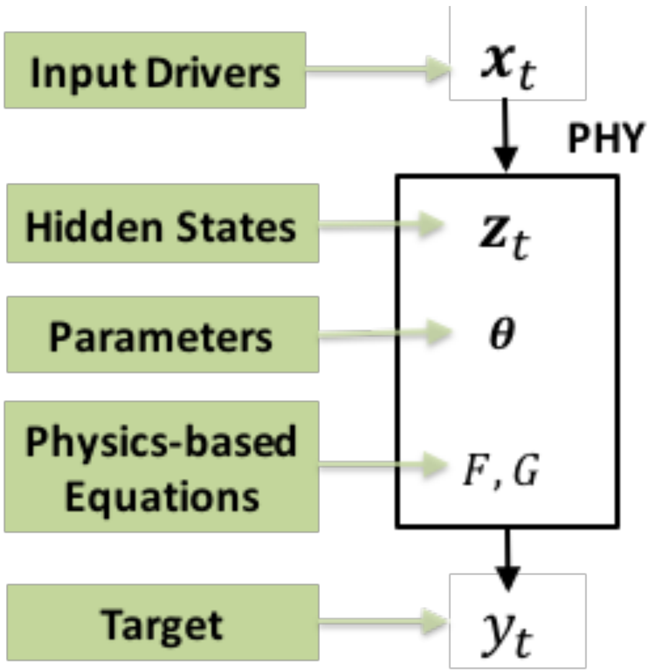
Physics-Guided Machine Learning :
A New Framework for Accelerating Scientific Discovery

Vipin Kumar
University of Minnesota
kumar001@umn.edu
www.cs.umn.edu/~kumar

Physics-based Models of Dynamical Systems

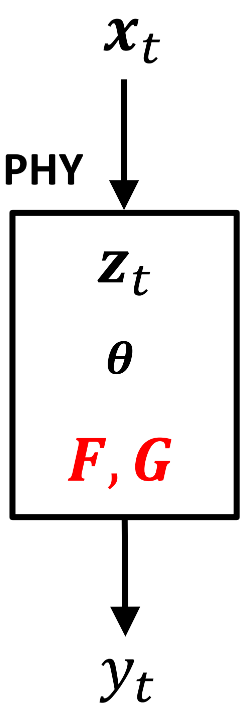
- Relationships b/w input & output variables governed by physics-based partial differential equations (PDEs)



Examples from Hydrology, Limnology, Fluid Dynamics, …

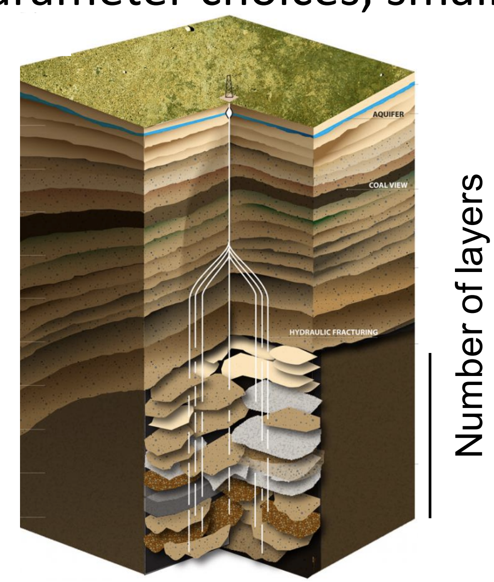
Input	Output	Parameters
Rainfall, topography, land use, river width	River discharge	Soil conductivity, channel flow
Solar radiation, air temp, wind speed	Lake quality	Lake bathymetry, water clarity
Pressure, strain rate tensor, kinetic energy	Velocity field, lift, drag	Reynolds stress, flow geometry

Limitations of Physics-based Models

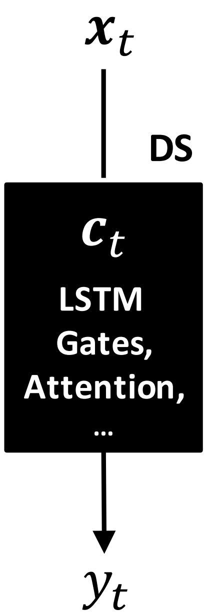


- Incomplete or missing physics ($\boldsymbol{F}, \boldsymbol{G}$)
  - Physics-based models often use approximate forms to meet "scale-accuracy" trade-off
  - Results in *inherent model bias*

- Unknown parameters ($\boldsymbol{\theta}$) need to be "calibrated"
  - *Computationally Expensive*
  - *Easy to overfit*: large number of parameter choices, small number of samples

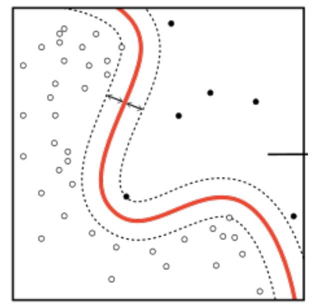


"Black-box" Data Science Models

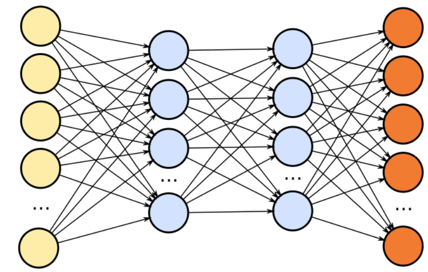


An alternative to modeling dynamical systems?

Choice of model family not governed by physics



Support Vector Machine



Deep Learning

- Hugely successful in commercial applications

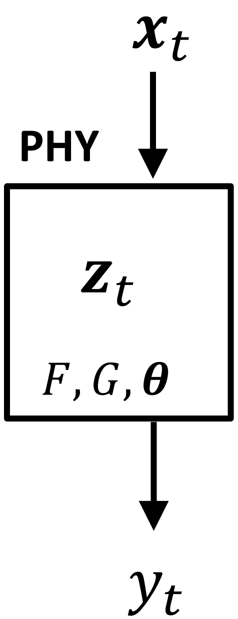
- But disappointing results in scientific domains!
  - Require lots of data
  - Can generate physically inconsistent results
  - Unable to generalize to unseen scenarios
  - Unable to provide valuable physical insights

Science

The Parable of Google Flu: Traps in Big Data Analysis

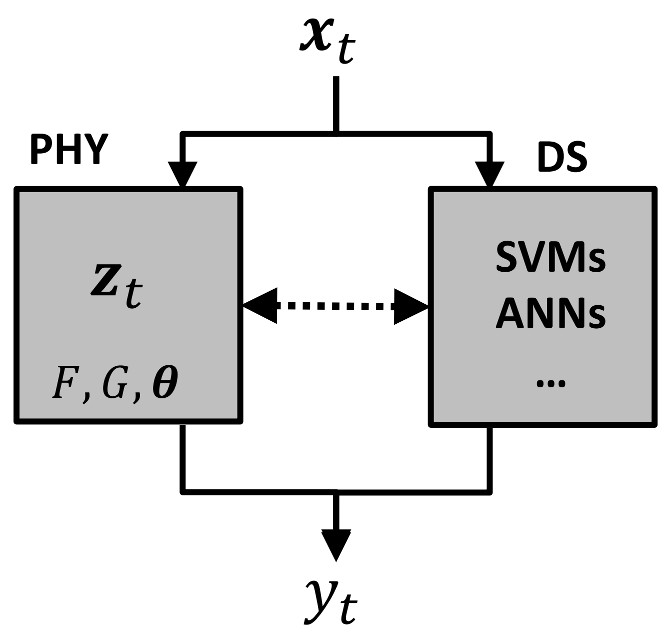
Hybrid-Physics-Data (HPD) Modeling:

A Paradigm Shift in Data Science



Physics-based Models

Data Science Models



HPD Models

Contain knowledge gaps in describing certain processes

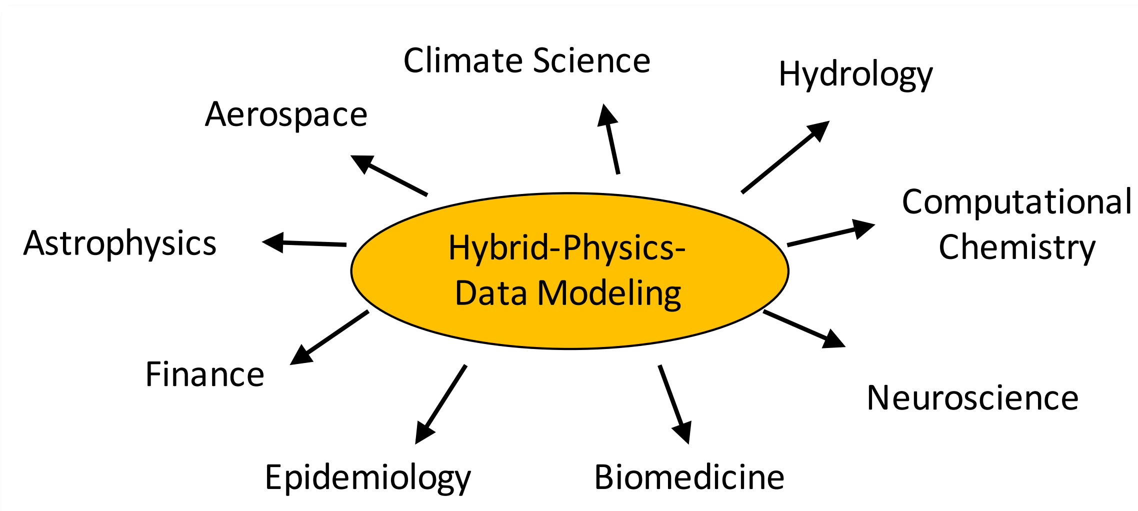
Require large number of representative samples

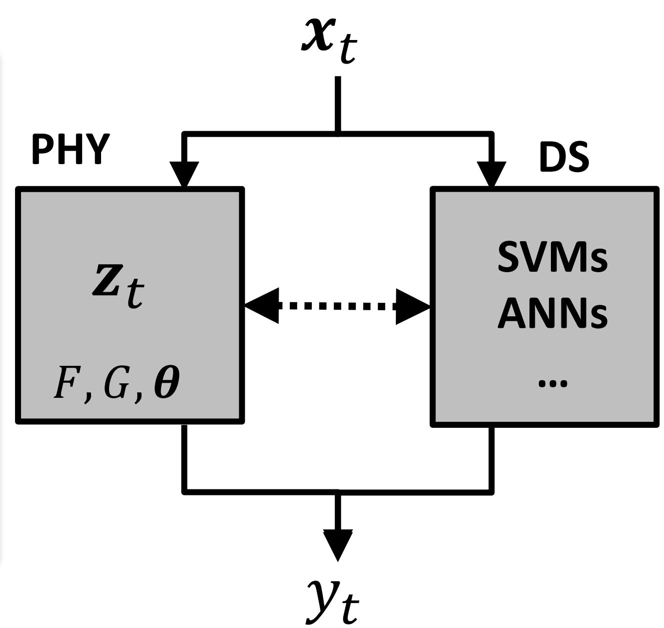
Overcome complementary weaknesses of both by combining PHY and DS in novel ways

Karpatne et al. "**Theory-guided data science: A new paradigm for scientific discovery**," TKDE 2017

Hybrid-Physics-Data (HPD) Modeling:

A Paradigm Shift in Data Science





Physics-based Models

Contain knowledge gaps in describing certain processes

Data Science Models

Require large number of representative samples

HPD Models

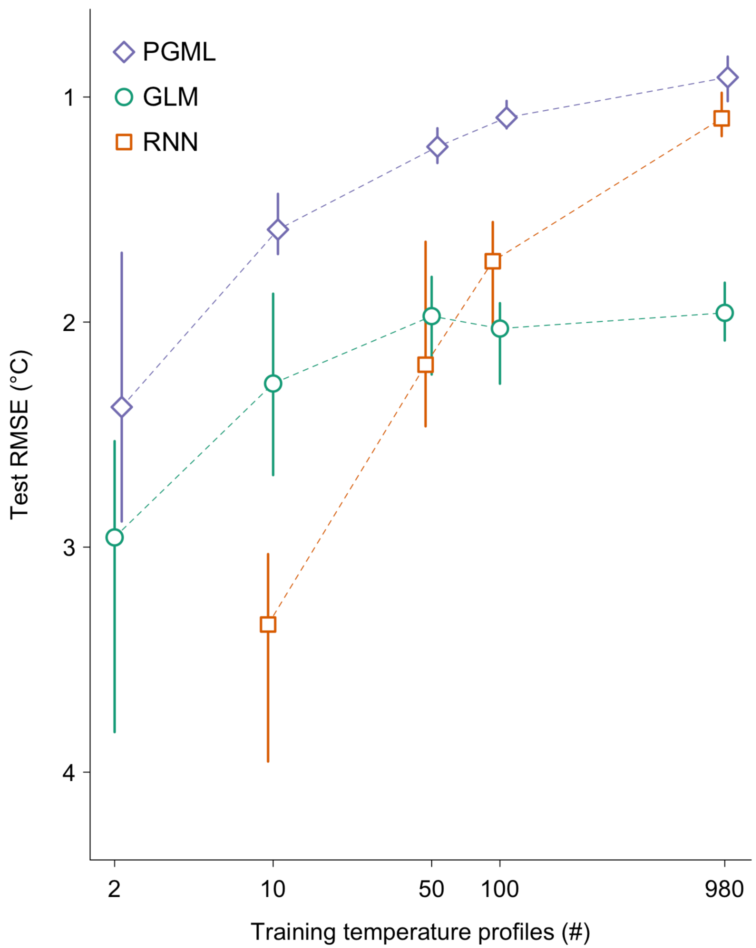
Overcome complementary weaknesses of both by combining PHY and DS in novel ways

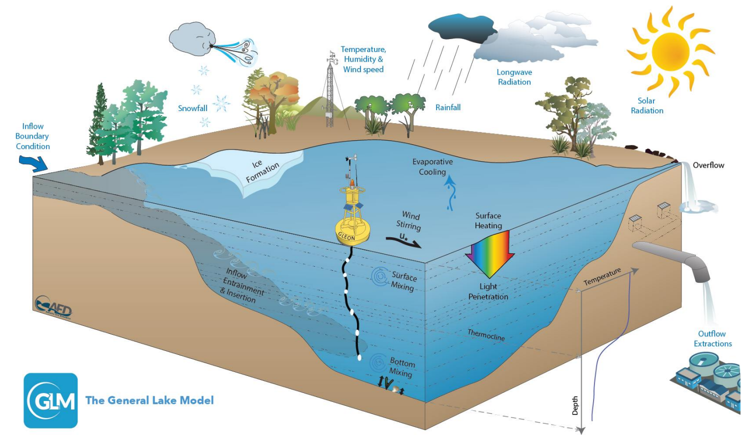
Karpatne et al. "**Theory-guided data science: A new paradigm for scientific discovery**," TKDE 2017

Questions

- Can machine learning (ML) models outperform physics based models given sufficient data?
- Can ML models leverage physics
  - to produce results that are physically consistent?
  - to learn with limited observation data?
  - To generalize to unseen scenarios
- Can physics guided ML models provide novel insights?

PGML for Modeling Lake Water Temperature:
Performance Under Data Sparse Conditions





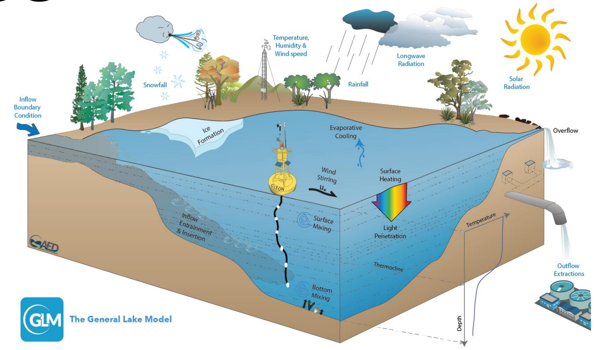
GLM: State of the Art physics based model used by USGS

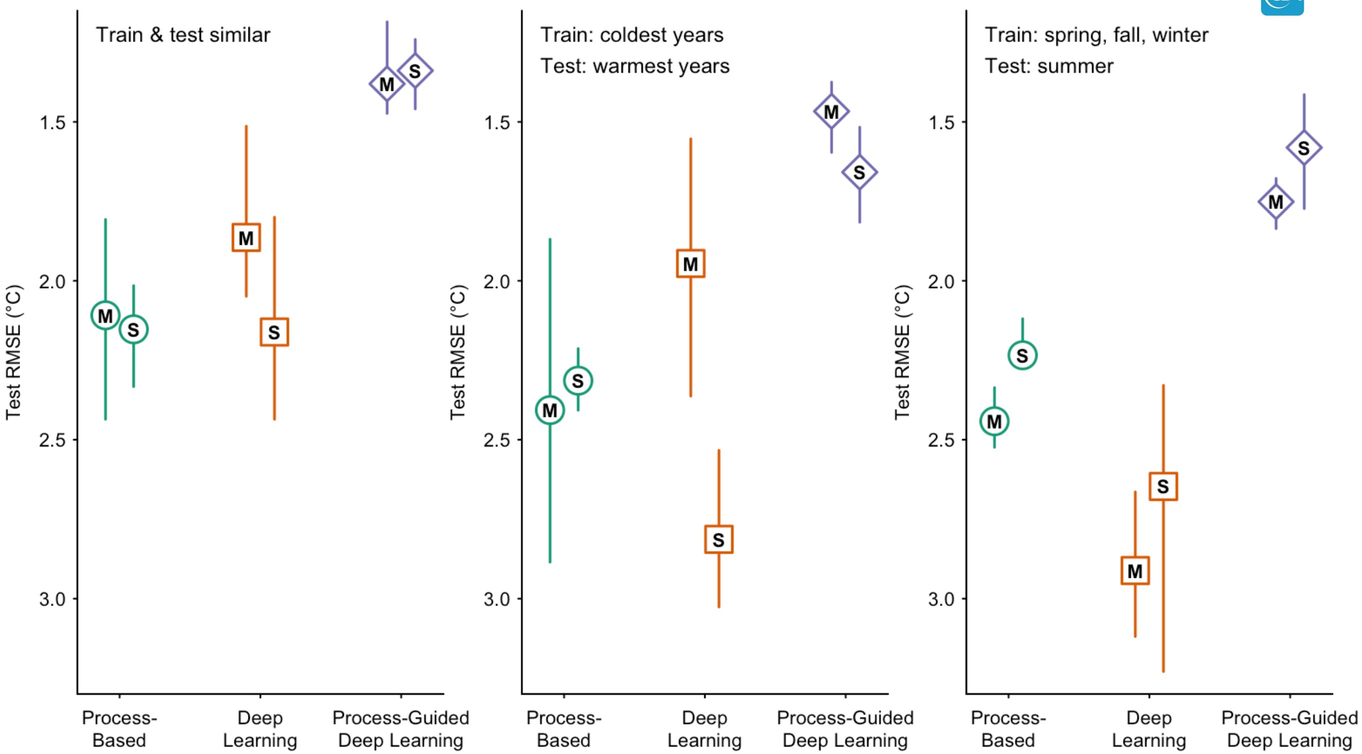
RNN: A black-box machine learning model that can incorporate time

PGML: A machine learning framework that leverages physics

Joint work with Jordan Read (USGS)

PGML for Modeling Lake Water Temperature:
Performance in Unseen Scenarios





S – Sparkling Lake
M – Lake Mendota

Joint work with Jordan Read (USGS)

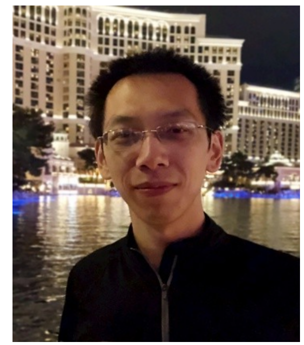
Acknowledgements

- Collaborators and Team Members

Anuj Karpatne
Virginia Tech

Jordan Read
USGS

Jacob Zwart
USGS



Xiaowei Jia
UMN

Jared Willard
UMN

Alison Appling (USGS), Samantha Oliver (USGS), Gretchen Hansen (UMN), Paul Hanson (U Wisconsin), William Watkins (USGS)