Science-guided Machine Learning (Part 2):

Case Studies, Recent Progress, and Future Prospects

Anuj Karpatne

Assistant Professor, Computer Science

Virginia Tech

karpatne@vt.edu

https://people.cs.vt.edu/karpatne/

HDR Grant # 1940247
EAGER Grant # 2026710

Illustrative Case Studies in SGML

- Case Study 1: Science-guided Learning and Hybrid-Science-ML Modeling for Lake Modeling
  - In collaboration with UMN, USGS, U Wisconsin
- Case Study 2: Science-guided Learning for Quantum Mechanics
  - In collaboration with SUNY Binghamton
- Case Study 3: Science-guided Architecture for Lake Modeling
  - In collaboration with USGS, VT Biological Sciences Dept.
- Case Study 4: Hybrid-Science-ML Modeling for Fluid Dynamics
  - In collaboration with VT Mechanical Eng. Dept.
- Case Study 5: Biology-guided NNs for Discovering Phenotyping Traits
  - In collaboration with Battelle, Tulane U., Drexel U., UW

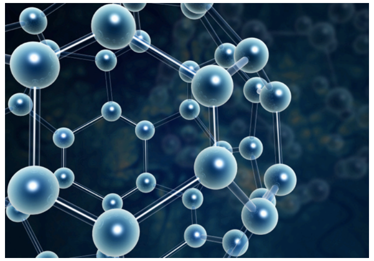
Case Study 1: Science-guided Learning and Hybrid-Science-ML for Lake Modeling

In Collaboration with:

- Motivation:

Growth and survival of fisheries

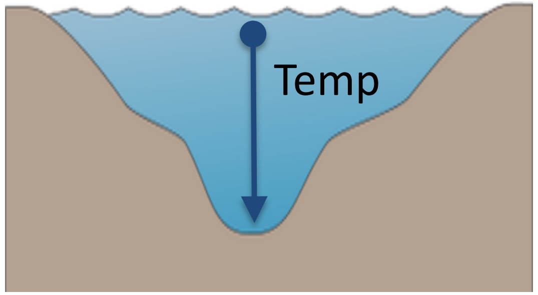
Harmful Algal Blooms



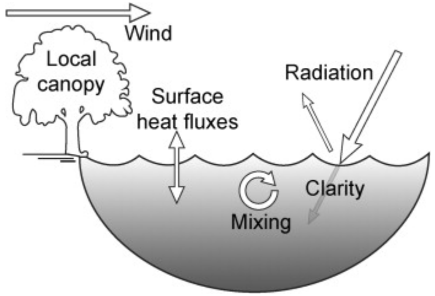
Chemical Constituents: O_2, C, N

- 1-D Model of Temperature:

Target: Temperature of water at every depth in a lake



General Lake Model[1] (GLM)



Input Drivers (observed):

Short-wave Radiation,
Long-wave Radiation,
Air Temperature, …

[1]Hipsey et al., 2014

Case Study 1: Science-guided Learning and Hybrid-Science-ML for Lake Modeling

In Collaboration with:

- Motivation:

Growth and survival of fisheries

Harmful Algal Blooms

Chemical Constituents: O_2, C, N

-

Can we combine physics-based models (e.g., GLM) with data science models to create hybrid-science-ML models?

Hybrid-Science-ML Modeling

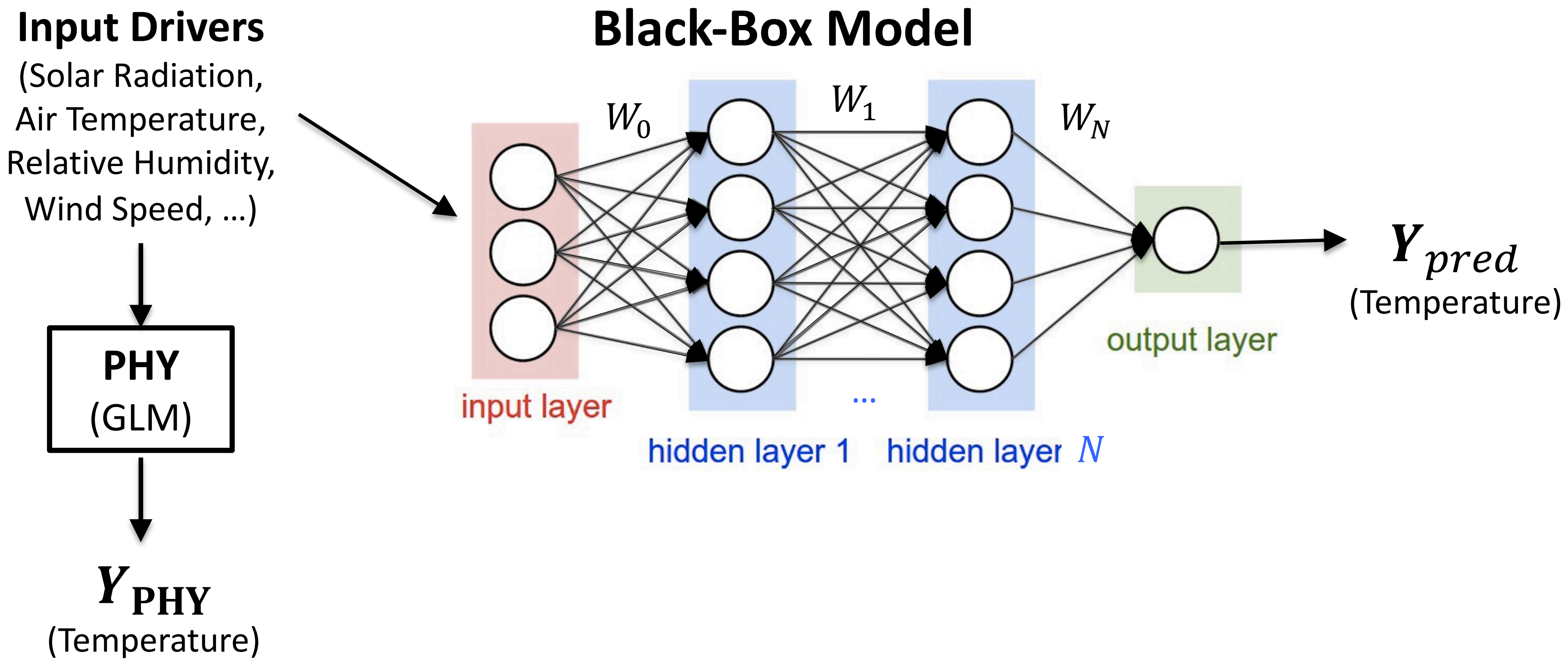
General Lake Model[1] (GLM)

Input Drivers (observed):

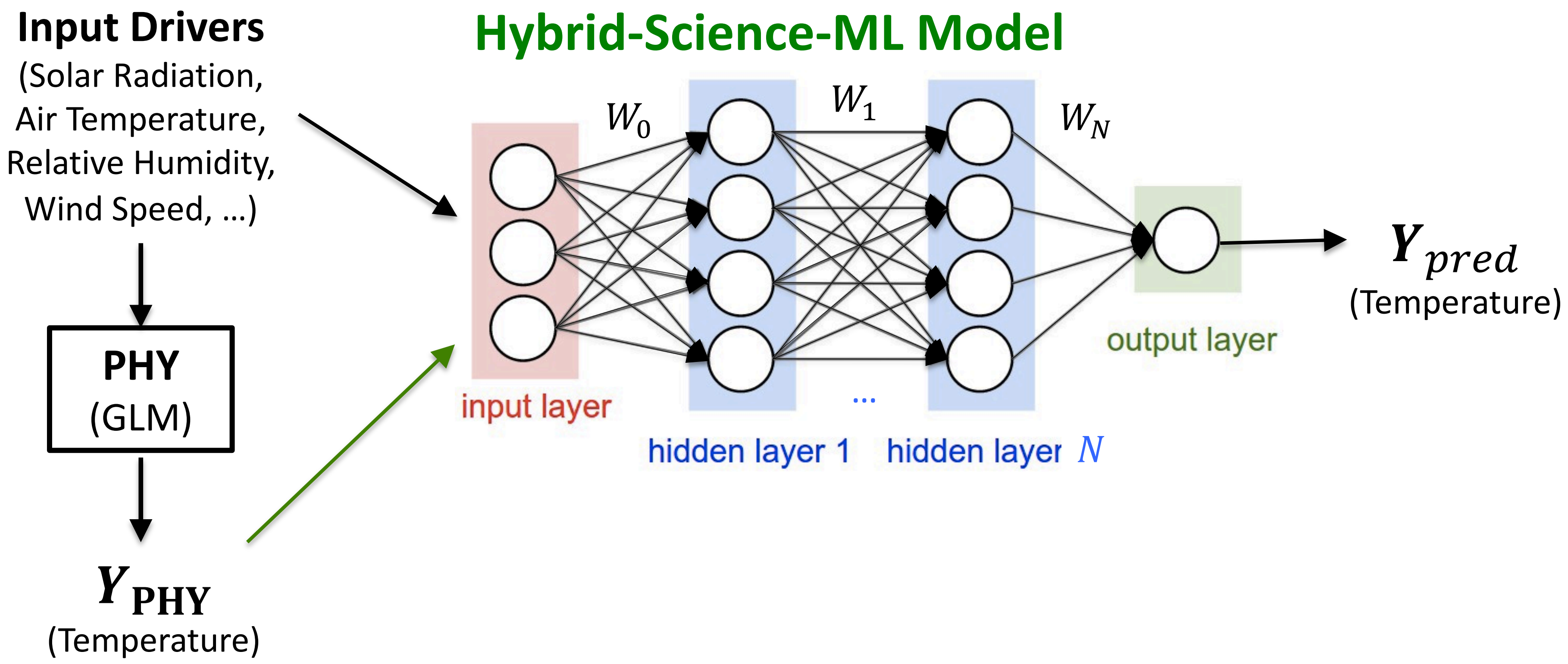
Short-wave Radiation,
Long-wave Radiation,
Air Temperature, …

[1]Hipsey et al., 2014

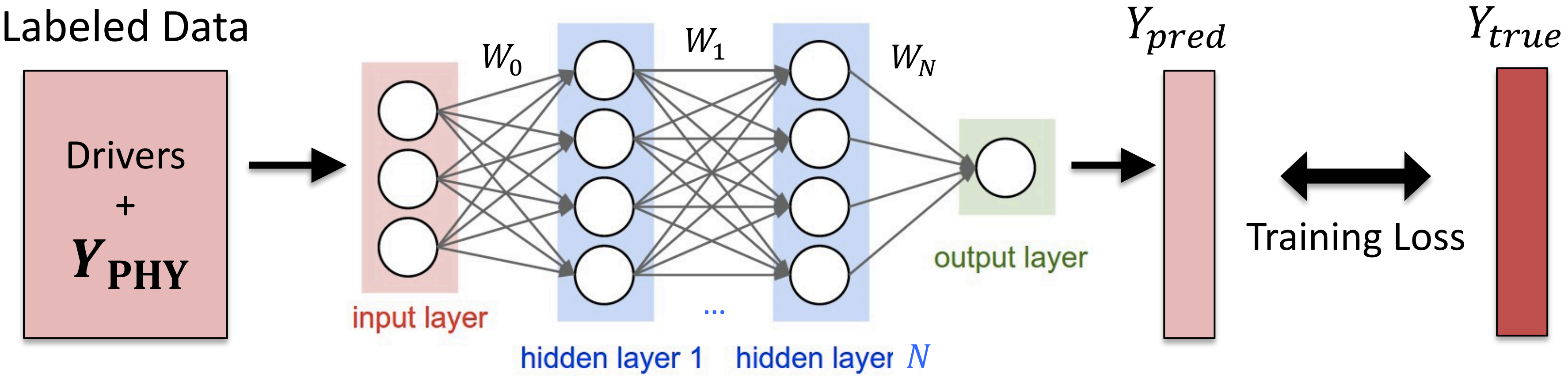
A Generic Framework for
Hybrid-Science-ML Modeling:



A Generic Framework for
Hybrid-Science-ML Modeling:



Training Hybrid-Science-ML Models



$$\text{Objective} := \text{Training Loss}\left(Y_{true}, Y_{pred}\right) + \lambda\, \mathrm{R}(W)$$

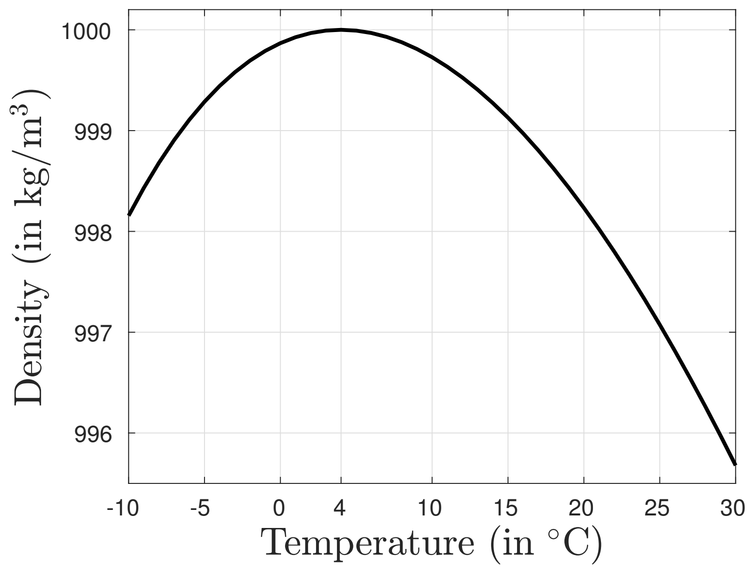
Regularization *(e.g., L1/L2-norm)*

Challenges:

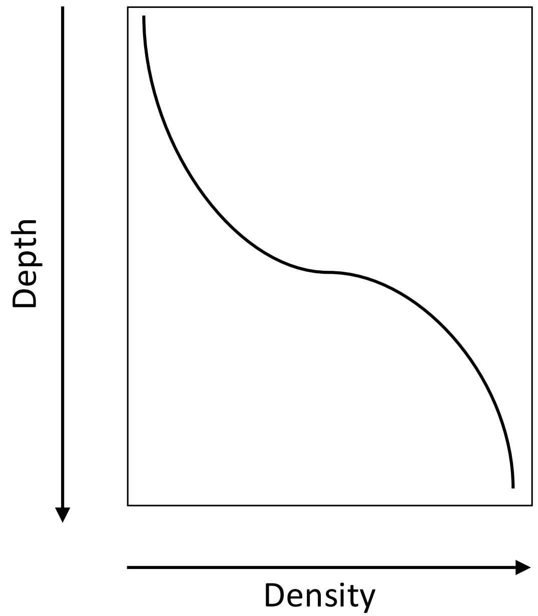
1. Labels (Y_{true}) are scarce
   - Difficult to train models with sufficient complexity
   - Standard methods for assessing generalization performance break down
2. Y_{pred} may violate **physical relationships** b/w Y and other variables

Physical Relationships of Temperature

Temperature directly related to density of water



Denser water is at higher depth



How can we ensure that Y_{pred} is physically consistent?

Use **physics-based loss functions:**

Measure violations of monotonic relationships b/w density and depth.

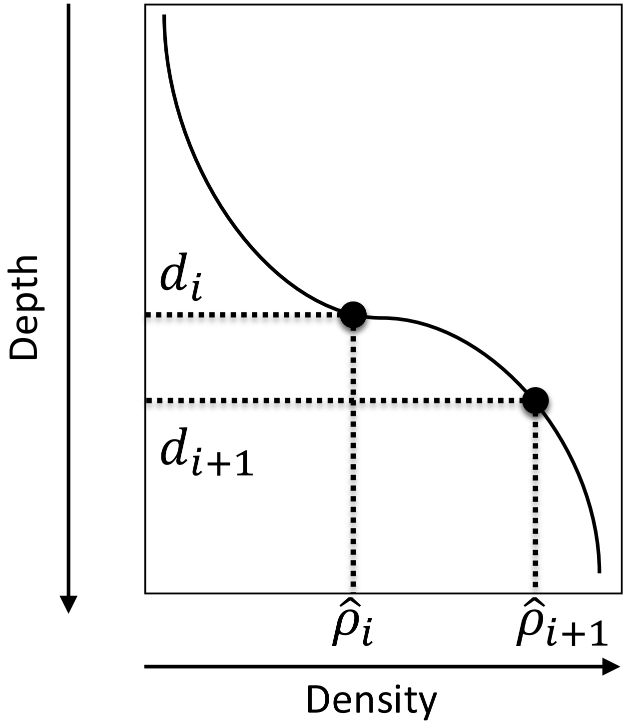
Science-guided Learning

Physics-based Loss for Modeling Temperature

Science-guided Learning

Physical Constraint:

$\hat{\rho}$ should increase with depth



For any consecutive depth pair, $d_i < d_{i+1}$

$$\Delta_i = \hat{\rho}_i - \hat{\rho}_{i+1} \leq 0$$

Physical Violation $= \mathrm{ReLU}(\Delta_i)$

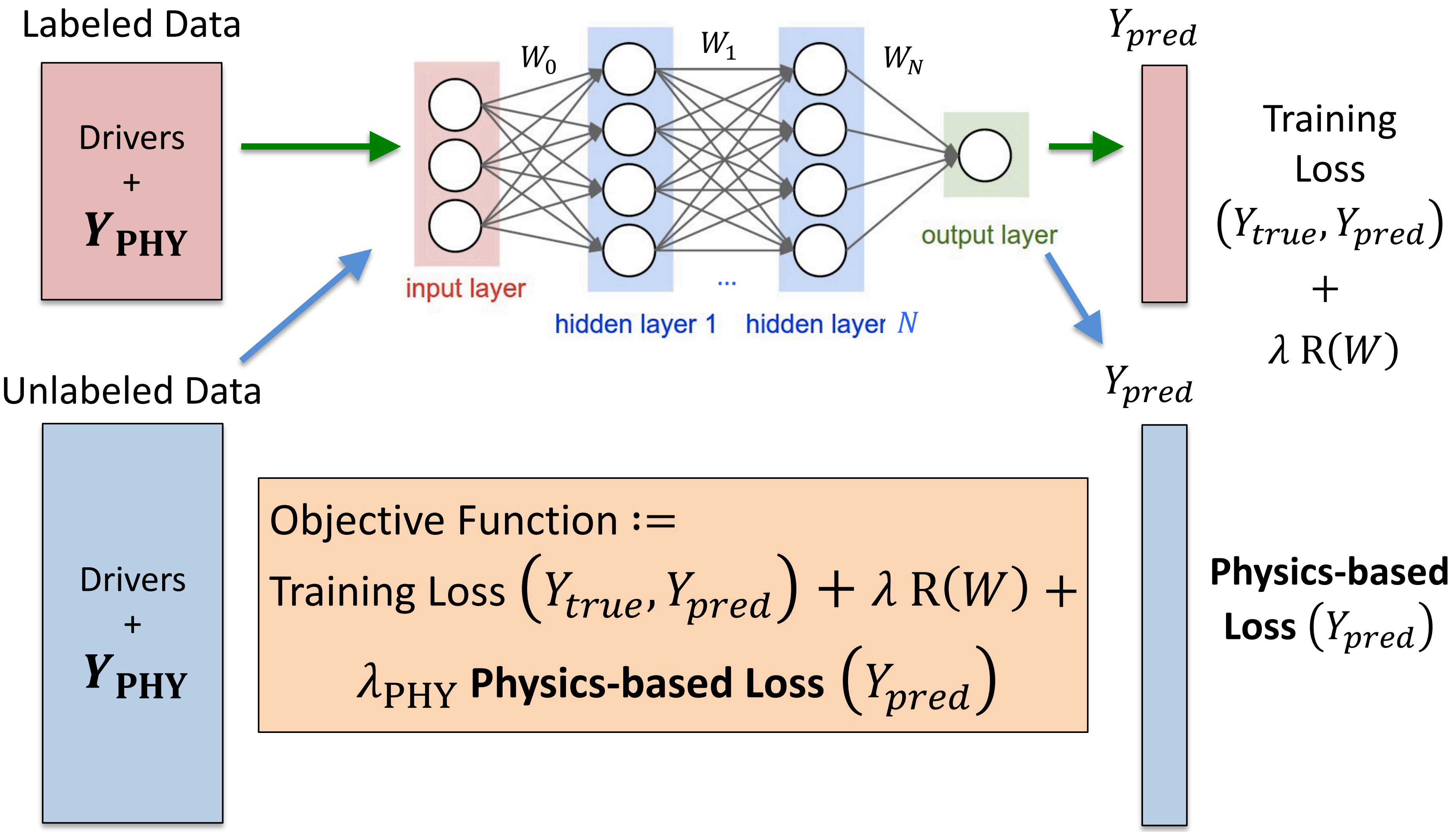
Physics-based Loss $\left(Y_{pred}\right) = \sum_i \mathrm{ReLU}(\Delta_i)$

- Does not require labels (Y_{true}) !
- Can be evaluated on unlabeled data

Physics-guided Neural Network (PGNN)[1]

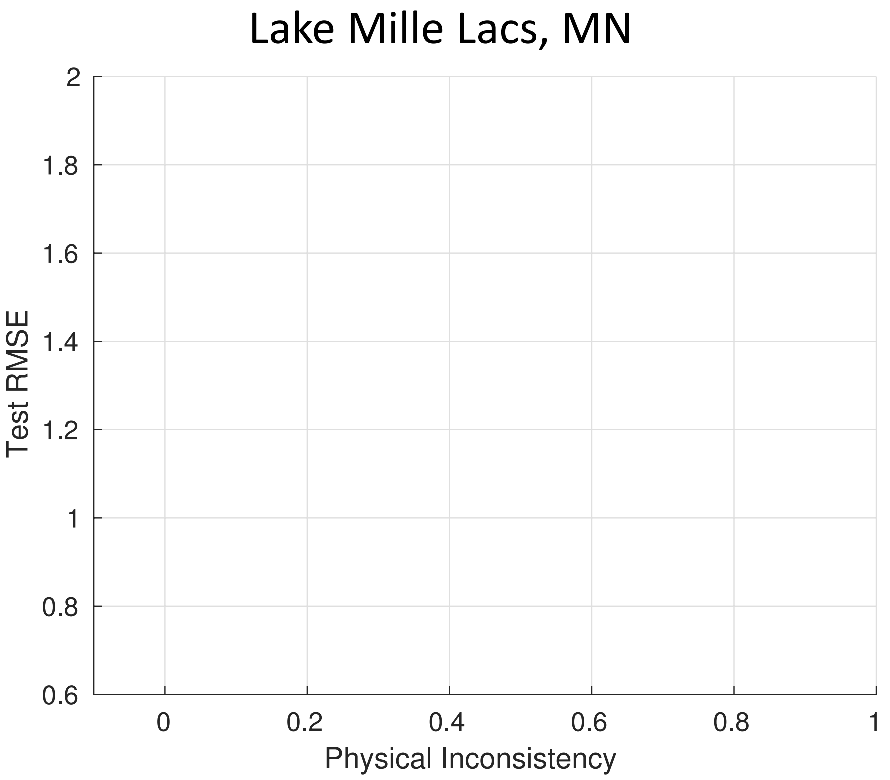
Hybrid-Science-ML

Science-guided Learning

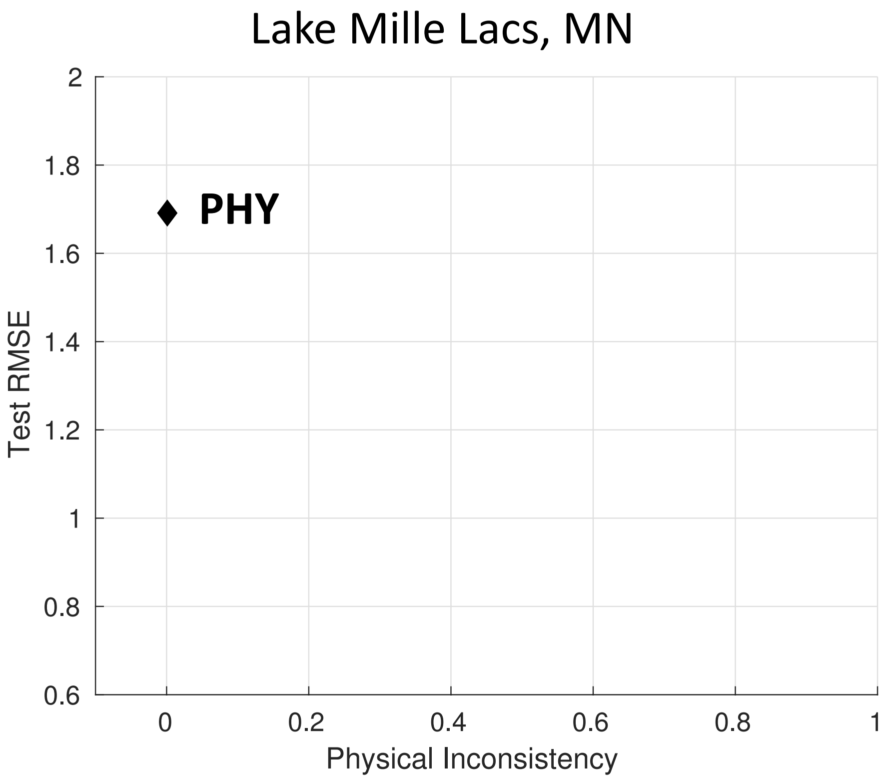


[1]Karpatne et al., "**Physics-guided neural networks (PGNN): An Application in Lake Temperature Modeling**," arXiv: 1710.11431, 2017.

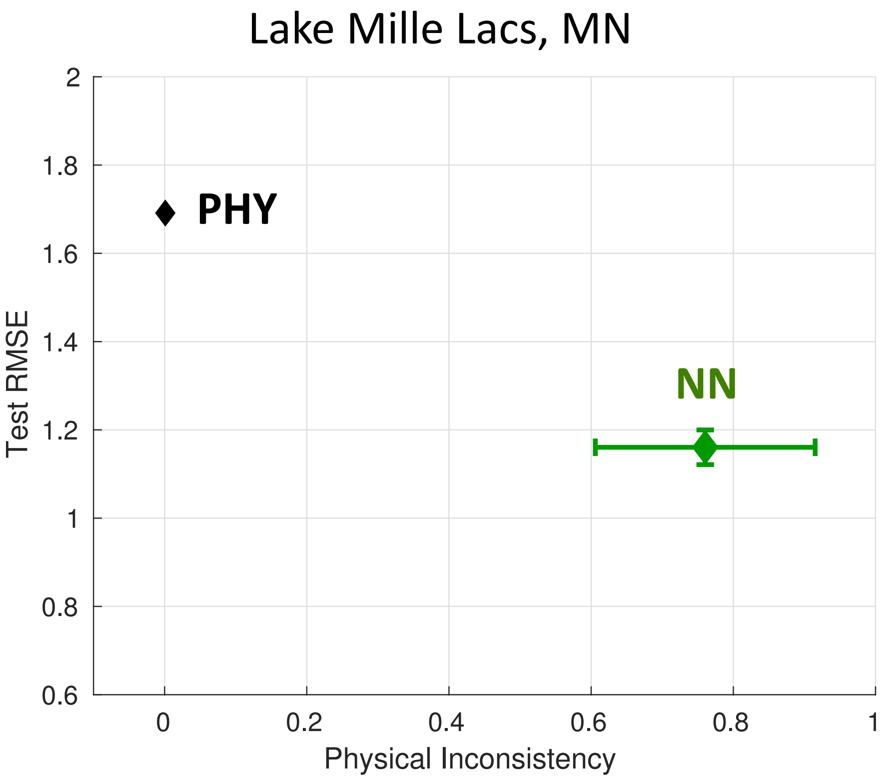
Experimental Results



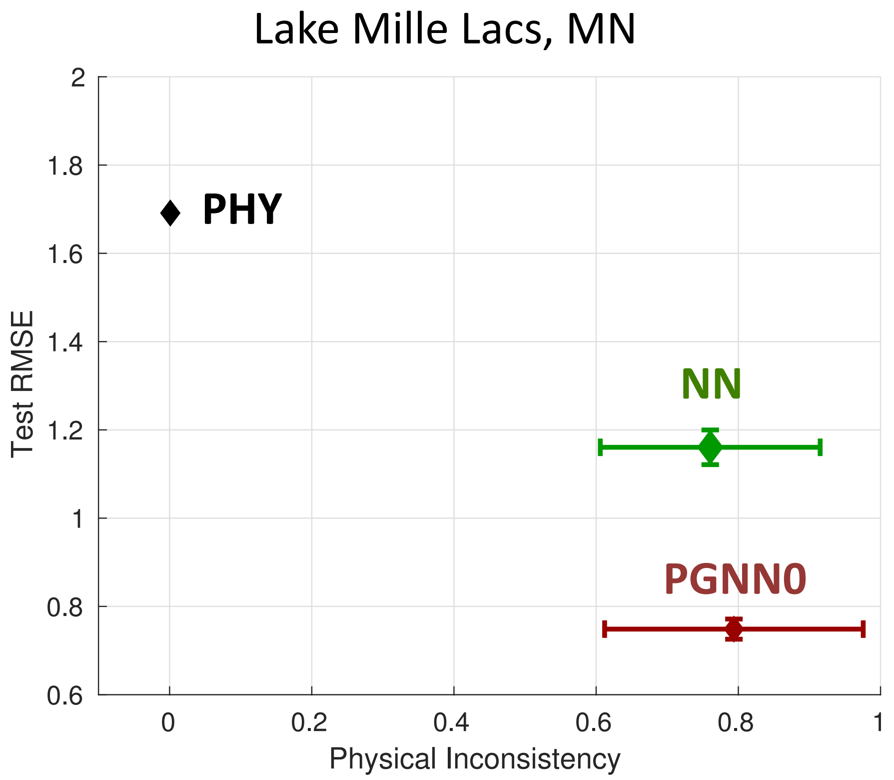
Experimental Results



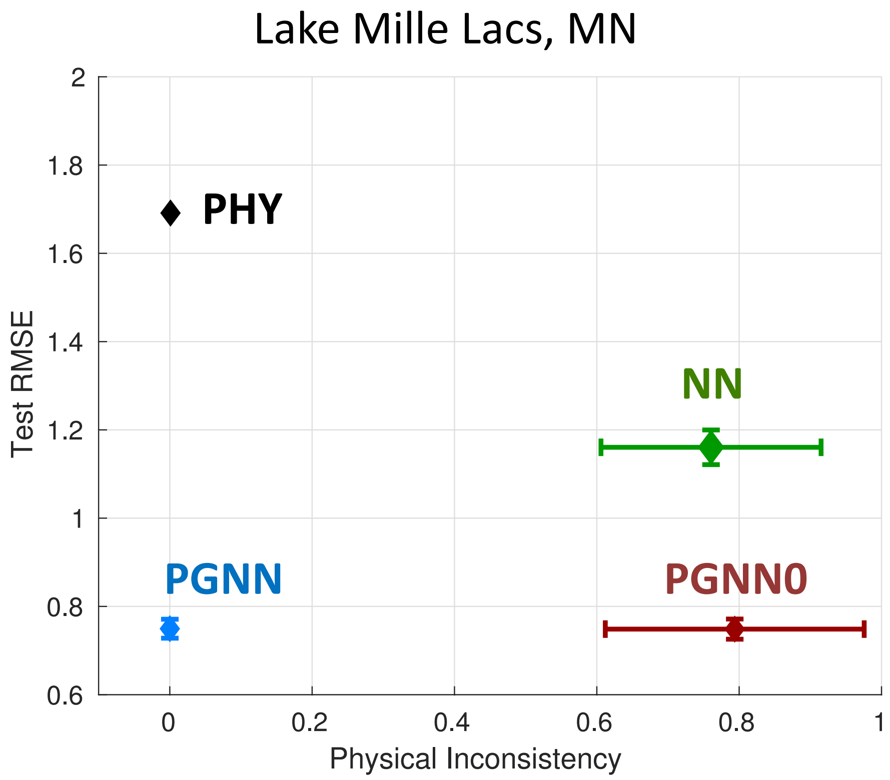
Experimental Results



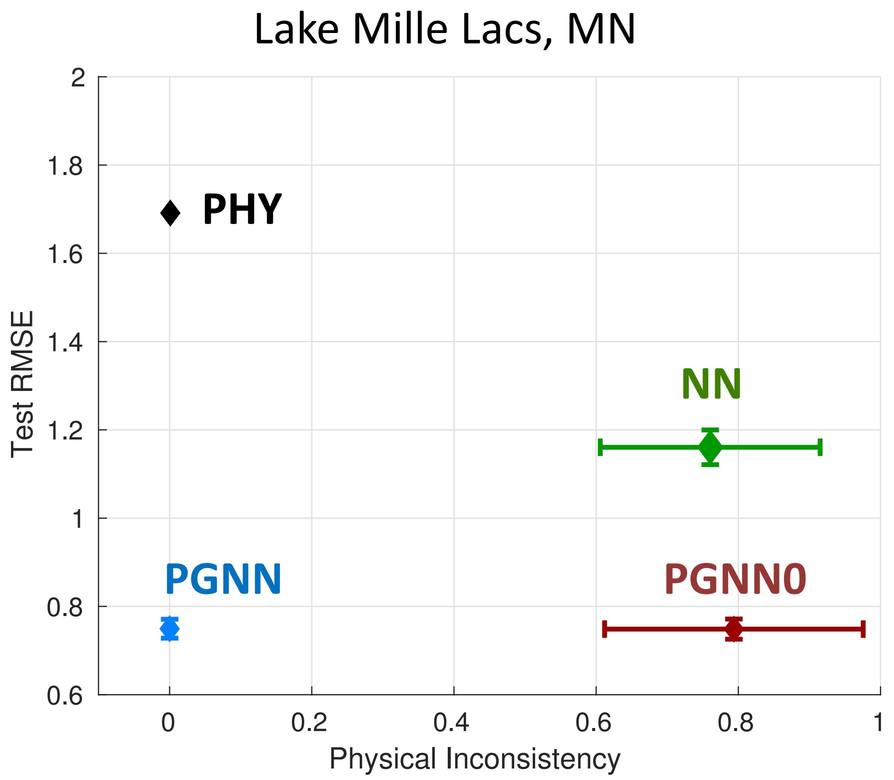
Experimental Results

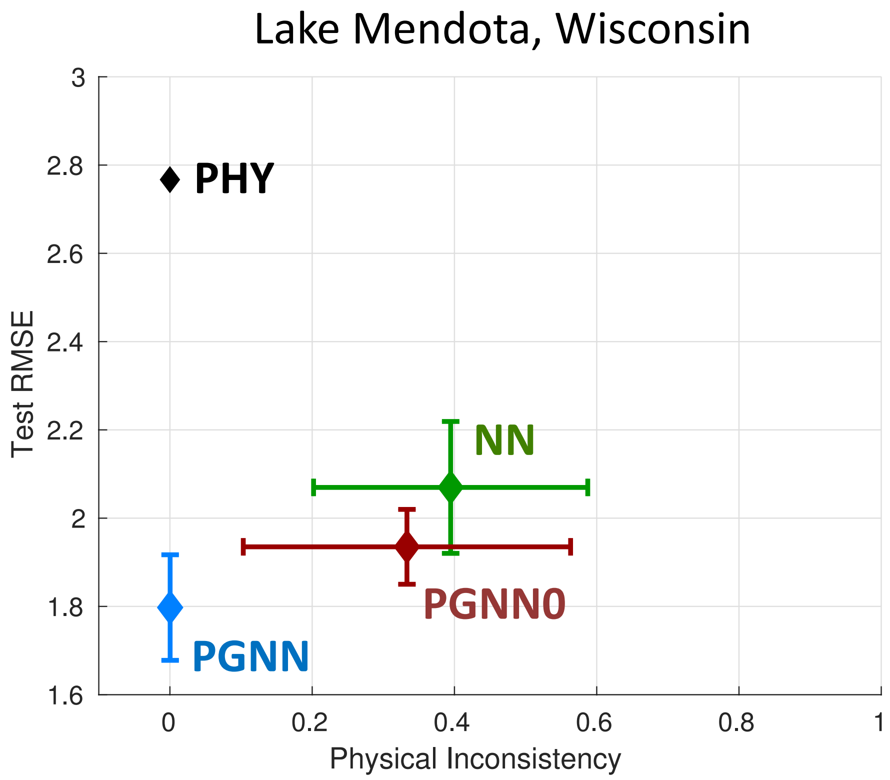


Experimental Results



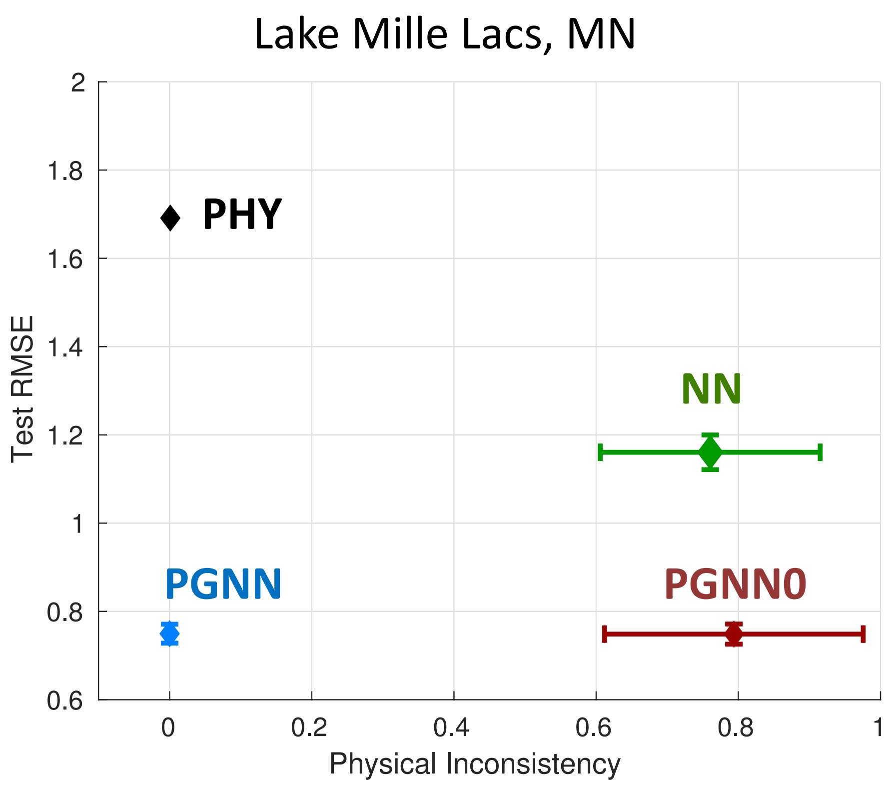
Experimental Results

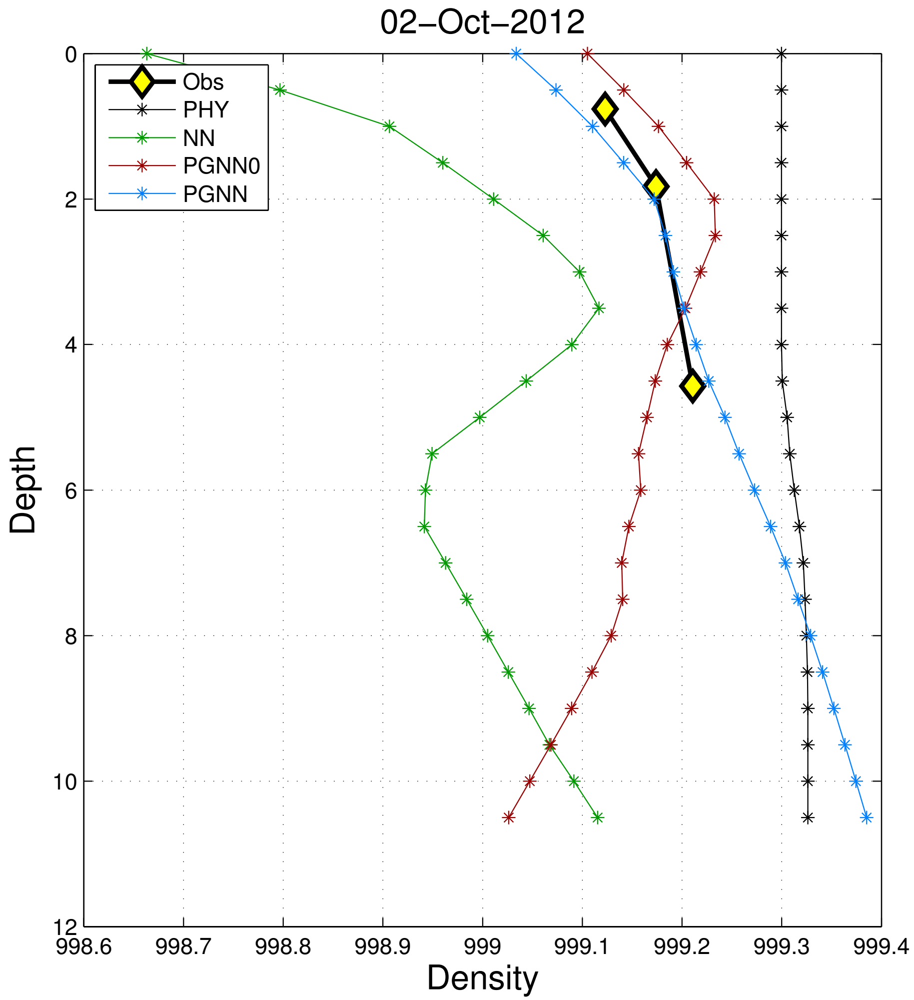




PGNN ensures Generalizability + Physical Consistency

Analyzing Physical Inconsistency

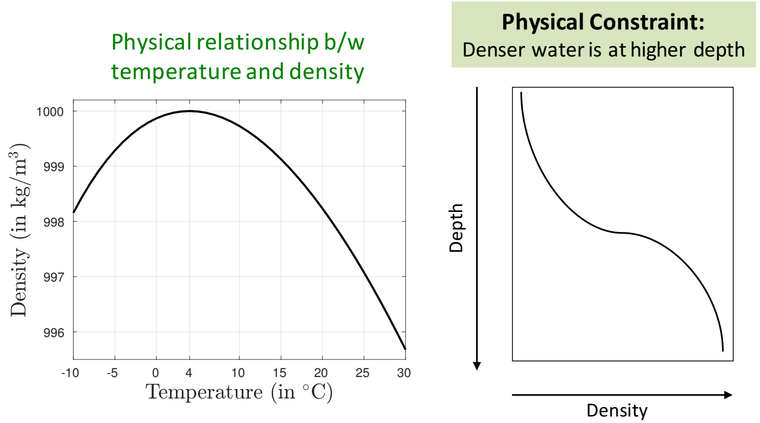




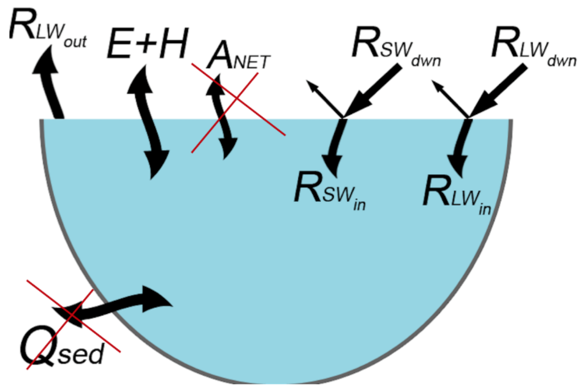
Include **physical consistency** as another evaluation criterion, going beyond standard metrics for test error

Alternate Ways of Incorporating Physics in ML

- Other Physics-based Loss Functions: **Science-guided Learning**



Depth-Density Constraint in Multi-layer Perceptron Network



$dU_t/dt =$
$R_{SW}(1-\alpha_{SW}) + R_{LWin}(1-\alpha_{LW}) - R_{LWout} - E - H$

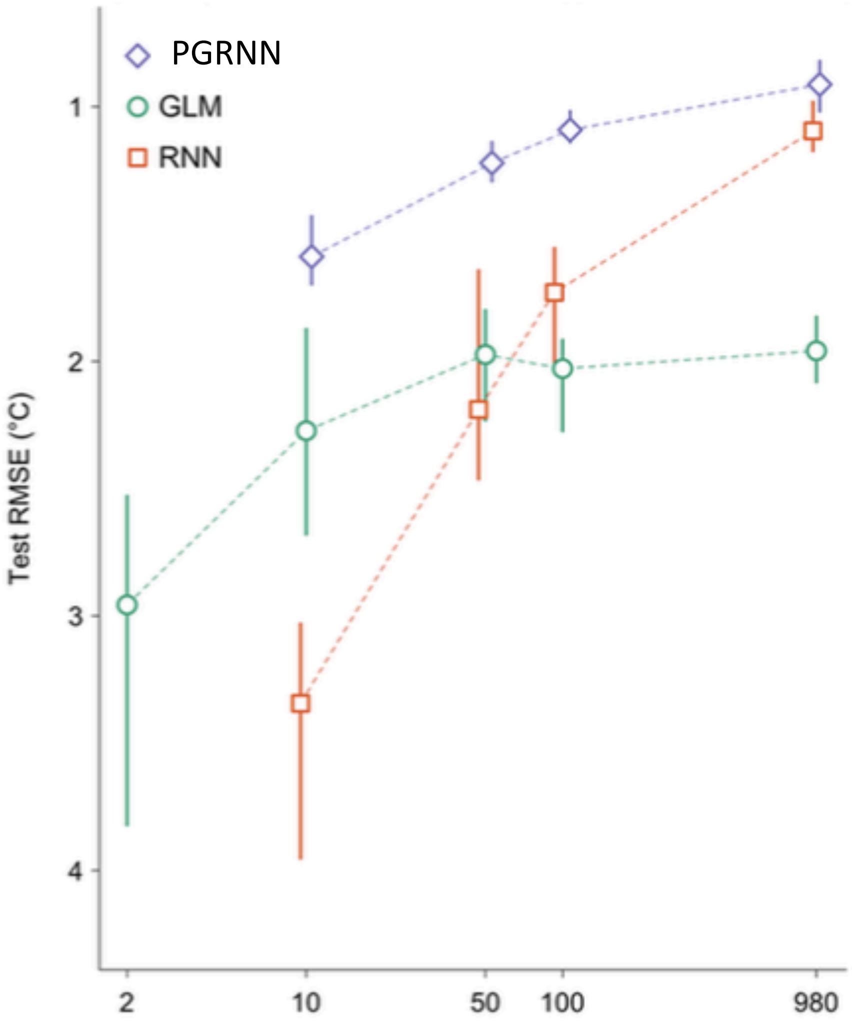
Conservation of Energy in Recurrent Neural Networks

- Pre-training ML models using Physics-based Simulations **Physics-guided Initialization**
  - Train ML methods using physical simulations
  - Fine-tune using observational data

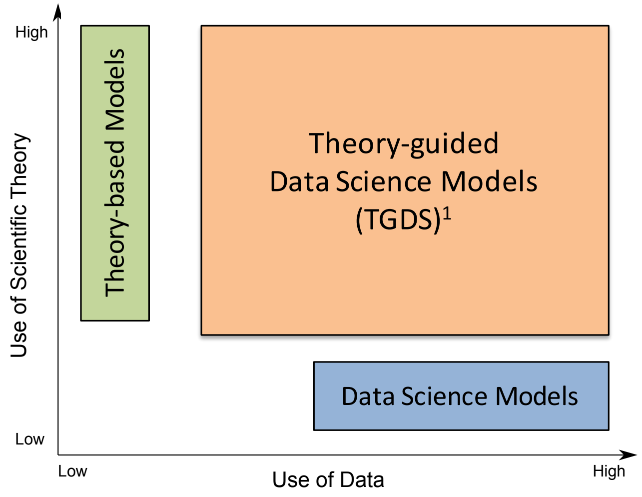
Physics-guided Recurrent Neural Networks (PGRNN)

Physics-guided Initialization

Science-guided Learning



Number of training samples



Jia et al., **Physics Guided RNNs for Modeling Dynamical Systems: A Case Study in Simulating Lake Temperature Profiles,** SDM 2019.

Science-guided Learning:

Recent Progress

$$\text{Prediction Loss}\,(y, \hat{y}) \; + \; \lambda\, \mathrm{R}(\theta) + \; \lambda_{\mathrm{PHY}}\, \textbf{Physics-guided Loss}\,(\hat{y})$$

D_{Tr} D_U

- Advantages of physics-guided neural networks (PGNNs) in lake modeling:
  - Requires far fewer samples in D_{Tr}
  - Better generalizability to novel testing scenarios
  - Ensures physical consistency of outputs

Read et al., “Process-guided deep learning predictions of lake water temperature." WRR 2019.

- Rapidly growing work on using physics-guided loss in various applications

See survey by Willard et al. 2020

  - Example: physics-informed neural networks (PINNs) Raissi et al. 2019
    - Construct physics-guided loss to measure consistency with PDE equations
    - “Label-free” learning only using physics-guided loss
    - Promising results on simplified PDEs

Science-guided Learning:

Open Questions

$$\text{Prediction Loss}\,(y,\hat{y}) + \lambda\,\mathrm{R}(\theta) + \lambda_{\mathrm{PHY}}\,\textbf{Physics-guided Loss}\,(\hat{y})$$

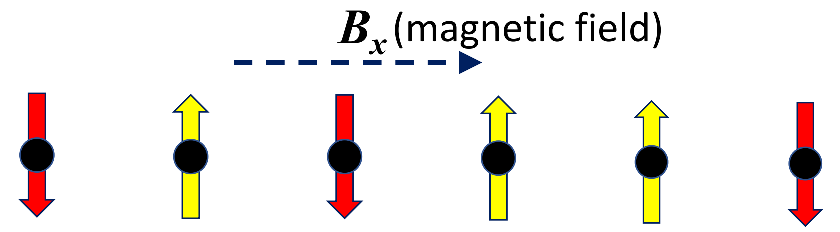
D_{Tr} D_U

- Can we entirely get rid of labeled data D_{Tr} and solely rely on physics-guided loss?
- How should we sample D_U and D_{Tr} to ensure generalizability on novel testing scenarios?
- How can we trade-off physics-loss and data-loss at different stages of ANN learning using adaptive λ_{PHY}?
- How can we deal with multiple physics-guided loss functions, each capturing a different (and possibly, competing) physics objective?

Case Study 2:
Science-guided Learning for Quantum Mechanics

In Collaboration with:
BINGHAMTON UNIVERSITY
State University of New York

Funded by:
NSF
EAGER Grant
2026710



Ising Chain Model

Applications: Comp. Chemistry, Quantum Computing, …

Schrodinger's equation

$$\mathbf{H}\psi = E\psi$$

Hamiltonian Energy Wave Function

Goal: Predict ground-state ψ (with lowest energy) given $\mathbf{H}$

$$\underset{\text{Input}}{\mathbf{H}} \longrightarrow \boxed{\text{ANN}} \longrightarrow \underset{\text{Output}}{\psi}$$

Prediction Loss $(\psi, \hat{\psi})$ + $\lambda\, \mathrm{R}(\theta)$ + λ_{PHY} **Physics-guided Loss** $(\hat{\psi})$

Multiple physics objectives:

Schrodinger Loss (S-Loss):

$$\lambda_S \frac{\left|\left|\mathbf{H}\hat{\psi} - \hat{E}\hat{\psi}\right|\right|^2}{\hat{\psi}^T\hat{\psi}}$$

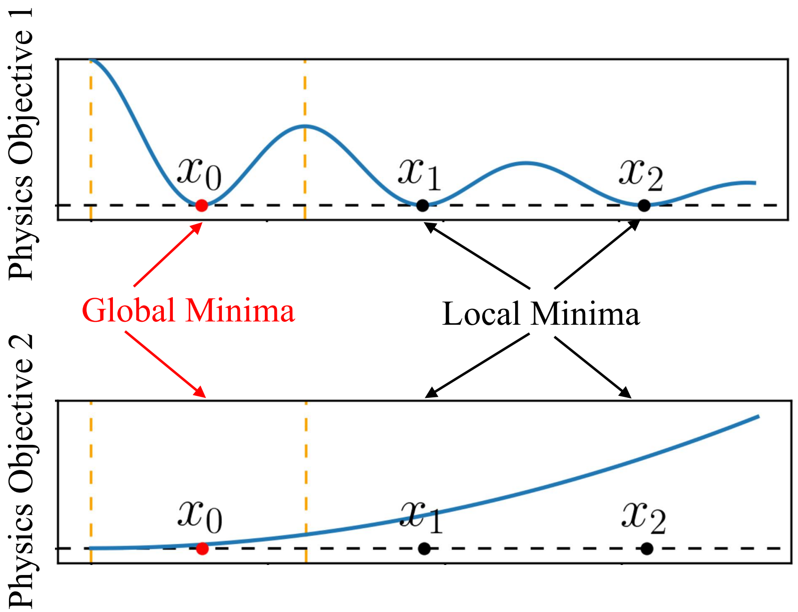
Energy Loss (E-Loss):

$$\lambda_E \exp(\hat{E})$$

How can we jointly incorporate S-Loss and E-loss in ANN learning to ensure generalizability?

Learning with Competing Physics Objectives

- S-Loss and E-Loss represent competing physics objectives
  - Can produce *conflicting* directions of gradient descent
  - Loss landscape of S-Loss is fraught with its own local minima
  - Non-trivial to balance data-loss, S-Loss, and E-Loss during learning

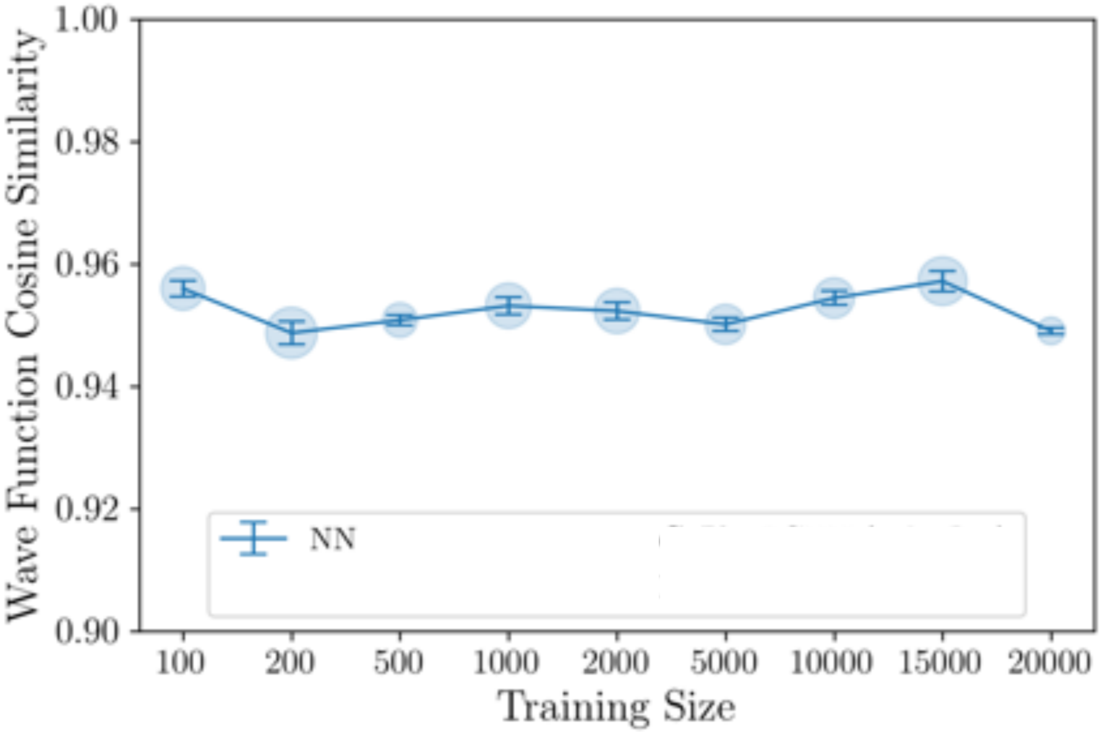


- **Key Question:** Can we adaptively tune the importance of S-Loss and E-Loss at different epochs (t) of ANN learning?

- **Solution:** PGNN with Competing Physics Objectives (CoPhy-PGNN)
  - *Annealing $\lambda_E(t)$:* Pay higher emphasis on E-Loss early on to avoid getting stuck at local minima of S-loss
  - *Cold-starting $\lambda_S(t)$:* Increase importance of S-Loss once we have zoomed in close to a generalizable solution

Elhamod et al., "CoPhy-PGNN: Learning Physics-guided Neural Networks withCompeting Loss Functions for Solving Eigenvalue Problems," Arxiv 2020

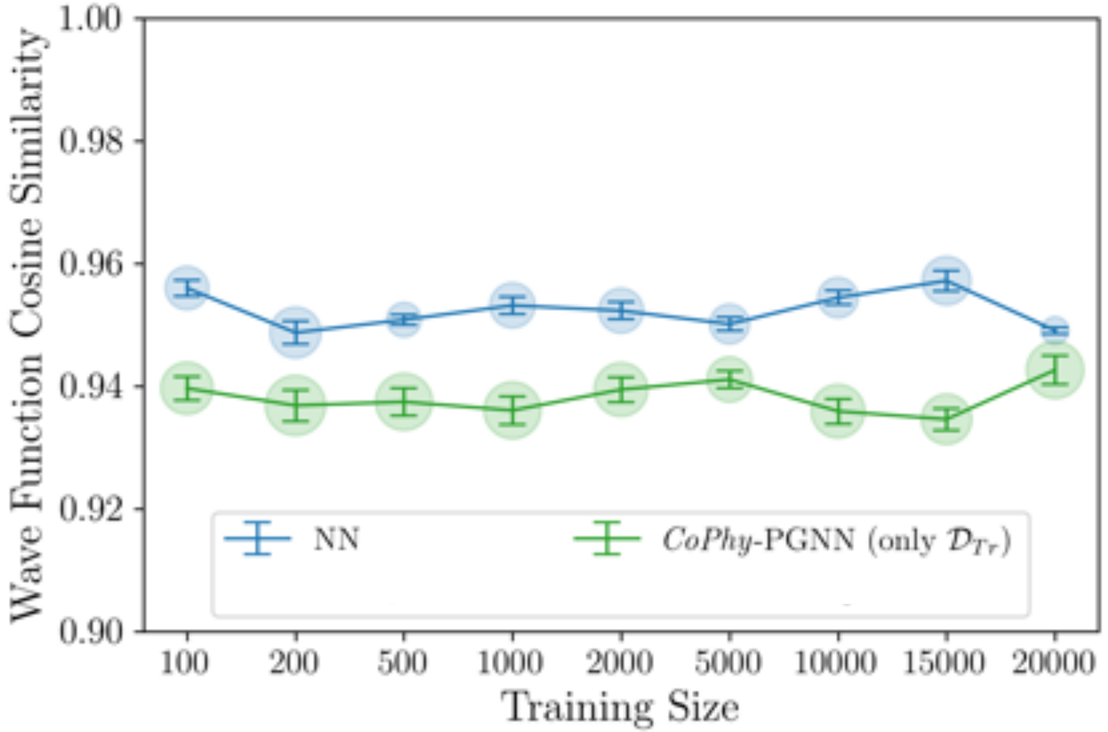
CoPhy-PGNN: Experimental Results



Evaluation Setup

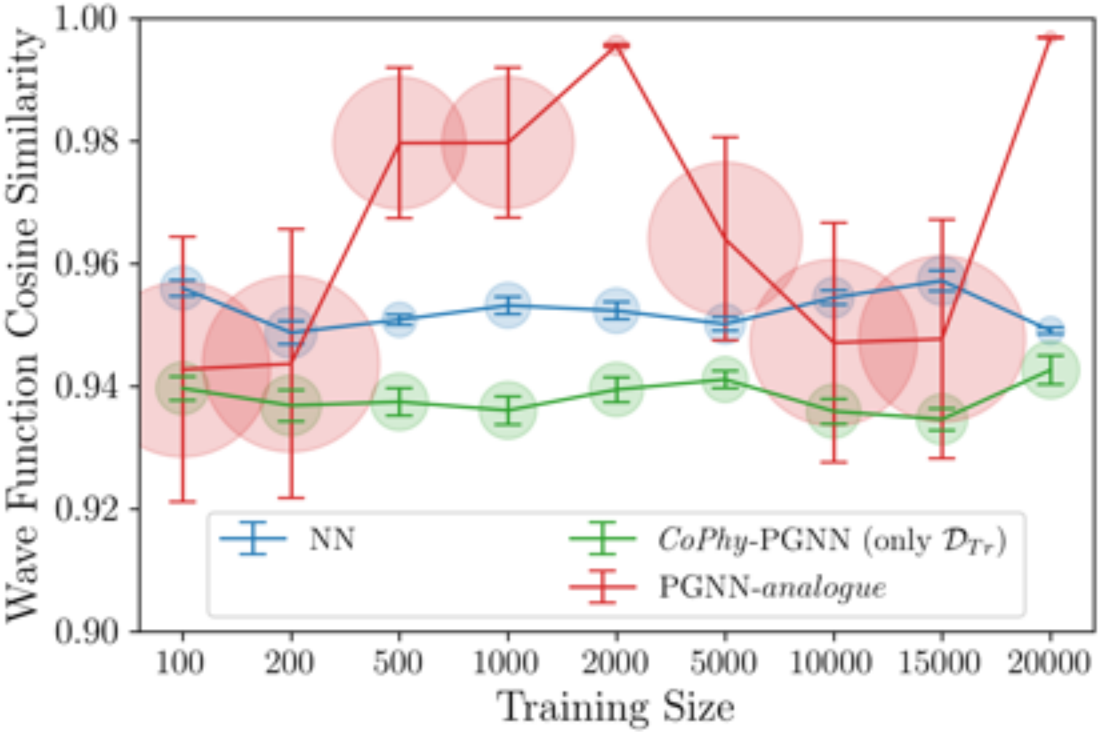
- 4-spin system
- Training data D_{Tr} sampled from Bx < 0.5 (ferromagnetic)
- Test data D_U sampled from Bx > 0.5 (ferromagnetic + paramagnetic)

CoPhy-PGNN: Experimental Results



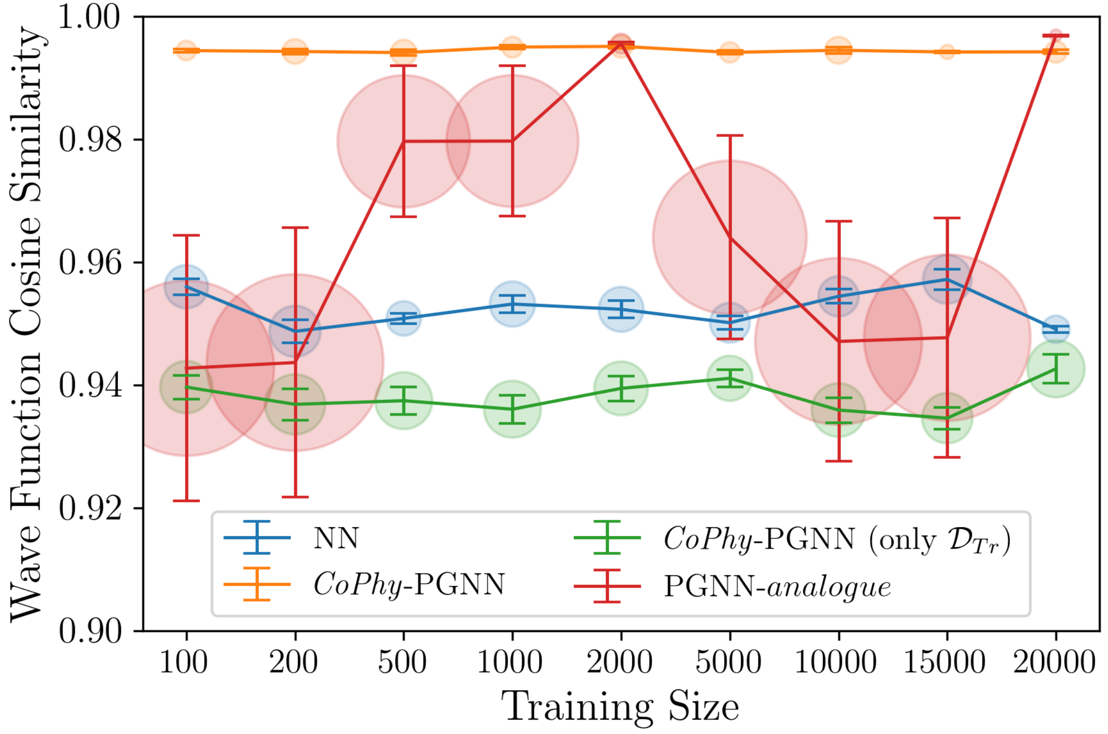
- Evaluating physics-guided loss on unlabeled samples from test scenarios is important

CoPhy-PGNN: Experimental Results



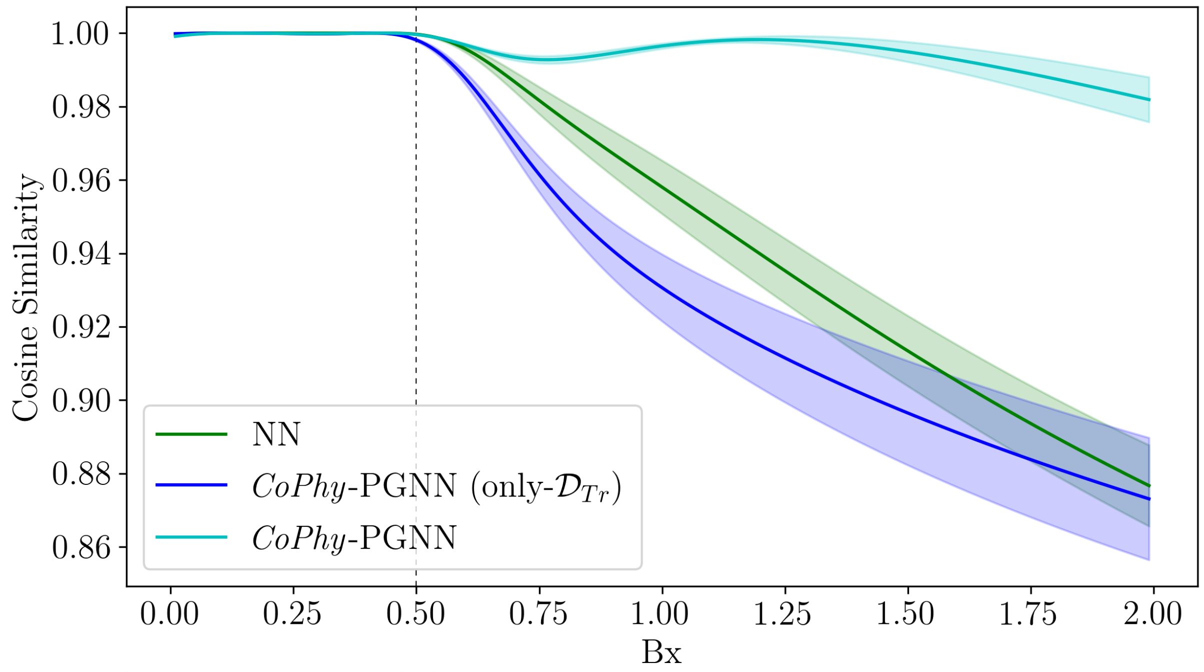
- Evaluating physics-guided loss on unlabeled samples from test scenarios is important
- Adding physics-guided loss with constant trade-off params can sometimes lead to spurious solutions

CoPhy-PGNN: Experimental Results



CoPhy-PGNN (Label-free) Cosine Similarity: 0.63

Evaluating generalizability on novel testing scenarios



Training (Ferromagnetic)

Testing (Ferromagnetic + Paramagnetic)

- Evaluating physics-guided loss on unlabeled samples from test scenarios is important
- Adding physics-guided loss with constant trade-off params can sometimes lead to spurious solutions
- CoPhy-PGNN achieves close-to-perfect performance even with 100 training examples

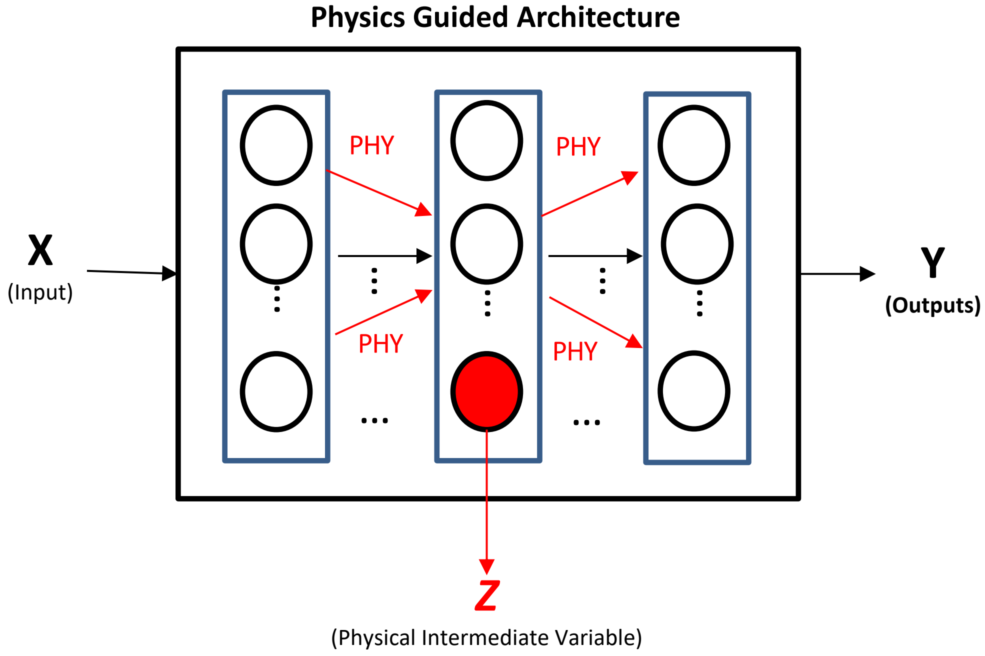
Case Study 3:

Science-guided Architecture for Lake Modeling

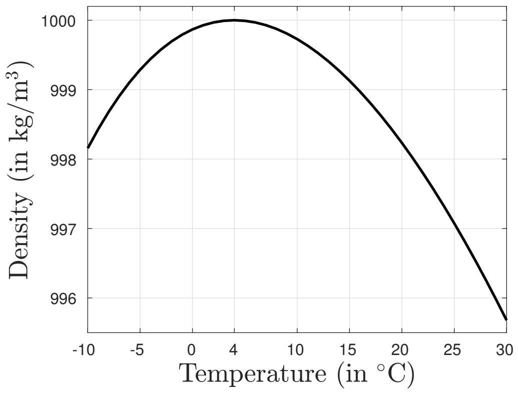
In Collaboration with: USGS, VT

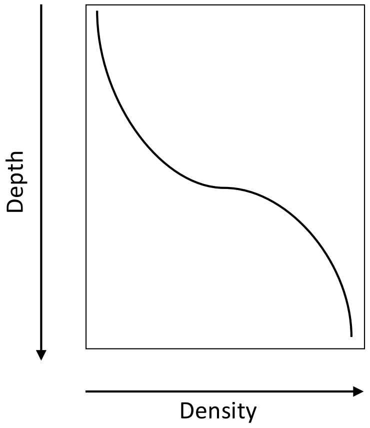
Biological Sciences
Forest Resources and Environmental Conservation

- Goal: "Bake in" physics in the architecture of neural networks
  - Ensure physical consistency during training as well as testing
    - In contrast to *science-guided learning* that only applies to training
  - Robust to minor perturbations in model weights
    - Critical for **uncertainty quantification** using MC Dropout



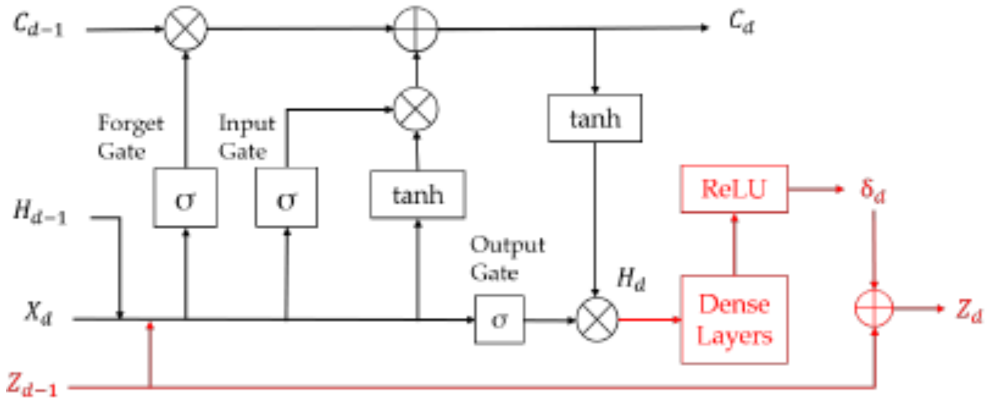
Can we hard-code density-depth physics in ANN models?



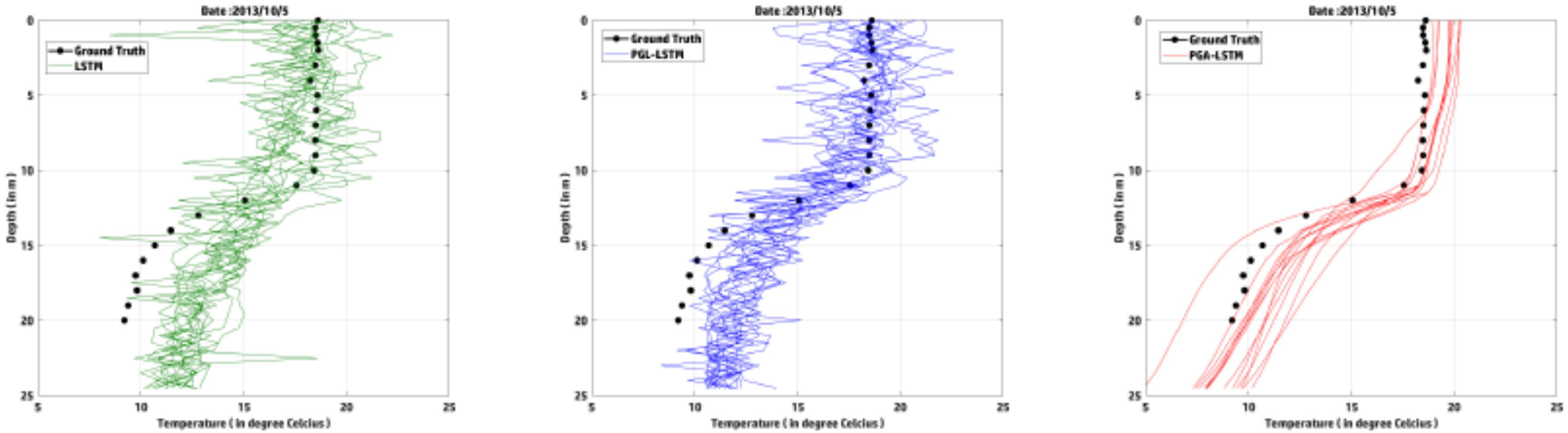


Physics-guided Architecture of LSTM models (PGA-LSTM)

- Use of physics-guided intermediate variables
  - Predict density as an intermediate variable in the ANN pathway
- Physics-guided connections among LSTM nodes
  - Monotonicity-preserving LSTMs ensures that density always increases with depth



Hard-coding physics in PGA-LSTM produces generalizable and physically consistent predictions, even after using MC dropout.



(a) LSTM Profiles (15 samples) (b) PGL-LSTM Profiles (15 samples) (physics-guided learning) (c) **PGA**-LSTM Profiles (15 samples)

Daw et al., "Physics-Guided Architecture (PGA) of Neural Networks for Quantifying Uncertainty in Lake Temperature Modeling," SDM 2020.

In Collaboration with:

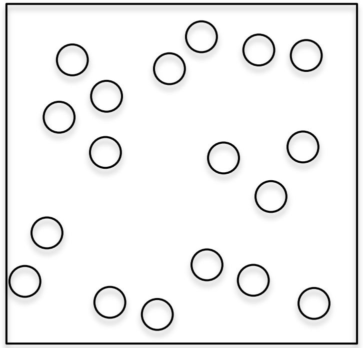
Computer Science,
Mech. Engineering

Case Study 4:
Hybrid-Science-ML Modeling for Fluid Dynamics

- Goal: Modeling drag force on particles suspended in a moving fluid

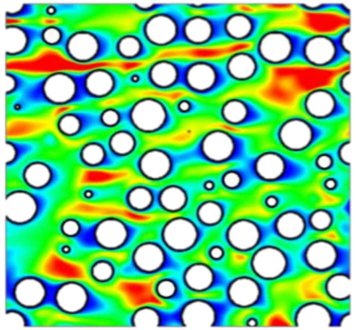
Applications: Gas separation, CO_2 capture, …

Alternate Modeling Approach



Coarse scale simulations of particle positions using Discrete Element Method (**DEM**)

Fast but incomplete



Particle Resolved Simulations (**PRS**) of pressure, velocity fields at *fine scale*

Accurate but expensive

Drag forces on particles

Inputs

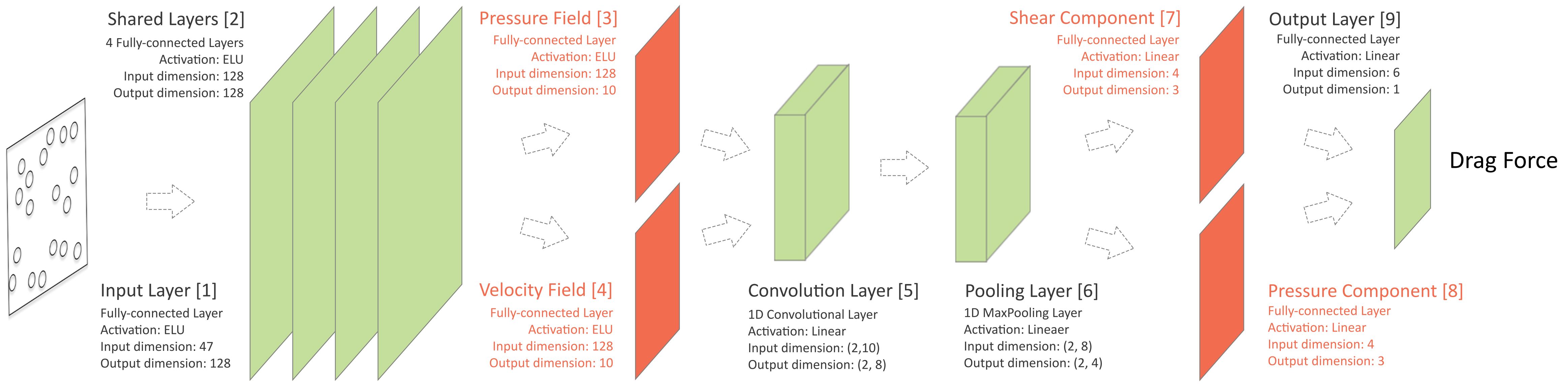
ANN

Outputs

Multi-scale Hybrid-Science-ML model

Can we ensure ANN produces pressure, velocity fields at intermediate layers?

Proposed Physics-guided Neural Net Architecture: PhyNet



Training internal layers to express physically meaningful variables results in better generalization with smaller number of samples

Muralidhar et al., "PhyNet: Physics Guided Neural Networks for Particle Drag Force Prediction in Assembly," SDM 2020.

Case Study 5:

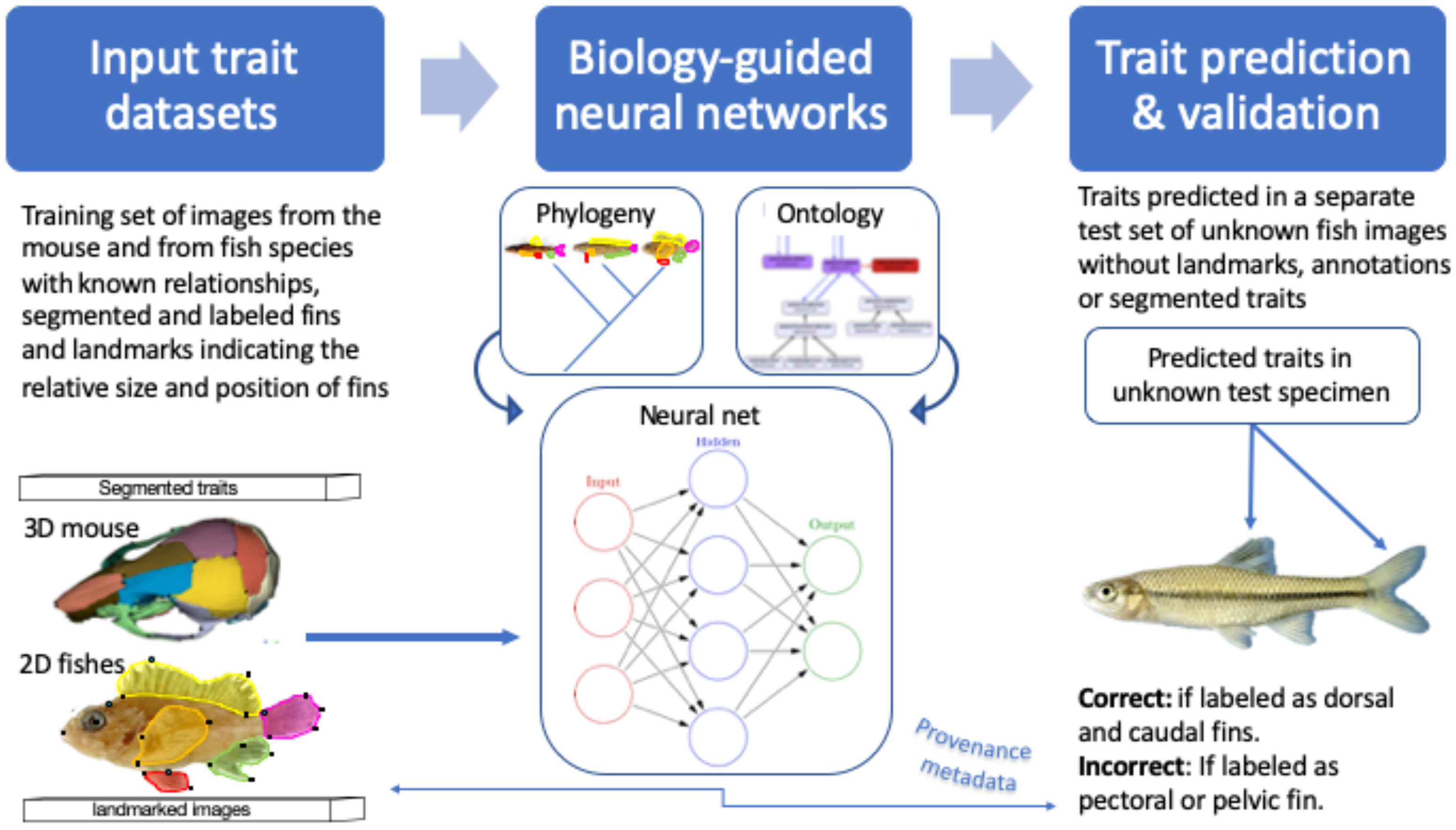
Biology-guided Neural Networks for Discovering Phenotypic Traits

Funded by:

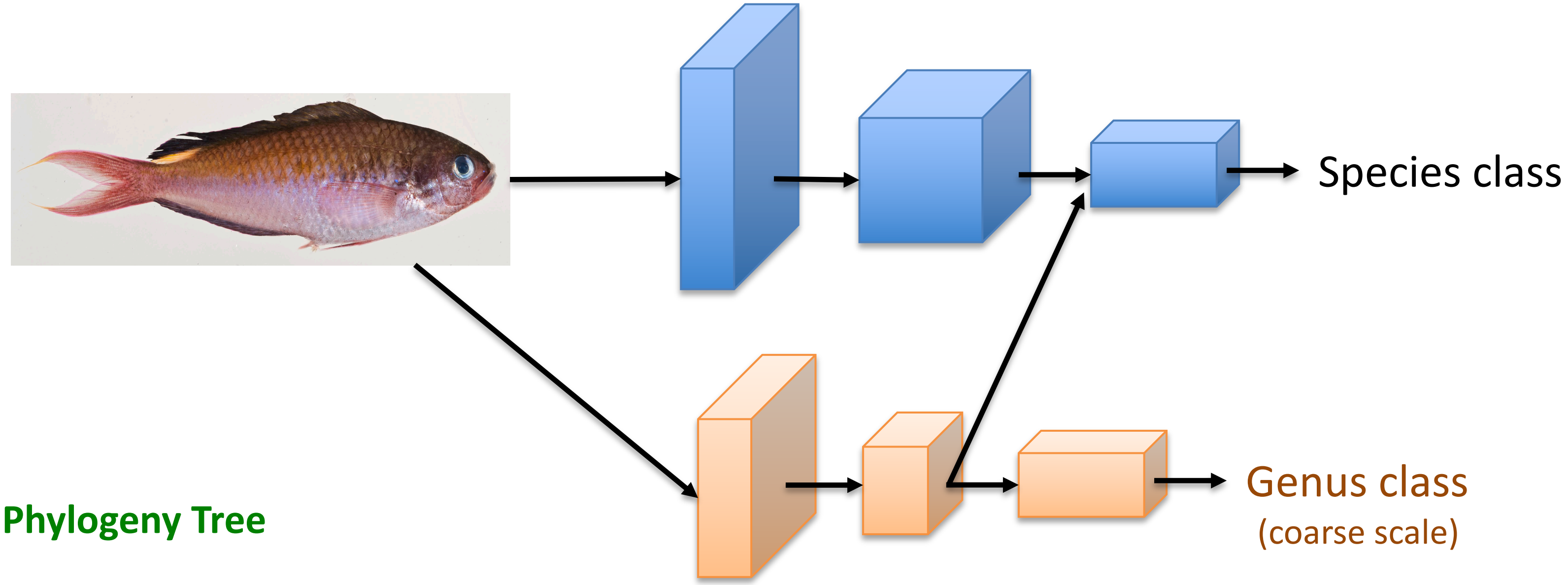
HDR Grant
1940247

Aims:

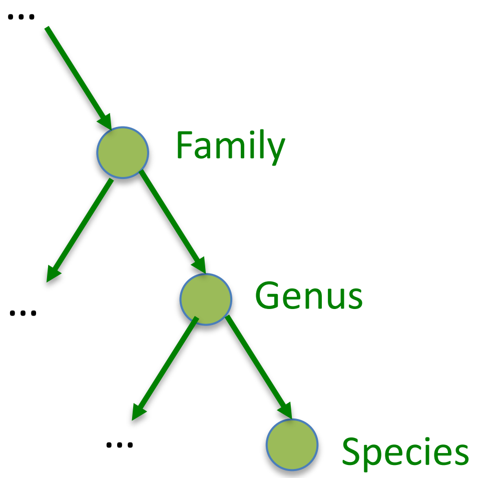
- Develop biology-guided neural networks (BGNN) for **species classification** and **trait segmentation**
- Apply BGNN to large volumes of unlabeled images to discover novel biological knowledge



Species Classification using Phylogeny Tree

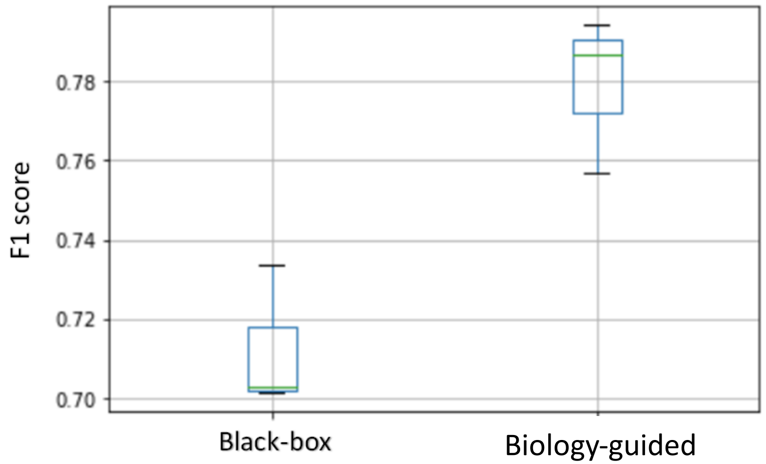


Phylogeny Tree

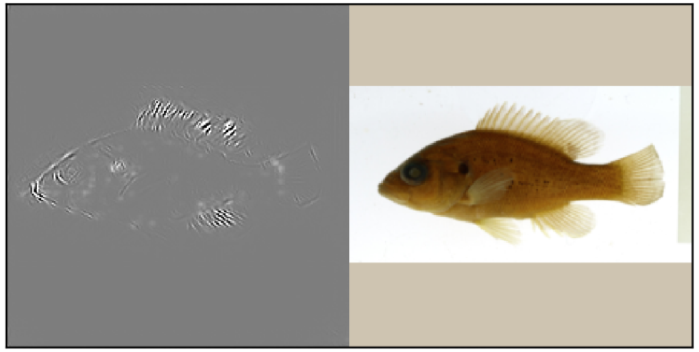


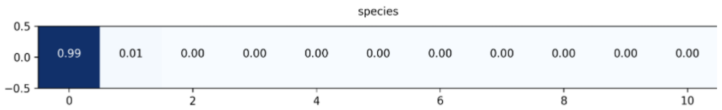
Key Idea: Species that belong to a common genus share similar ANN features

Using biology leads to better classification accuracy



Saliency Map





Robustness to Adversarial Attacks

- Adversarial occlusion procedure: incrementally occlude image patches with highest contribution to saliency maps

Saliency Map

Step 1

Step 2

Saliency Map:

Black-box	0.80	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
BGNN	0.88	0.10	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00

Step 1:

Black-box	0.49	0.04	0.04	0.04	0.04	0.08	0.04	0.10	0.04	0.04	0.07
BGNN	0.87	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

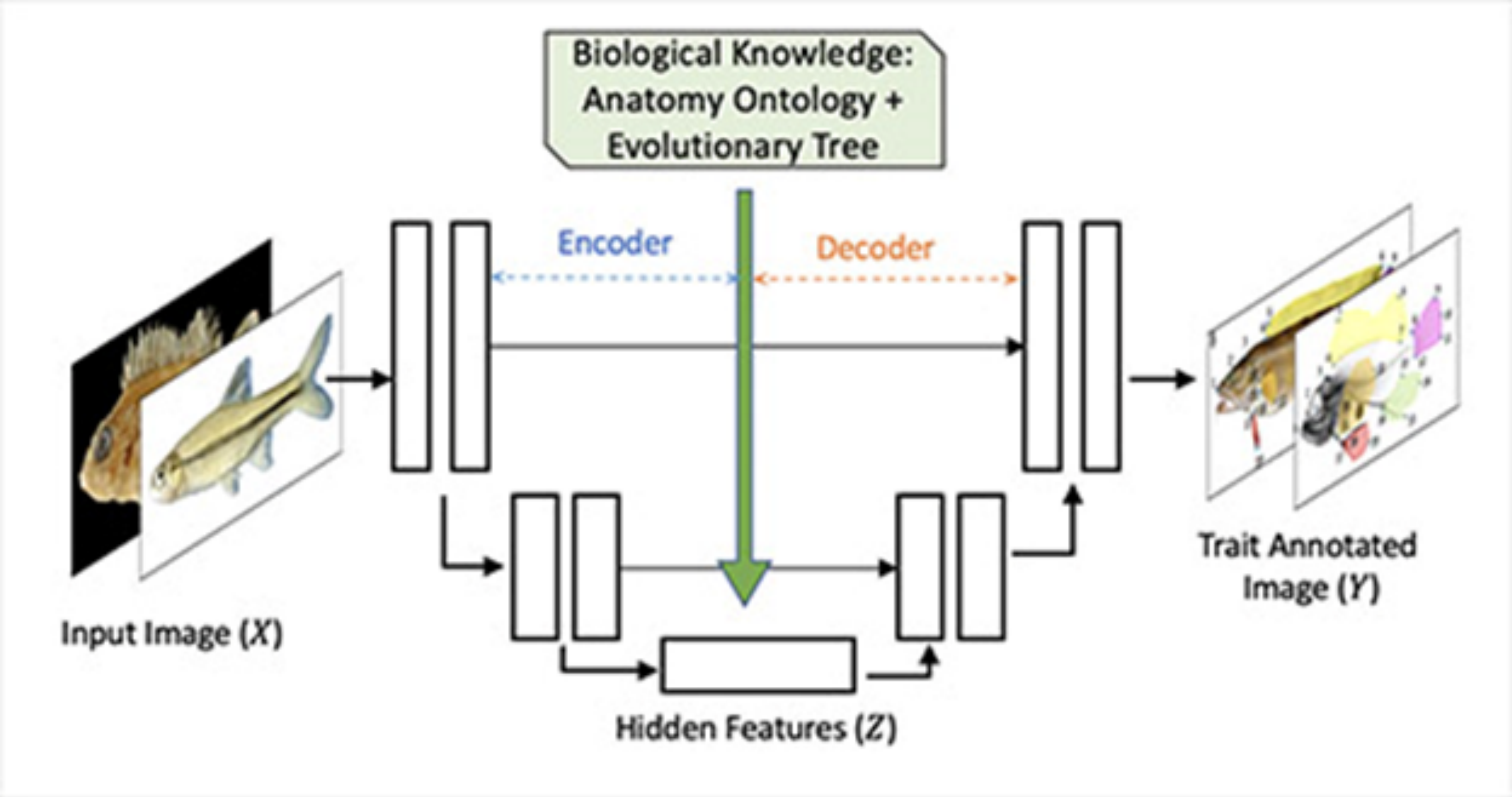
Step 2:

Black-box	0.09	0.06	0.20	0.07	0.06	0.12	0.07	0.10	0.06	0.07	0.09
BGNN	0.53	0.12	0.08	0.14	0.02	0.04	0.03	0.01	0.00	0.01	0.02

By forcing ANN features to comply with biological knowledge, we can be more robust to adversarial occlusions

Trait Segmentation using Anatomy Ontology (Ongoing)

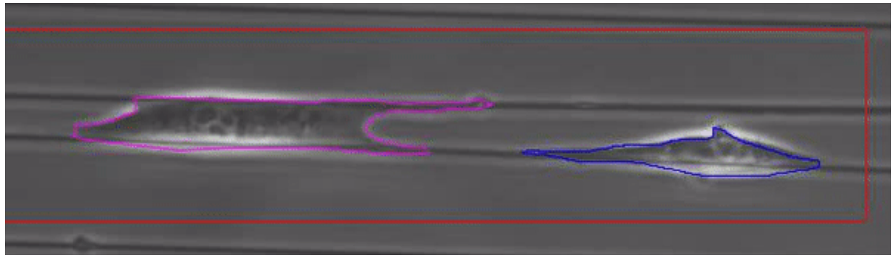
- Verify if the predicted traits from the neural network *violate* known ontological relationships and minimize such violations during training (as additional loss functions in objective function)



Other Ongoing Projects in SGML

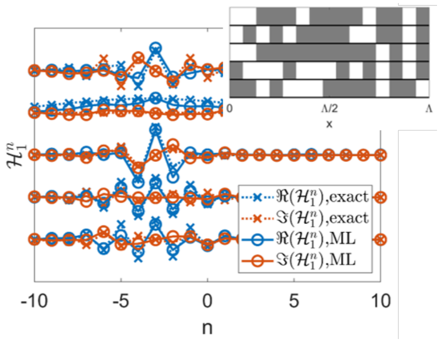
Physics-guided Tracking of Living Cells in Mechanobiology

Collaborators: Mechanical Engineering at VT

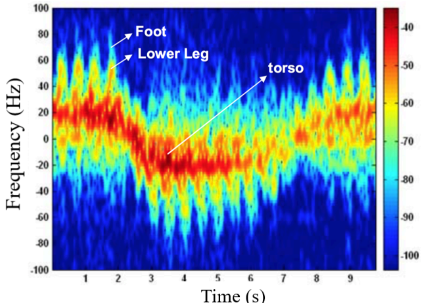


Physics-guided Learning for Quantum Mechanics, Optics, and Radar Physics

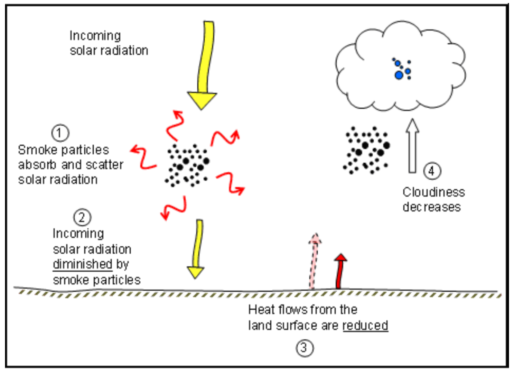
Collaborators: Ohio State U., U. Mass. Lowell, SUNY Binghamton



EAGER Grant
2026710



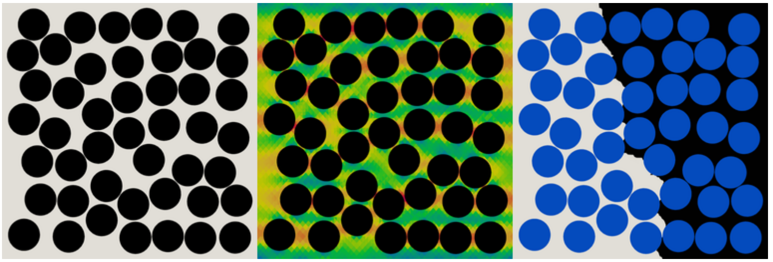
Inverse Modeling of Aerosol Properties from Spectroscopy Data



Collaborators: ECE at VT

Crack Prediction in Composites using Physics-guided ANN Architecture

Collaborators: Civil and Environmental Engineering at VT



Summary

- Research Themes in SGML
  - **Diverse forms of scientific knowledge**
    - First-principle equations, Model simulations, Ontologies, …
  - **Diverse ways of integrating scientific knowledge with ML**
    - Science-guided Learning
    - Science-guided Architecture
    - Hybrid-science-ML modeling
  - **Diverse scientific applications**
    - Lake modeling, Quantum mechanics, Fluid dynamics, Biology (ichthyology)

- Upcoming Activities in SGML:
  - AAAI Spring Symposium Series on "Combining Artificial Intelligence and Machine Learning with Physical Sciences", March 22-24, 2021, https://sites.google.com/view/aaai-mlps
  - Editing Book on ""Science-guided Machine Learning: Emerging Trends in Combining Scientific Knowledge with Data-driven Methods," CRC Press, to appear in Aug 2021

Acknowledgments

- Funding Agencies:
  - NSF # 2026710, # 1940247

- Collaborators:
  - VT: Cayelan Carey, Quinn Thomas, Danesh Tafti, Naren Ramakrishnan, Amrinder Nain, Sohan Kale, Elena Lind, Maryam Shakiba
  - Vipin Kumar (UMN), Jordan Read and Alison Appling (USGS), Paul Hanson (U Wisc.), Wei-Cheng Lee (SUNY Binghamton), Viktor Podolskiy (U. Mass Lowell), Anish Arora (OSU), Paula Mabee (Battellee), Hank Bart (Tulane U.), Jane Greenberg (Drexel U.), Murat Maga (UW)

- PGML Lab Members
  - PhD Students: Jie Bu, Arka Daw, Mohannad Elhamod, Md. Al Maruf, Snehal More
  - MS Students: Ioannis Papakis, Prathamesh Mandke, Arya Shahadi
  - Alumni: Sandhya Bhaskar (MS), Zheng Li (MS)

karpatne@vt.edu
https://people.cs.vt.edu/karpatne/

PGML Lab
Physics-guided Machine Learning

Publications

- A. Karpatne, G. Atluri, J. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar, “Theory-guided data science: A new paradigm for scientific discovery,” In IEEE Transactions on Knowledge and Data Engineering, 29(10): 2318-2331, 2017. DOI: 10.1109/TKDE.2017.2720168.
- J. Read, X. Jia, J. Willard, A. Appling, J. A. Zwart, S. K. Oliver, A. Karpatne, G. J. Hansen, P. C. Hanson, W. Watkins, M. Steinbach, and V. Kumar, “Process-guided deep learning predictions of lake water temperature,” Water Resources Research 55, no. 11 (2019): 9173-9190.
- A. Karpatne, W. Watkins, J. Read, and V. Kumar, “Physics-guided Neural Networks (PGNN): An Application in Lake Temperature Modeling,” arXiv: 1710.11431, 2017.
- X. Jia, J. Willard, A. Karpatne, J. Read, J. Zwart, M. Steinbach, and V. Kumar, “Physics Guided RNNs for Modeling Dynamical Systems: A Case Study in Simulating Lake Temperature Profiles,” In Proceedings of the 2019 SIAM International Conference on Data Mining, 558-566, 2019. DOI: 10.1137/1.9781611975673.
- A. Daw, Q. Thomas, C. Carey, J. Read, A. Appling, and A. Karpatne, “Physics-Guided Architecture (PGA) of Neural Networks for Quantifying Uncertainty in Lake Temperature Modeling,” SIAM SDM, 532-540, 2020. DOI: https://doi.org/10.1137/1.9781611976236.60.
- M. Elhamod, J. Bu, C. Singh, M. Redell, A. Ghosh, V. Podolskiy, W.C. Lee, and A. Karpatne, “CoPhy-PGNN: Learning Physics-guided Neural Networks withCompeting Loss Functions for Solving Eigenvalue Problems,” arXiv:2007.01420, 2020.
- N. Muralidhar, J. Bu, Z. Cao, L. He, N. Ramakrishnan, D. Tafti, and A. Karpatne, “PhyNet: Physics Guided Neural Networks for Particle Drag Force Prediction in Assembly,” SIAM SDM, 559-567, 2020. DOI: https://doi.org/10.1137/1.9781611976236.63.
- J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, “Integrating physics-based modeling with machine learning: A survey,” arXiv preprint arXiv:2003.04919, 2020.

Thank you!!