Standing on the shoulders of giants, a perspective from and for modeling natural systems

Mary C Hill

University of Kansas

Recorded for a 2022 Workshop at the University of Alabama

Challenges

- Can model analysis methods be built for complicated systems and noisy, insufficient data to explore "Why?"
- Model results are important but often isolated in development and results from the people and communities involved
  - Need, for example, social engagement and economic consequences

Methods I am finding useful

- Machine Learning/Artificial Intelligence – to explore how climate change might affect renewable energy and crop production
  - use extensive data sets and process-based DSSAT crop model results

It is not either/or. It is standing on the shoulders of giants to reach further.

- Giants from process-based modeling, agent-based modeling, machine learning…
- Address problems that process-based models alone really are not able to address
- This is a fun and exciting challenge that needs those with knowledge of process-based modeling!

What I learned from using process-based models

- Programming matters. Speed, accuracy.
- Transparency matters. Documentation, model analysis methods.
- Process-based models can be used to explore "Why?"
- Process-based models incorporate the understanding of equations and statistics programmed into them. The models ideally can be used to explore the consequences of these equations and statistics.
- Natural systems are complicated, data are insufficient and noisy, and precise results are desired but unattainable
- It is easier to investigate "Why?" with simple models, but people trust complicated models that match the conditions with which they are most familiar, like water levels in their area. This dichotomy is fundamental and difficult to address.

Robust Inverse Framework using Knowledge-guided Self-Supervised Learning: An application to Hydrology

Rahul Ghosh
University of Minnesota - Twin Cities
Minneapolis, Minnesota, USA
ghosh128@umn.edu

Arvind Renganathan
University of Minnesota - Twin Cities
Minneapolis, Minnesota, USA
renga016@umn.edu

Kshitij Tayal
University of Minnesota - Twin Cities
Minneapolis, Minnesota, USA
tayal007@umn.edu

Xiang Li
University of Minnesota - Twin Cities
Minneapolis, Minnesota, USA
lixx5000@umn.edu

Ankush Khandelwal
University of Minnesota - Twin Cities
Minneapolis, Minnesota, USA
khand035@umn.edu

Xiaowei Jia
University of Pittsburgh
Pittsburgh, Pennsylvania,
xiaowei@pitt.edu

Christopher Duffy
Penn State University
State College, Pennsylvania, USA
cxd11@psu.edu

John Nieber
University of Minnesota - Twin Cities
Minneapolis, Minnesota, USA
nieber@umn.edu

Vipin Kumar
University of Minnesota - Tw
Minneapolis, Minnesota,
kumar001@umn.edu

ABSTRACT

Machine Learning is beginning to provide state-of-the-art performance in a range of environmental applications such as streamflow prediction in a hydrologic basin. However, building accurate broad-scale models for streamflow remains challenging in practice due to the variability in the dominant hydrologic processes, which are best captured by sets of process-related basin characteristics. Existing basin characteristics suffer from noise and uncertainty, among many other things, which adversely impact model performance. To tackle the above challenges, in this paper, we propose a novel Knowledge-guided Self-Supervised Learning (KGSSL) inverse framework to extract system characteristics from driver and

KEYWORDS

Self-supervised Learning, Inverse Modeling, Forward

ACM Reference Format:
Rahul Ghosh, Arvind Renganathan, Kshitij Tayal, Xiang Li, delwal, Xiaowei Jia, Christopher Duffy, John Nieber, and Vipi Robust Inverse Framework using Knowledge-guided Self-Sup ing: An application to Hydrology. In *Proceedings of the 28th Conference on Knowledge Discovery and Data Mining (KD 14–18, 2022, Washington, DC, USA.* ACM, New York, NY, U https://doi.org/10.1145/3534678.3539448

1 INTRODUCTION

Knowledge-guided Self-Supervised Learning (KGSSL)

Method	Mean NSE
Baseline(actual characteristics)	0.560
Baseline ($0.5\sigma_i$ noise)	0.474
Baseline ($1\sigma_i$ noise)	0.245
KGSSL (1 year)	0.460
KGSSL (2 year)	0.535
KGSSL (3 year)	0.554
KGSSL (9 year)	0.582

Table 4: Forward model performance with corrupted characteristics and using KGSSL embeddings. In KGSSL (n-year), the n refers to the number of years of data utilized to learn the embedding.

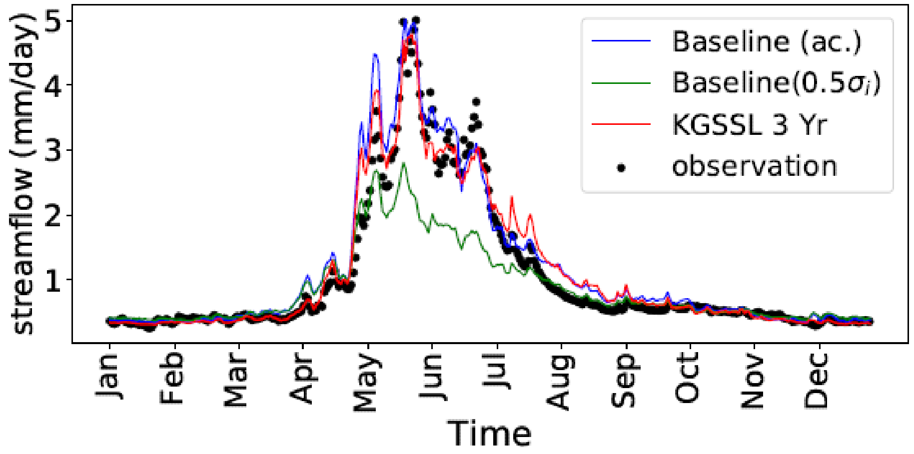

Figure 8: Basin at year 1992 (Best seen in color)

NSF IS-GEO (Intelligent Systems for the Geosciences)

- RCN: Research Coordination Network
- Led by Suzanne Pierce (UT Austin), Yolanda Gil (USC), Basil Tikhoff (U Wisconsin)
- Funded 2016-2021 with added no-cost extension. Just renewed for another 5 years
- Videos of talks organized by this program are available at
  - https://is-geo.org/resources/research-presentations/
- Suggestions for the U of Alabama meeting
  - First 15 minutes of this 28-minute video by Vipin Kumar
    - https://www.youtube.com/watch?v=udFhxyhas58
  - First 4 minutes of this 48-minute video by Mary Hill
    - https://www.youtube.com/watch?v=dwH9QzXIZpg

References on integrating process-based modeling and ML/AI

- Explainable AI: Interpretting, explaining, and visualizing deep learning, 2019
- Theory-guided data science: A new paradigm for scientific discovery from data Anuj Karpatne, Gowtham Atluri, James H Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, Vipin Kumar 2017/6/27 IEEE Transactions on knowledge and data engineering
- Machine learning for the geosciences: Challenges and opportunities A Karpatne, I Ebert-Uphoff, S Ravela, HA Babaie, V Kumar, IEEE Transactions on Knowledge and Data Engineering 31 (8), 1544-1554
- Intelligent systems for geosciences: an essential research agenda. Yolanda Gil, Suzanne A Pierce, Hassan Babaie, Arindam Banerjee, Kirk Borne, Gary Bust, Michelle Cheatham, Imme Ebert-Uphoff, Carla Gomes, Mary Hill, John Horel, Leslie Hsu, Jim Kinter, Craig Knoblock, David Krum, Vipin Kumar, Pierre Lermusiaux, Yan Liu, Chris North, Victor Pankratius, Shanan Peters, Beth Plale, Allen Pope, Sai Ravela, Juan Restrepo, Aaron Ridley, Hanan Samet, Shashi Shekhar (2019) Communications of the ACM, 62(1):76-84. DOI:10.1145/3192335

What I have done using process-based models

- Here are two references that represent my contributions, out of over 150
- Effective groundwater model calibration: with analysis of data, sensitivities, predictions, and uncertainty. MC Hill, CR Tiedeman – 2006. Wiley and Sons.
- Practical use of computationally frugal model analysis methods. Hill, M. C., Dmitri Kavetski, Martyn Clark, Ming Ye, Mazdak Arabi, Dan Lu, Laura Foglia, and Steffen Mehl. 2016. Groundwater. 54(2):159-170, DOI: 10.1111/gwat.12330