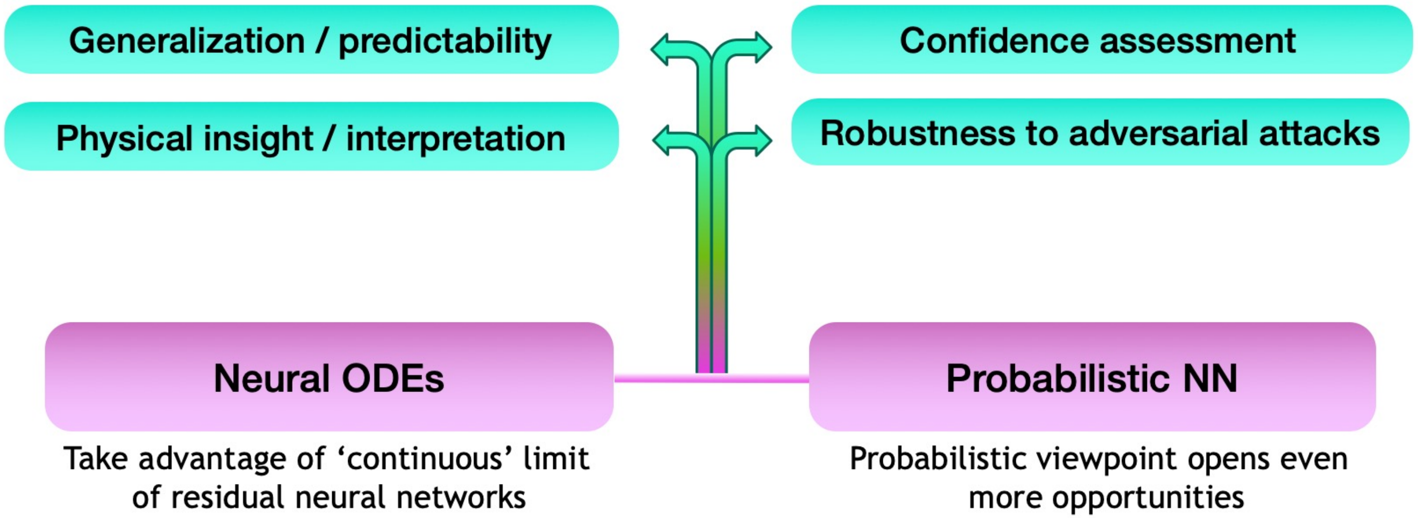
Ending project review, August 29, 2023

FY21-24, $525K/yr

PURPOSE, GOALS AND APPROACH

Despite all the success, there was (and still is) many recognized challenges and unknowns in the behavior of neural networks (NNs) in SciML context.

- Can we improve training methods of NNs?
- Can we achieve better generalization of NNs?
- Can we augment NN predictions with uncertainties?



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

Goal:
Analyze and improve the NN performance *[training, generalization, predictive confidence]* relying on dynamical and probabilistic viewpoints

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
- ✓ Weight Parameterization
- ✓ Probabilistic Weights

QUiNN library

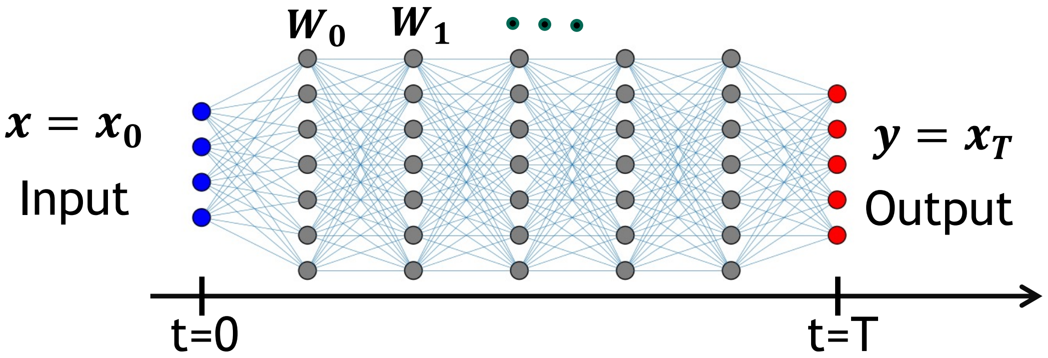
Methods

- ✓ Climate Land Modeling
- ✓ Materials Science

Applications

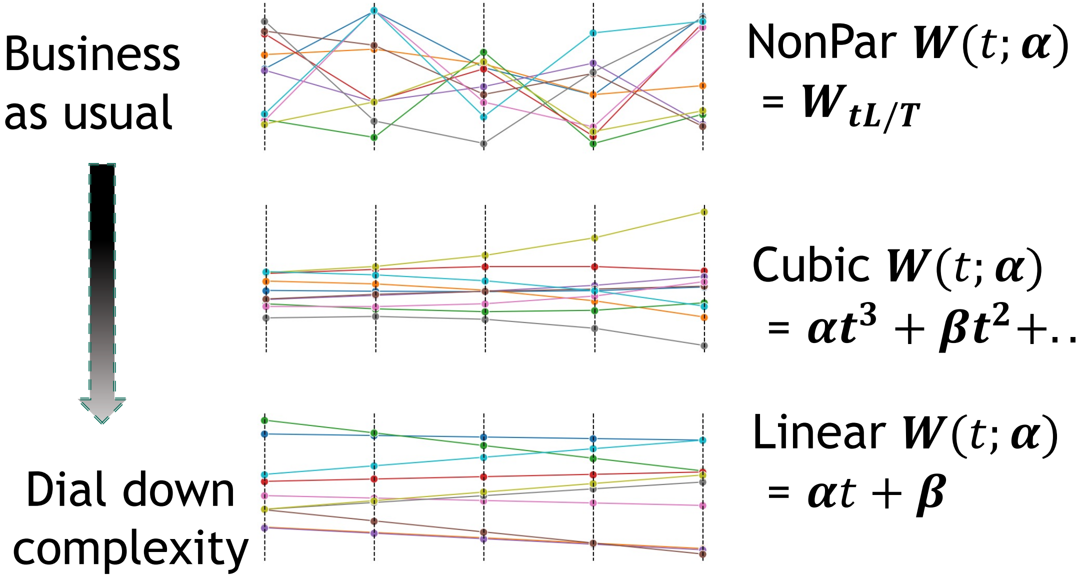
TA: Weight parameterization setup

ResNet: $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \boldsymbol{\alpha}_n \boldsymbol{\sigma}(\boldsymbol{W}_n \boldsymbol{x}_n + \boldsymbol{b}_n)$



Training for weight matrices $\boldsymbol{W}_0, \boldsymbol{W}_1, \ldots$

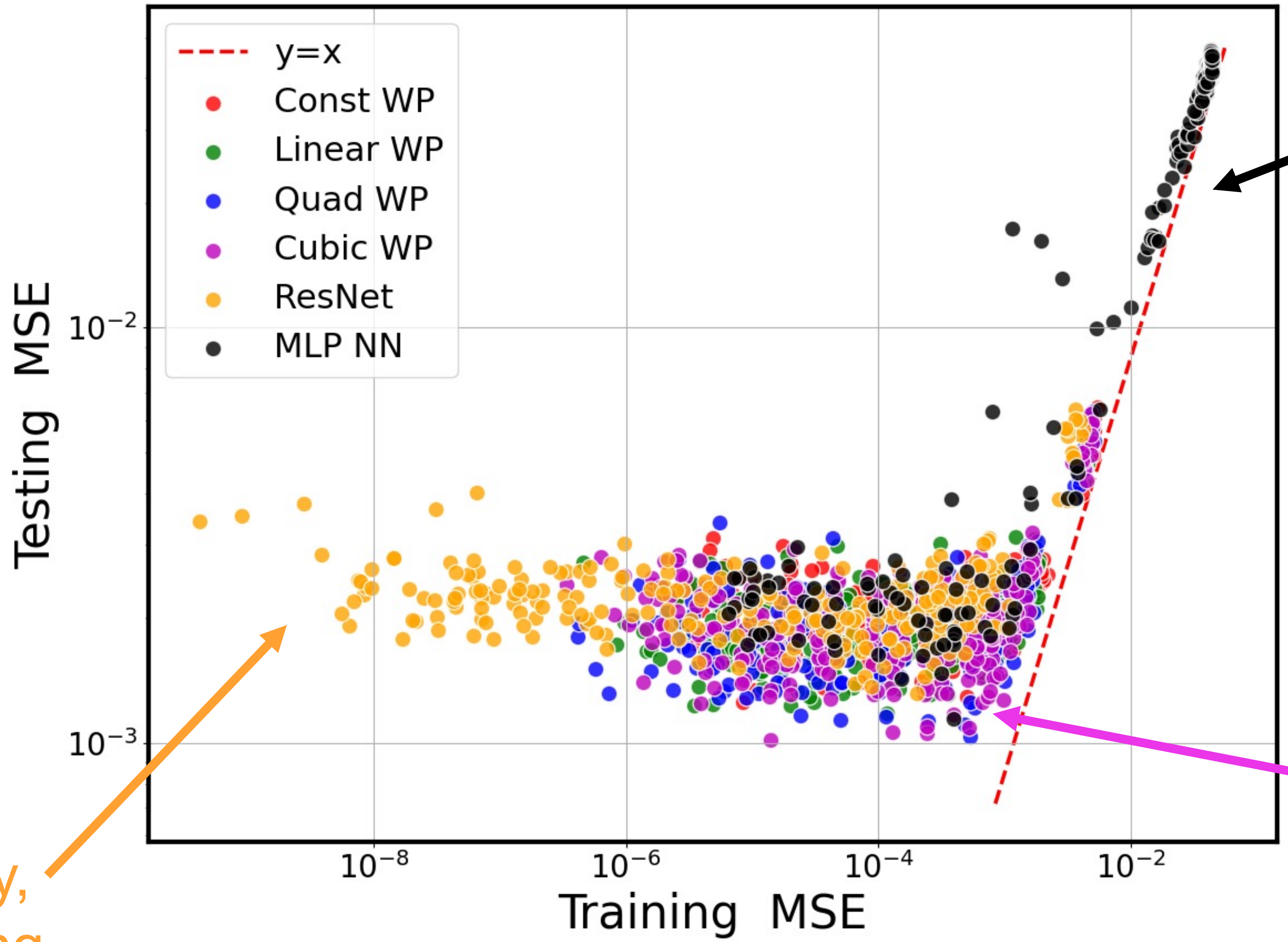
Heavily overparameterized, does not generalize well



Parameterize $\boldsymbol{W}(t;\boldsymbol{\alpha})$ and train for $\boldsymbol{\alpha}$'s.

Parameterization of weight functions reduces capacity and improves generalization

TA: Weight parameterization improves generalization



MLP:
bad accuracy

ResNet:
ok accuracy,
but overfitting

WP:
perfect tradeoff

QUiNN: probabilistic wrapper for NNs

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 $\boldsymbol{p}(\boldsymbol{W_0}), \boldsymbol{p}(\boldsymbol{W_1}), \ldots$
- Three classes of options:

Full Bayesian → Approximate Bayesian → Ensemble methods

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

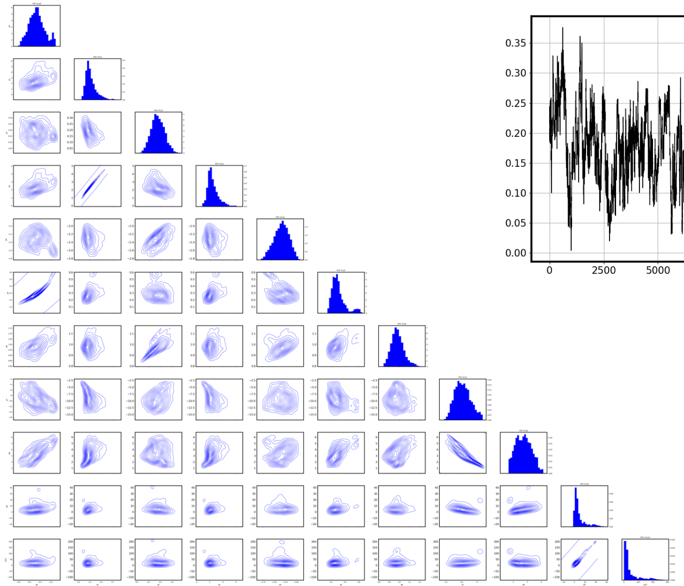
Approximate Bayesian
- ➢ Variational methods
- Typically underestimates extrapolative predictions

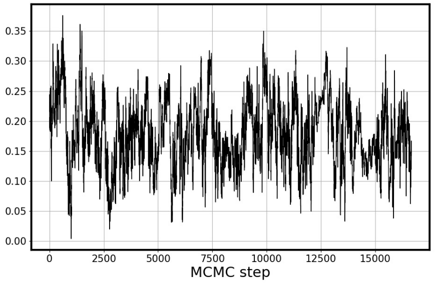
Ensemble methods
- ➢ Heuristic, but works
- Many recent papers viewing deep ensembles as Bayesian approximation

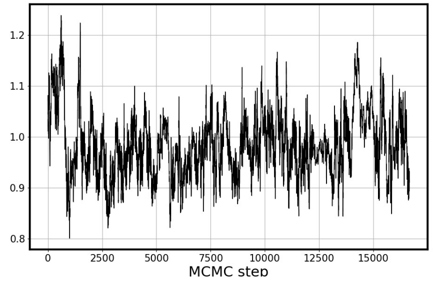
Weight-parameterized ResNets allow these methods to scale much better than before

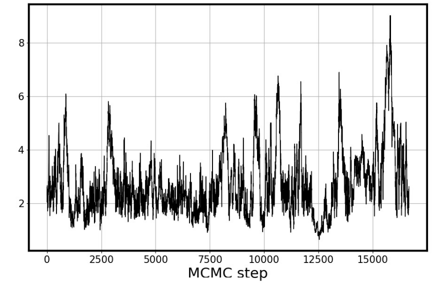
TA: ResNet + WP enables full Bayesian treatment

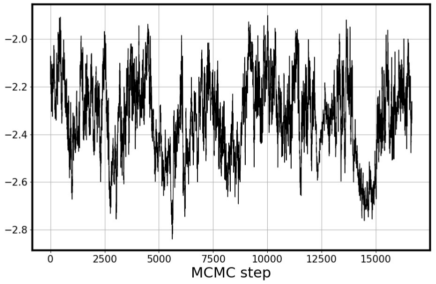
- Number of parameters in ResNets, as well as MLPs, **grows with linearly depth**.
- Number of parameters in weight-parameterized ResNets is **independent of depth**.
- We can easily achieve regimes with manageable MCMC dimensionality and posterior PDFs that out-of-box MCMC methods can easily sample.

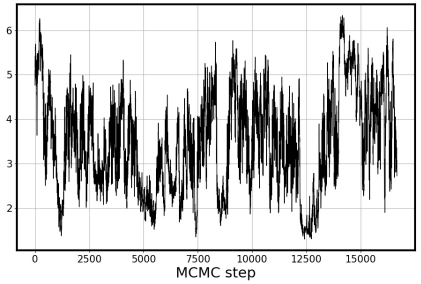






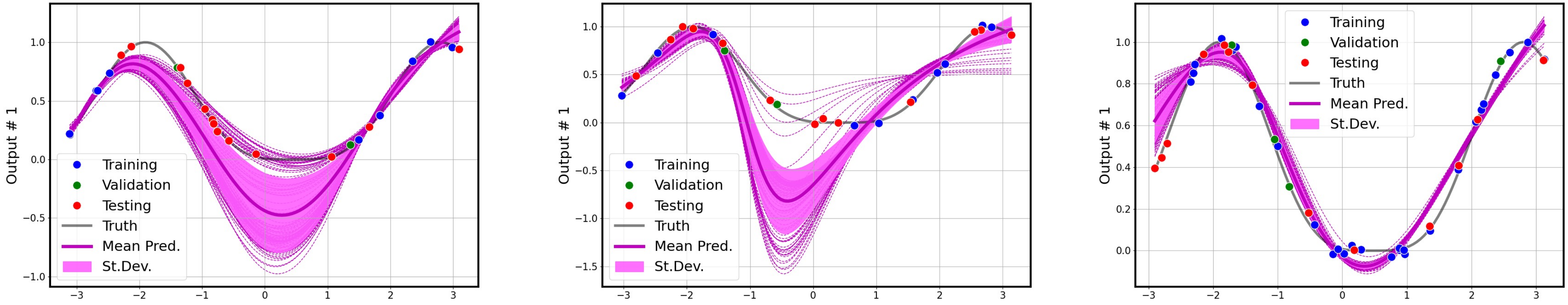






TA: ResNet + WP enables full Bayesian treatment

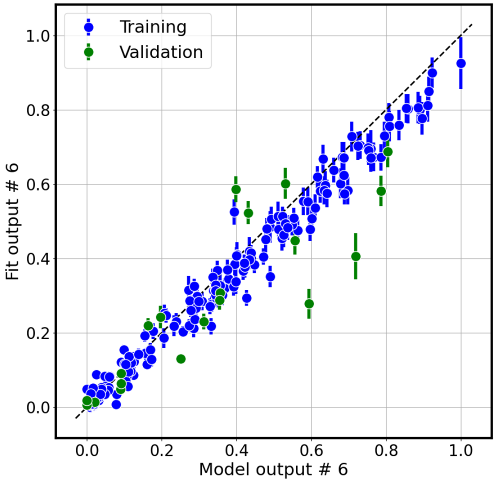
- Number of parameters in ResNets, as well as MLPs, **grows with linearly depth**.
- Number of parameters in weight-parameterized ResNets is **independent of depth**.
- We can easily achieve regimes with manageable MCMC dimensionality and posterior PDFs that out-of-box MCMC methods can easily sample.



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

Climate modeling

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

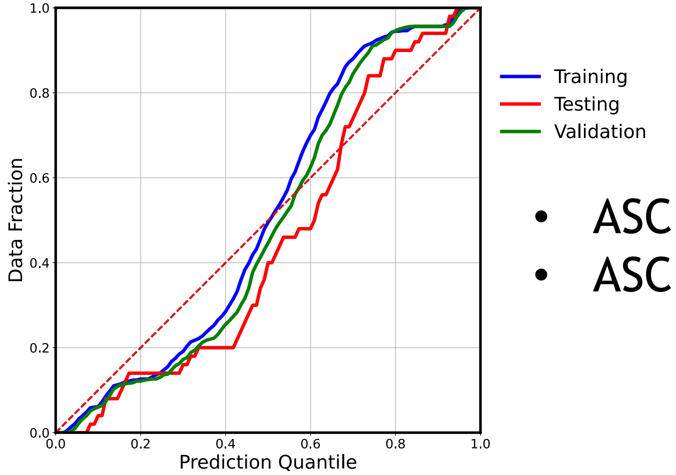


- E3SM, SciDAC efforts require ELM surrogate
- Promising avenues: AI4ESP, BER funding

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

Materials science

- FitSNAP Aluminium dataset, ~10k training
- Bayes-by-backprop for uncertainty estim.
- Energy/force/stress output
- 50:20 WP ResNet



- ASCR/FES project;
- ASC interest

PI's PROJECT LEGACY: Technical

<u>Inspired by ODEs:</u>

- Stiffness penalization made NNs generalize better but not as strongly as we hoped for. Two publications.
- Fractional NODEs: Nonlocal kernel networks, a stable and resolution-independent NN (M. D'Elia).
- Weight parameterization (WP) - worked very well. Reduces overparameterization and improves NN training and prediction accuracy. Two publications in progress. Partnership w Emory University.

<u>Inspired by UQ:</u>

- Uncertainty-augmented NNs are becoming more and more popular, but not industry-standard yet
- UQ and Bayesian NNs made feasible by weight-parameterized ResNets (i.e. discrete NODEs)
- Home-grown software allowed in-depth studies of various UQ methods

PI's PROJECT LEGACY: Technical/Programmatic/Reproducibility

Applications impacted:

- Climate modeling: spatio-temporal surrogate construction for E3SM land model (ELM)
  - Current and future development/deployment guaranteed (since PI is ELM UQ lead)
- Materials science: ML interatomic potentials equipped with uncertainties
  - SciDAC FES-ASCR partnership funded: uncertainty-augmented ML interatomic potentials

Hiring challenges / Lessons learned:

- We never managed to hire the second postdoc; we lost Marta D'Elia to industry; returned $50K
  - Internal hiring / university partnership / student internships
  - Having a software to stand on turned out very useful!

Trace / Reproducibility:

- Wiki https://wiki.sandia.gov/display/CLIACFI/Analysis+of+Neural+Networks+as+Random+Dynamical+Systems
- Internal repository for all documents/codes gitlab-ex.sandia.gov/ksargsy/nnrds
- MICA up-to-date; SAND report in progress.

PROJECT OUTPUTS: Software and publications

- QUiNN library: https://github.com/sandialabs/quinn
  - Quantification of Uncertainties in Neural Networks (QUiNN) is a python library with probabilistic wrappers over PyTorch modules to provide uncertainty estimation in Neural Network (NN) predictions.

- Publications
  - H. You, Y. Yu, M. D'Elia, T. Gao, S. Silling, 'Nonlocal Kernel Network (NKN): a Stable and Resolution-Independent Deep Neural Network', Journal of Computational Physics, 469, 111536, November, 2022.
  - J. Hudson, M. D'Elia, H. Najm, K. Sargsyan: 'The Role of Stiffness in Training and Generalization of ResNets', Journal of Machine Learning for Modeling and Computing, 4(2):75–103, 2023.
  - J. Hudson, M. D'Elia, H. Najm, K. Sargsyan: 'Measuring Stiffness in Residual Neural Networks', RAMSES: Reduced order models; Approximation theory; Machine learning; Surrogates; Emulators and Simulators; Workshop Proceedings, accepted, 2023.
  - H. Rosso, L. Ruthotto, K. Sargsyan: 'Weight parameterization in Neural ODEs', in prep, 2023.
  - O. Diaz-Ibarra, H. Najm, K. Sargsyan: 'Land model surrogate construction via weight-parameterized ResNets', in prep 2023.

PROJECT OUTPUTS: Presentations

- J. Hudson, K. Sargsyan, M. D'Elia, H. Najm: 'Analysis of Neural Networks as Dynamical Systems', MLDL Workshop, SNL, July 22, 2021.
- J. Hudson, K. Sargsyan, M. D'Elia, H. Najm: 'Detecting Stiffness in ResNets Inspired by Neural ODEs', RAMSES Workshop, virtual, Dec 14-17, 2021.
- K. Sargsyan, J. Hudson, O. Diaz-Ibarra, M. D'Elia, H. Najm: 'Training and Generalization of Residual Neural Networks as Discrete Analogues of Neural ODEs', MLDL Workshop, SNL, July 26, 2022.
- H. Rosso, L. Ruthotto, K. Sargsyan: 'Weight-Parameterization in Neural ODEs for Surrogate Modeling and Sampling', SIAM MDS, San Diego, CA, Sep 26-30, 2022.
- J. Hudson, K. Sargsyan, M. D'Elia, H. Najm: 'Examining Stiffness in ResNets through Interpretation as Discretized Neural ODEs', SIAM MDS, San Diego, CA, Sep 26-30, 2022.
- K. Sargsyan, J. Hudson, O. Diaz-Ibarra, M. D'Elia, H. Najm: 'Quantifying Uncertainties in Residual Neural Networks and Neural ODEs', UNCECOMP, 5th International Conference on Uncertainty Quantification in Computational Science and Engineering, Athens, Greece, June 12-14, 2023.
- O. Diaz-Ibarra, K. Sargsyan, H. Najm: 'Dimensionality Reduction and Weight-Parameterized Neural Network Surrogates for Climate Models', USNCCM, 17th U. S. National Congress on Computational Mechanics, Albuquerque, NM, July 23-27, 2023.

CAPABILITIES DEVELOPMENT

- QUiNN Library: https://github.com/sandialabs/quinn
  - Includes implementation of weight-parameterized ResNets
  - Allows modular addition of other UQ-for-NN methods
  - Development will be continued under other existing projects

- Career Development
  - Joshua Hudson (postdoc) just started as Assistant Prof. at U of Arkansas
  - Oscar Diaz-Ibarra transitioned to 1446 staff
  - Haley Rosso (grad student from Emory U partnership) did internship with SNL-NM
  - Student intern Javier Murgoitio-Esandi (USC) – year-around now

TEAM BUILDING AND PARTNERSHIPS

- Collaboration with **Emory University**:
  - Prof. Lars Ruthotto is a leader in the field
  - Grad. student Haley Rosso visited SNL-CA in 2022;
  - … also did summer internship w SNL-NM in 2023
- Two SciML proposals have been heavily fueled by this work
  - **UMichigan**: Profs. Karthik Duraisamy, Alex Gorodetsky, Xun Huan
  - **USC**: Profs. Audrey Olivier, Assad Oberai, Roger Ghanem
  - **MIT**: Prof. Youssef Marzouk; **JHU**: Prof. Yannis Keverekdis
- SIAM UQ Minisymposium co-organization on the way
  - Trying to mark our territory and serve as a bridge in UQ-for-NN community
  - Mix of conventional UQ/Computational Science and core ML/Data Science communities
  - Overwhelming response: maxing out to 16 talks (4 mini-sessions)

IA/PM PROJECT LEGACY

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), proposal work with UMich/USC/MIT/JHU

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 works with application data to hone the methods

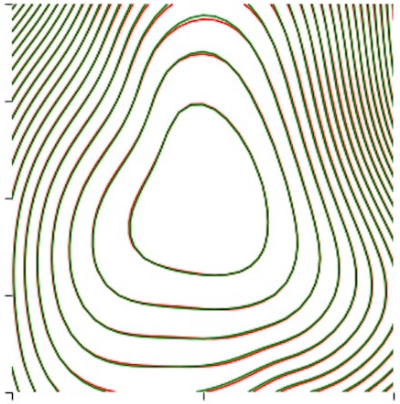
Technology insertion and follow-on funding for potential and realized ROI

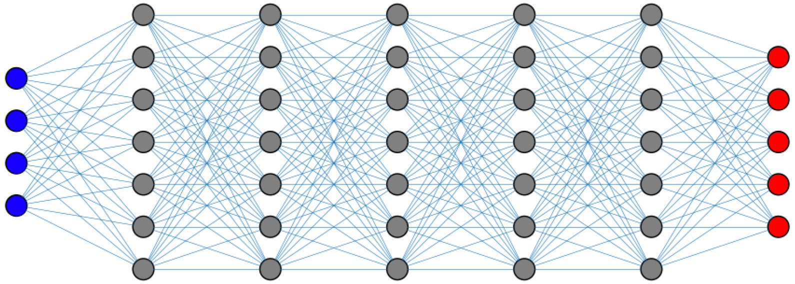
- Fundamental mathematical work, with both code development and impact on real applications
- Strong potential to help Sandia establish leadership in the ML/UQ landscape
- Weight-parameterized ResNets as well as probabilistic NN wrapper software can serve as a base for future funding, both for ASCR (theory) 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 (8739)

Project Goal(s)

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





FY23 Technical Milestones

- Extend regularization analysis under uncertainty
- UQ enabled via weight-parametrization
- 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

- Both algorithmic and software development central to ongoing DOE proposals
- Follow-on funding (BER, FES, BES) for applications of interest highly likely
- QUiNN library released as a bi-product which will serve us well for future DOE funding and ASC/mission engagement

Extra Materials

ResNet and NODE in a regression setting (supervised ML)

ResNet (discrete)

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

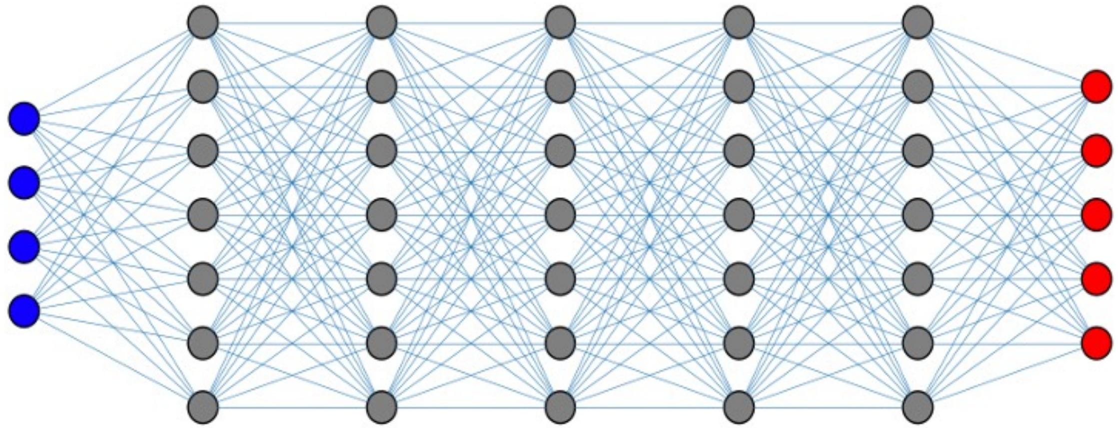
Neural ODE (continuous)

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

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

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

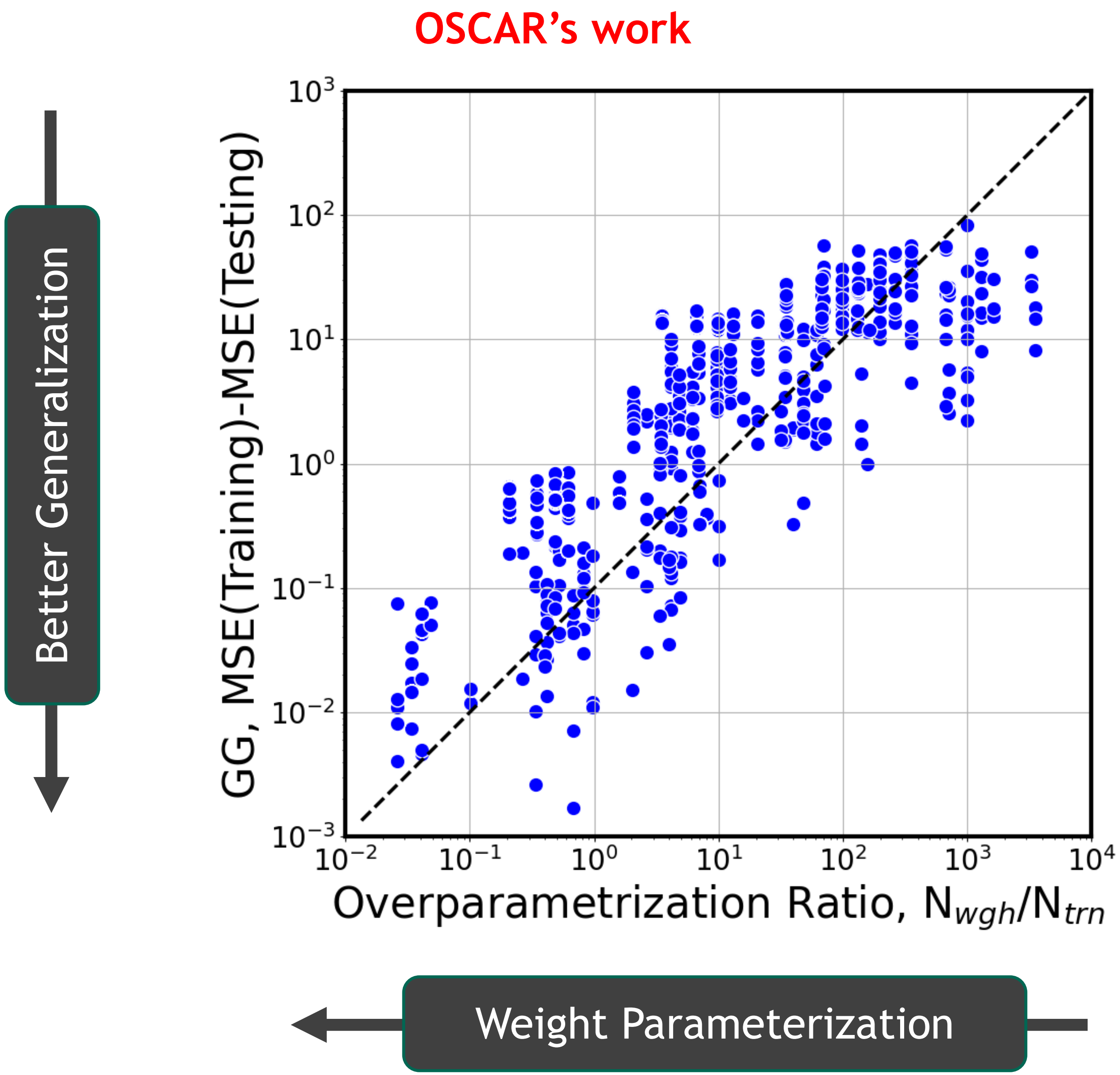
Input



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

Output

TA: Weight parameterization improves generalization



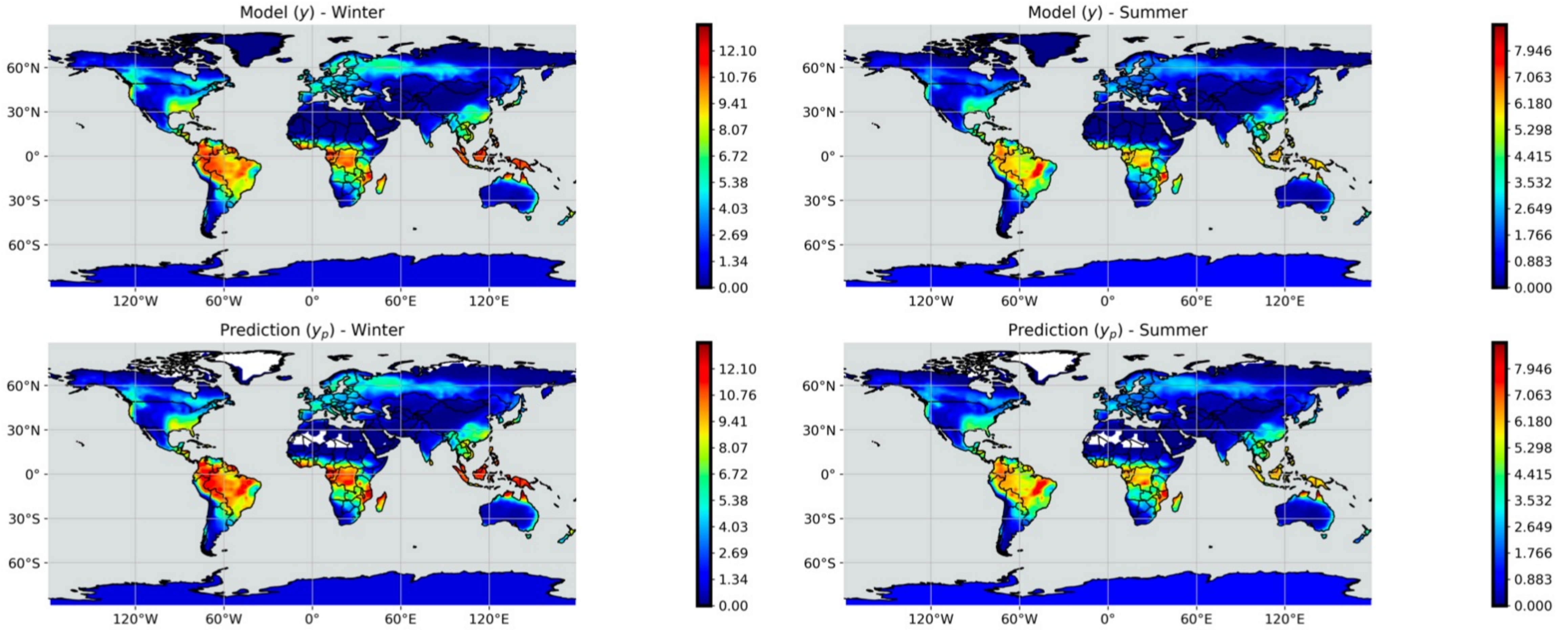
- Generalization Gap correlates with overparameterization

- Weight-parameterized ResNets reduce Generalization Gap

Each dot is a training run with varying weight parameterization functions

E3SM Land Model (ELM) surrogate

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



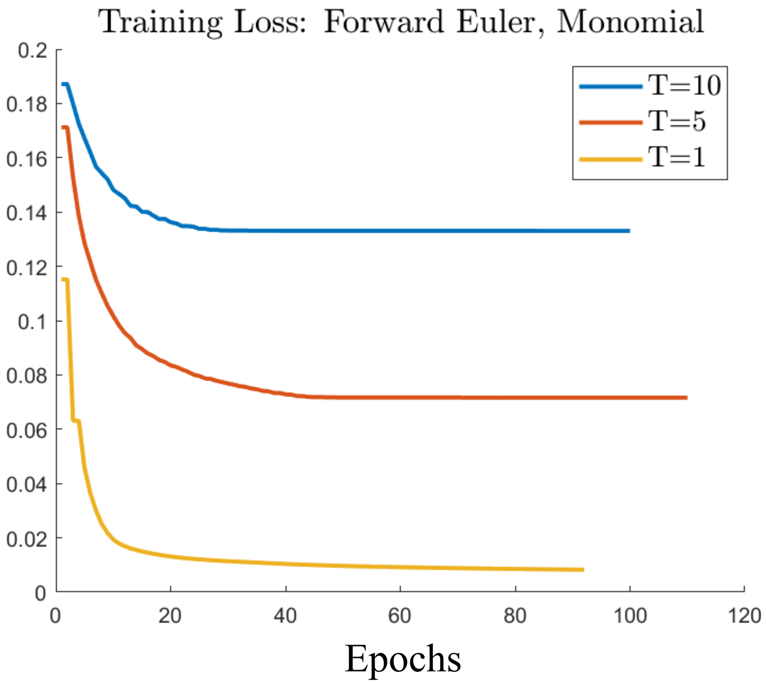
Orthogonal-basis weight parameterization

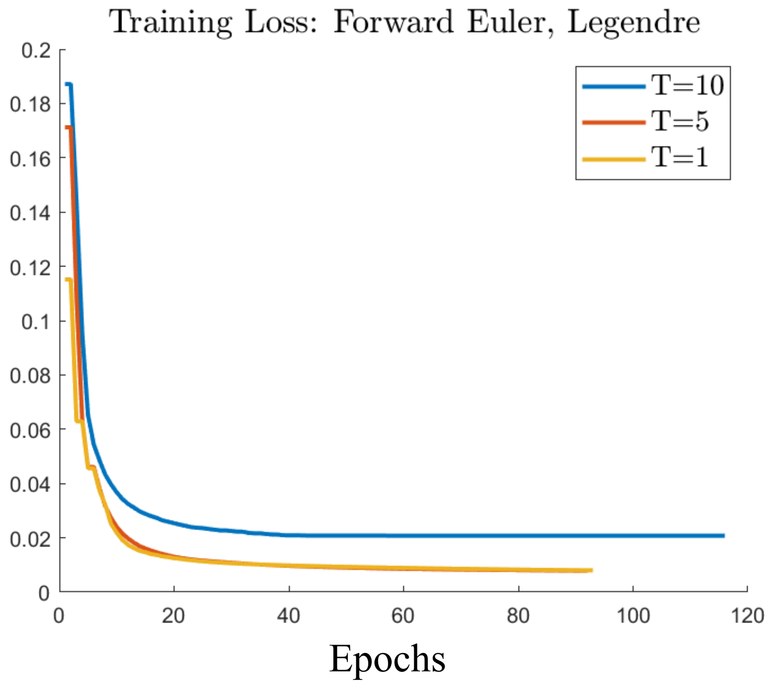
Prof. Lars Ruthotto
Ph.D. student Hailey Rosso
EMORY UNIVERSITY

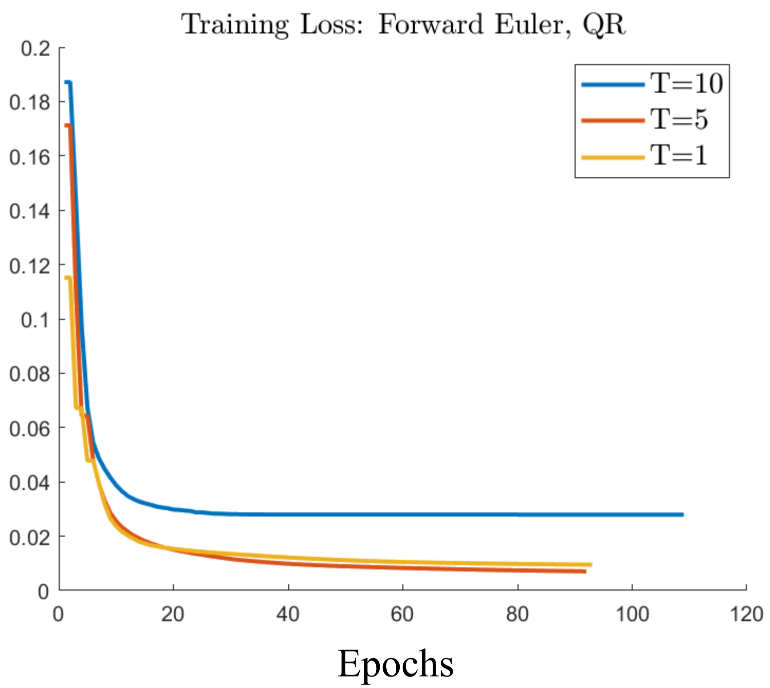
- Basis expansions as weight parameterization
- Training deteriorates as time (*i.e.* NN depth) increases
- Orthogonal basis for NN weight parameterization helps

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

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

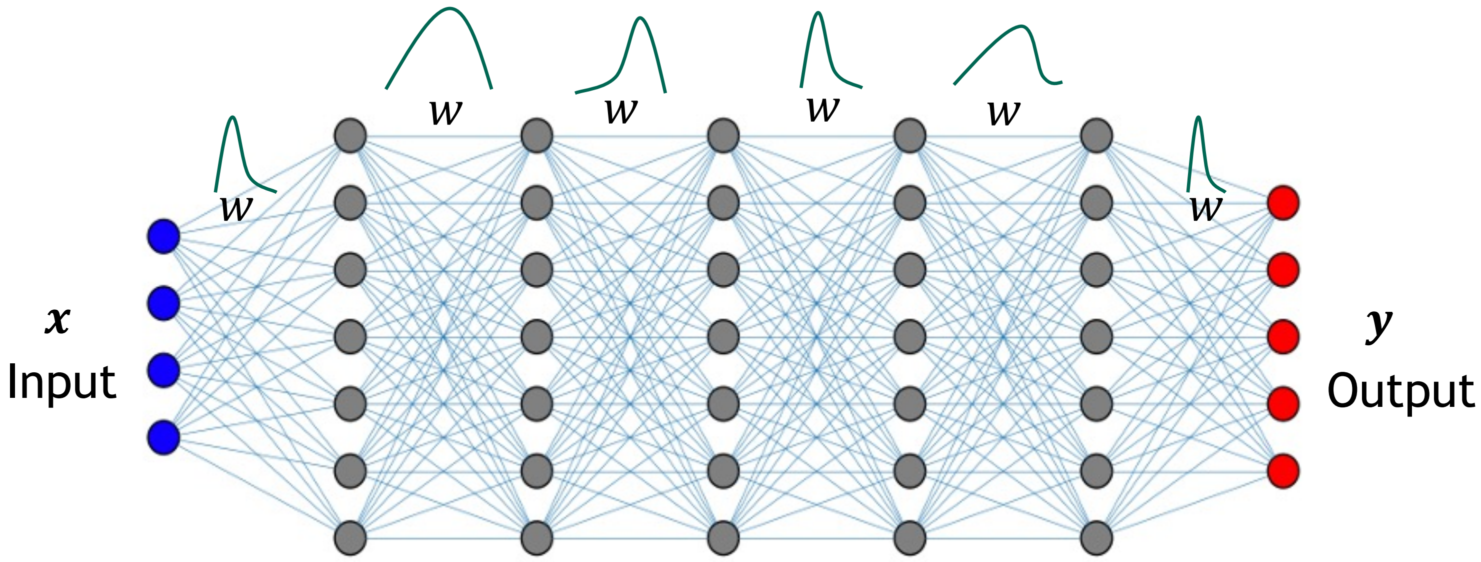






UQ-for-NN: state of the art

- **True Bayesian:** Sampling methods with true posterior distribution



Posterior / Likelihood / Prior

$$p(w\,|\,y) \propto p(y\,|\,w)\,p(w)$$

$$\propto \exp\left(-\frac{||y - f_w(x)||^2}{2\sigma^2}\right)$$

Negative log-posterior ⟷ Deterministic loss function

✓ Markov chain Monte Carlo (MCMC) sampling of posterior; Hamiltonian MC *[Levy, 2018]*

❑ Tuning is an art: essentially infeasible outside academic examples

QUiNN: probabilistic wrapper for NNs

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 $\boldsymbol{p}(\boldsymbol{W_0}), \boldsymbol{p}(\boldsymbol{W_1}), \ldots$
- 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

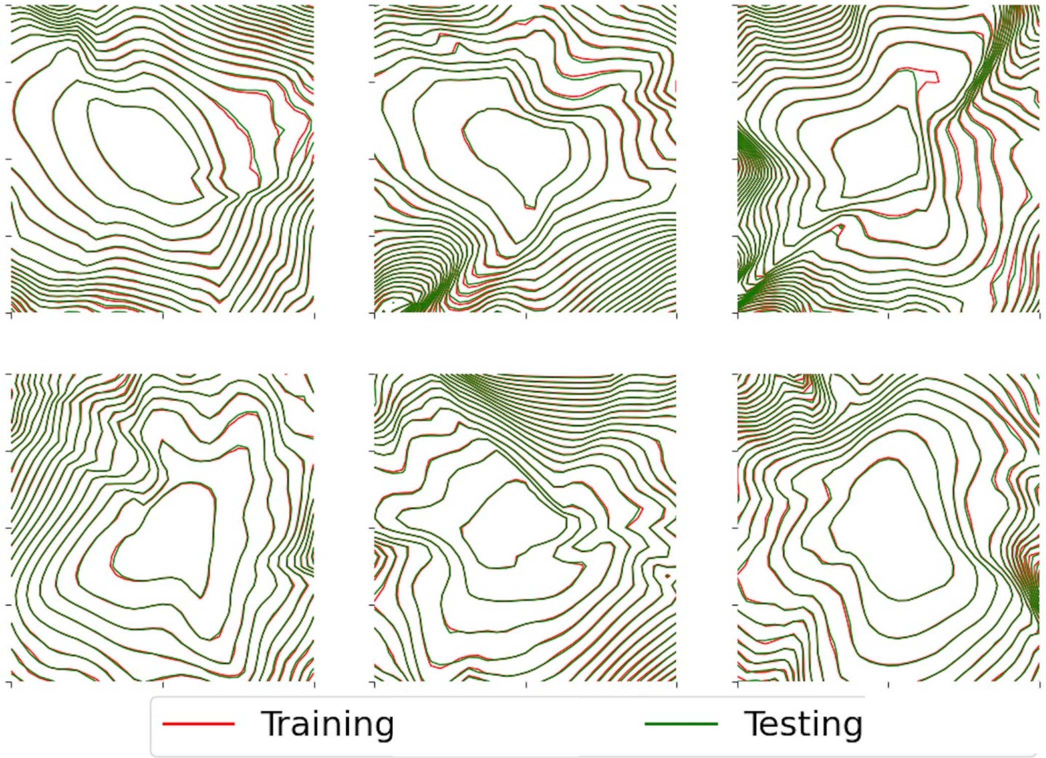
Weight-parameterized ResNets allow these methods to scale much better than before

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

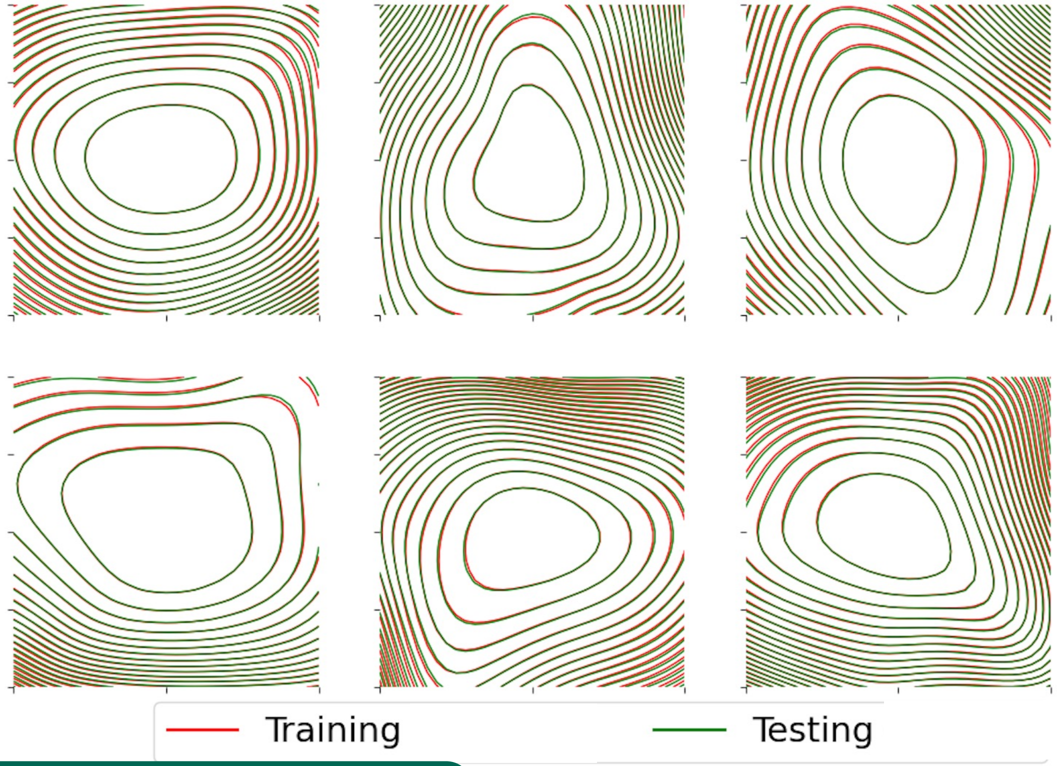
MLP NN: $x_{n+1} = \sigma(W_n x_n + b_n)$

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

Multilayer Perceptron (learning the layer)



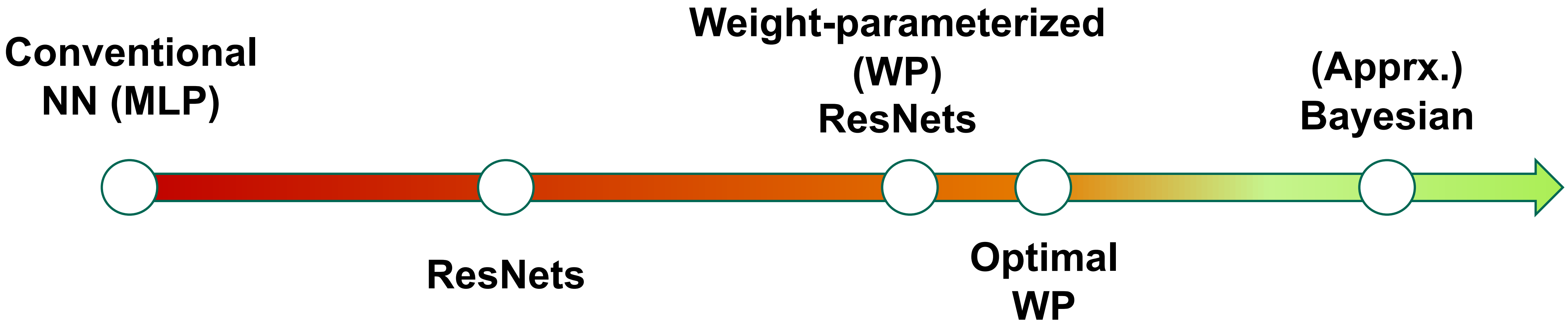
ResNets (learning the layer diff.)



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

Architectural regularization allows UQ

path toward better generalization and confidence assessment



- [Work-in-progress with Lars Ruthotto, Emory U]

orthogonal expansions for WP

work better than monomials

QUiNN: Quantifying Uncertainty in NN

github.com/sandialabs/quinn

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
```

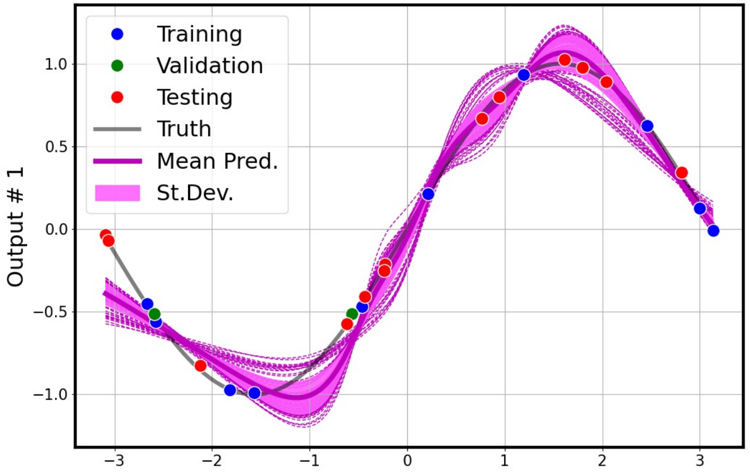
```
uqnet = VI_NN(nnet)
```

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

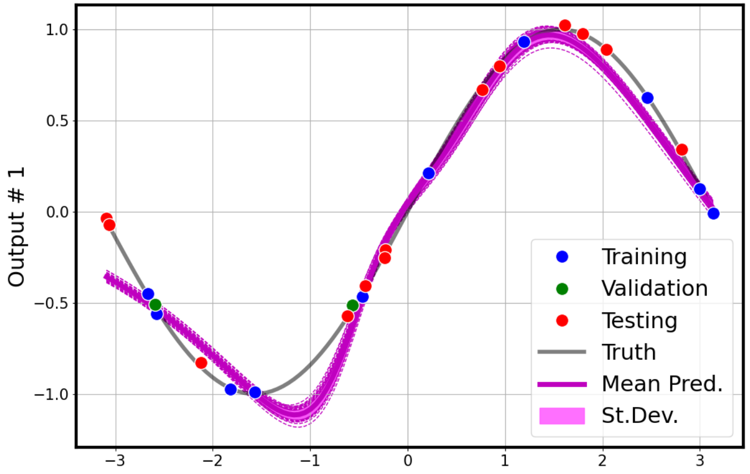
```
uqnet = Ens_NN(nnet, nens=nmc)
```

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

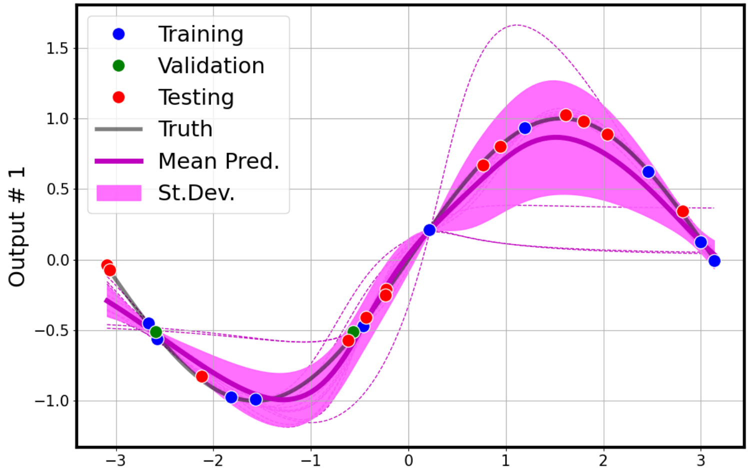
Option 1: MCMC



Option 2: Variational Inference



Option 3: Ensembling



TECHNICAL ACCOMPLISHMENTS: Overview

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

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

ResNet regularization via weight parameterization — *Paper in prog. w Emory U*

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

Probabilistic augmentation for the continuous NODE — *Current work w Emory U.*

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

QUiNN: software for quantifying uncertainties in NNs — *Released at github.com/sandialabs/quinn*

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

Materials science — *FES-funded work*

ASC interest, data from Aidan Thompson and co