Paul Hanson
University of Wisconsin

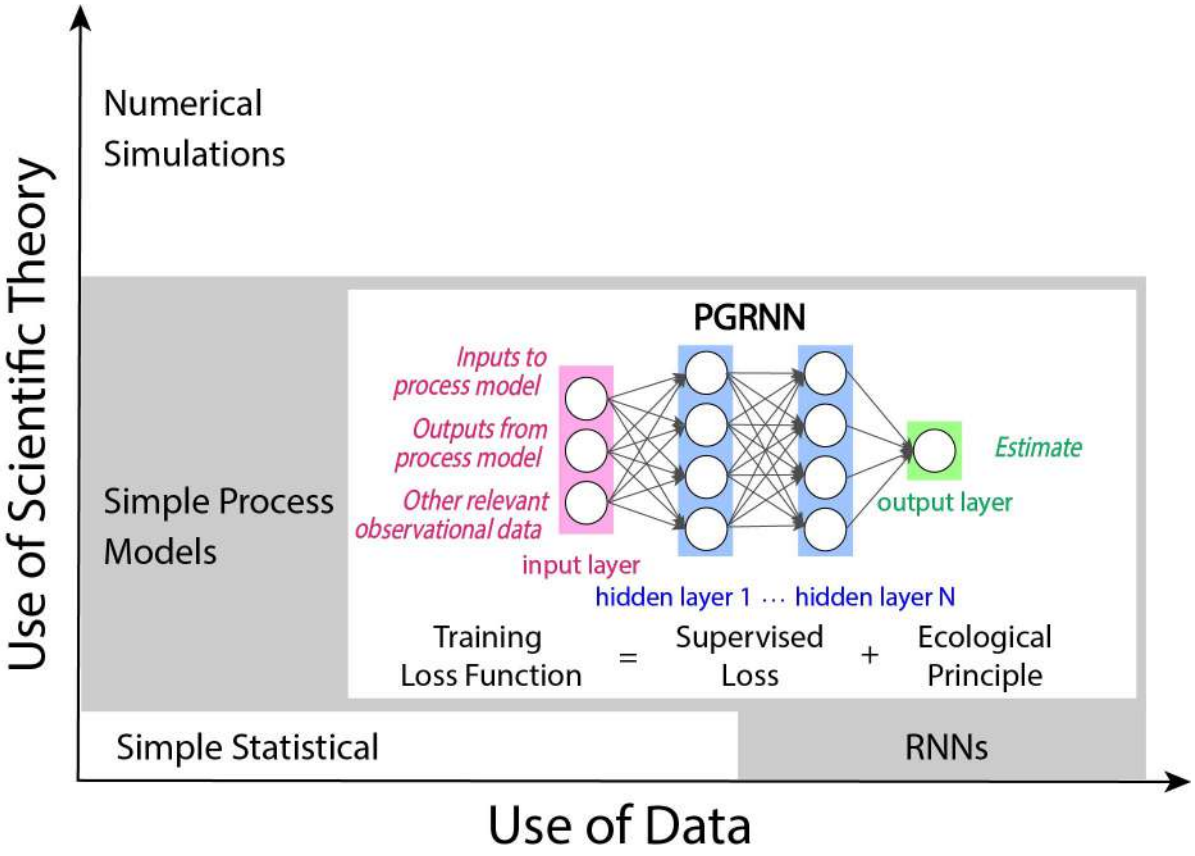
FAIR, AI, and Ecology

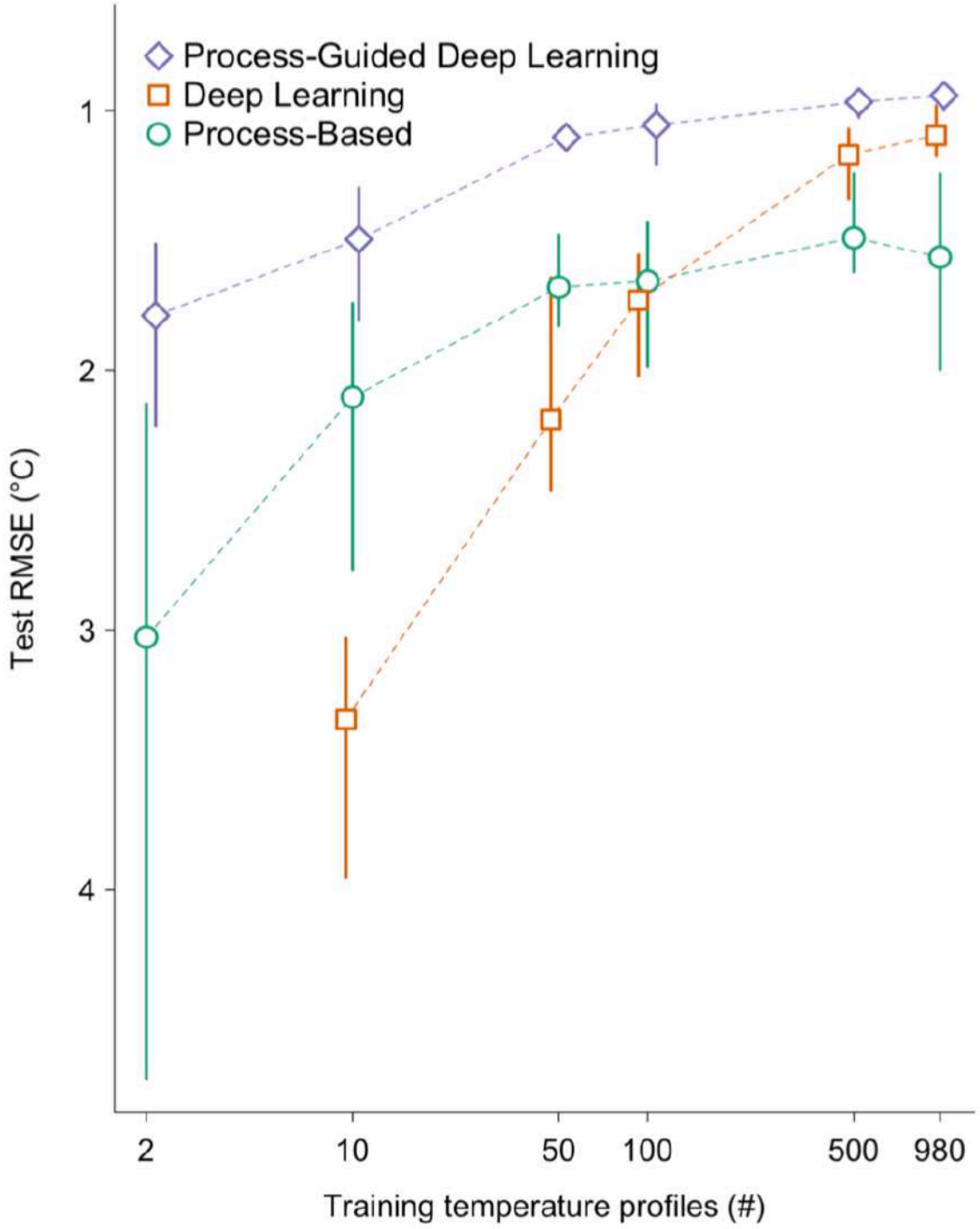
Paul's background

- Ecosystem scientist who has always worked at the science-technology interface
- Co-chair of GLEON, along with Kathie Weathers, for nearly 10 years
- Co-PI of the Environmental Data Initiative (EDI)
- Have worked with Beth Plale and her student on symbolic representation of time series data
- Have worked with Miron Livny on HTCondor
- Have worked with Prof. Yu Hen Hu on neural networks and support vector machines
- Currently funded projects
  - GRAPLEr with Renato Figueiredo (NSF++) and his team (UF)
  - Entity matching research (NSF CISE) with AnHai Doan (UW)
  - Theory guided data science (NSF HDR) with Vipin Kumar (UM)
  - Environmental Data Initiative (NSF ABI) with Corinna Gries (UW)
  - UW Data Sciences initiative, working with CS and the USGS
  - FAIR data for lake observatories of the future workshop (NSF) with Kathie Weathers (Cary Institute), Cayelan Carey (Virg Tech), and Renato Figueiredo (UF)

Three stories

1. When your phone rings answer it
    1. Shelly Stall, 2017, AGU and FAIR data
    2. All the vectors pointed in the same direction
2. FAIR does not mean easy
    1. Entity matching project with AnHai Doan
    2. USGS Water Quality portal
3. Sometimes, good can come from a hurricane
    1. It takes a year or more to learn and calibrate GLM
    2. We have done amazing work in GRAPLEr to advance that work
    3. Yolanda Gil
    4. Vipin Kumar


Use of Scientific Theory
Numerical Simulations
Simple Process Models
PGRNN
Inputs to process model
Outputs from process model
Other relevant observational data
input layer
hidden layer 1 ⋯ hidden layer N
output layer
Estimate
Training Loss Function
=
Supervised Loss
+
Ecological Principle
Simple Statistical
RNNs
Use of Data


Process-Guided Deep Learning
Deep Learning
Process-Based
Test RMSE (°C)
1
2
3
4
2
10
50
100
500
980
Training temperature profiles (#)

Beth's Questions

- Where do you see the FAIR principles reducing the research burden for international collaboratories (such as GLEON)
  - Access to and reuse of data is a major bottleneck
  - Great questions can die when data are too difficult
  - FAIR principles are inspiring standardization and innovation
- What AI applications motivate your work
  - Lake science (limnology) and water quality
  - Using what we know about a few well-studied lakes to make predictions for thousands of lakes with sparse data
  - Anoxia: depletion of oxygen in lakes that support cold water fisheries
- What are the unique infrastructure or sociotechnical challenges in AI research involving international partners
  - Compared to other fields, ecology has relatively small data sets that are highly heterogeneous
  - Differing views on the value and reuse of data
  - For some scientists, data are the main asset they bring to the table
  - Technologies that take address each letter in FAIR will help
  - Understanding "the way the world works" is usually the goal
- What are the unique training challenges and needs for AI research? For FAIR data?
  - Very few in my field know how to use these techniques
  - Requires partnering with experts