Scientifically-Interpretable Reasoning Network (ScIReN): Uncovering Hidden Relationships in the Soil Carbon Cycle

Joshua Fan[1]*, Haodi Xu[2]*, Feng Tao[3,4]*, Md Nasim[1], Marc Grimson[1], Yiqi Luo[2], Carla P. Gomes[1]

[1]Cornell University, Computer Science [2]Cornell University, Soil & Crop Science [3]Cornell University, Ecology and Evolutionary Biology
[4]Pennsylvania State University, Informatics and Intelligent Systems & Institute of Energy and the Environment *Equal contribution

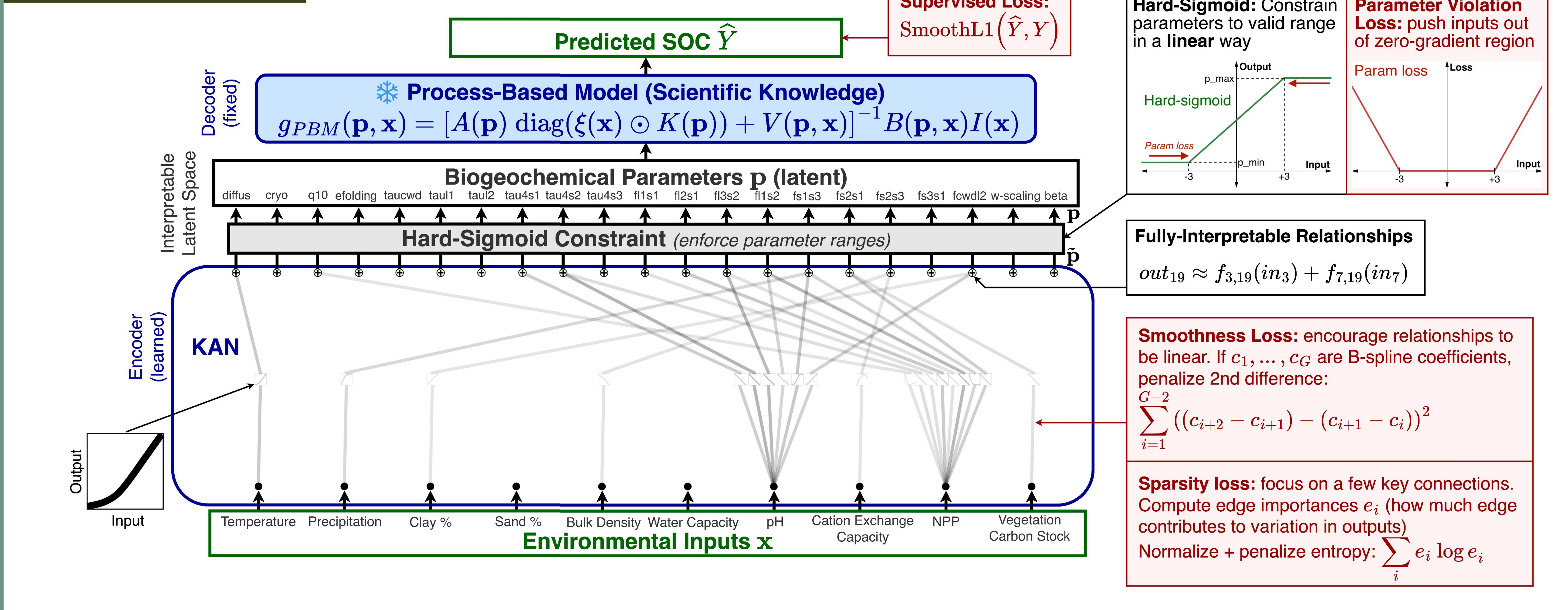
Motivation

- Soils store vast amounts of carbon (more than atmosphere+plants) [1]
  - Potential to remove CO2 from the atmosphere, mitigate climate change
- However, the **soil carbon cycle** remains poorly understood, leading to major uncertainties in climate projections [2]
  - e.g. How much carbon is currently stored in the soil? How long does it stay there?
- Scientists encode their scientific understanding into **process-based models** that simulate how carbon moves through the soil
  - Many unknown parameters that vary across space :(

Approach

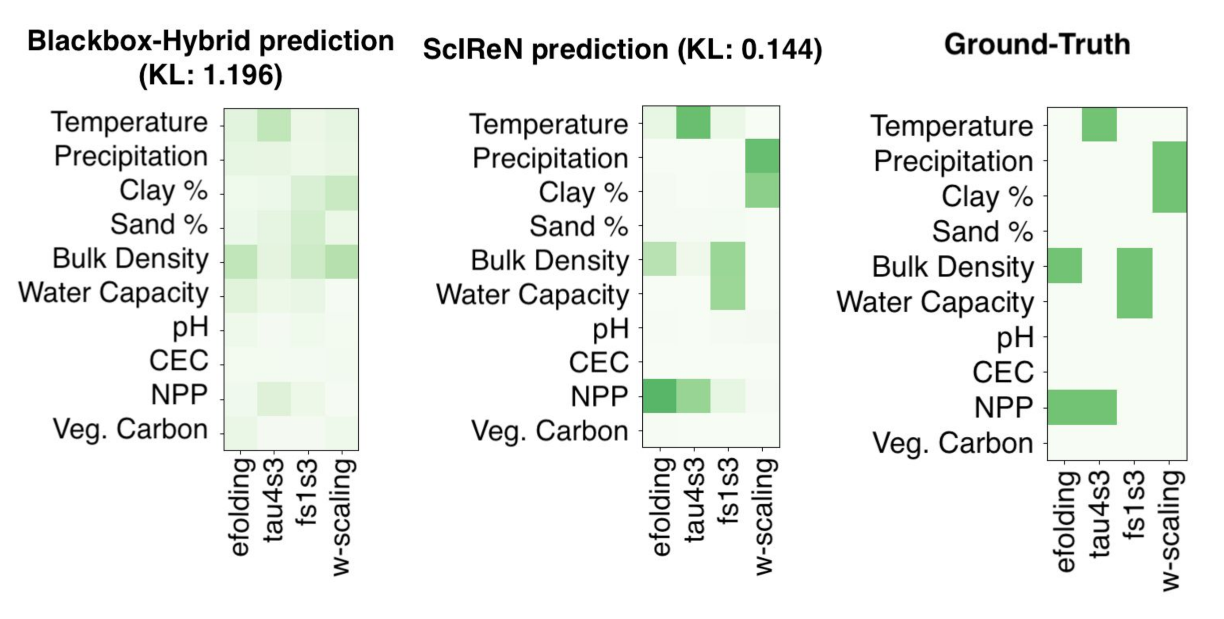
- Idea: embed differentiable process-based model inside neural network [3]
- **Encoder:** Given environmental features, predict latent parameters
- **Process-based "decoder":** given predicted parameters, uses **scientific knowledge** to simulate carbon flows and predict carbon amount in each pool
- However, the encoder is **not interpretable**
  - Scientists want **functional relationships** between inputs and parameters
  - Idea: use sparse Kolmogorov-Arnold network [4][5]
- We propose **Scientifically-Interpretable Reasoning Network (ScIReN)**, a transparent model that **respects existing knowledge** and **reveals new insights**

ScIReN Overview



Functional Relationships

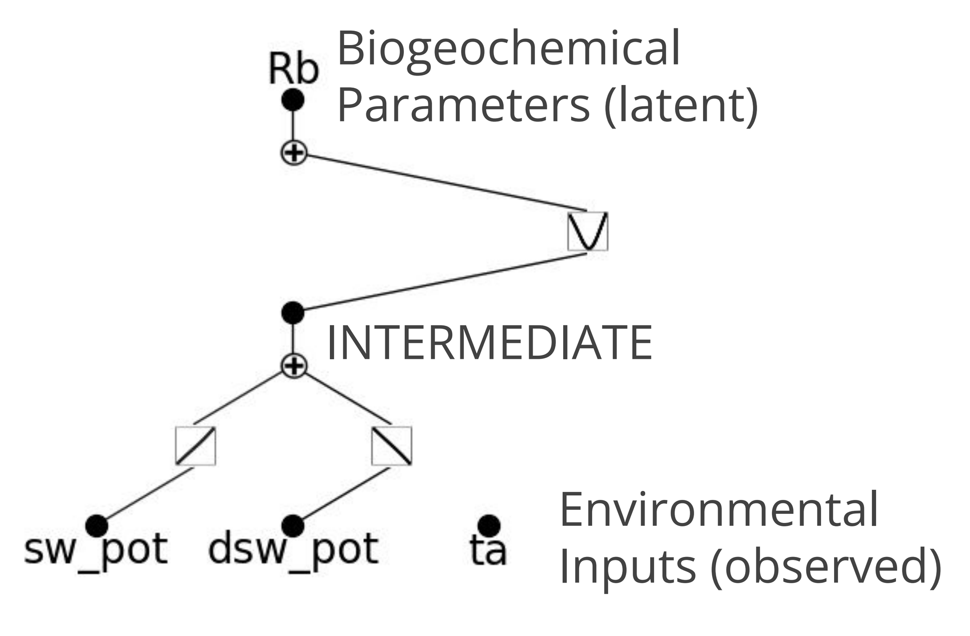
- Each curve reveals how changing one environmental variable will affect a biogeochemical process
- Each output is the sum of interpretable contributions from each input, e.g.
  - **output1 = (input1)^2 + exp(input6) + ...**
- How accurate are these? Prescribed functional relationships between 10 inputs and 4 most sensitive biogeochemical parameters (right)



- BINN (left) does not reveal functional relationships. If we apply a post-hoc interpretation method, it did not even implicitly learn the correct relationships.
- ScIReN (center) revealed the relationships accurately. **(Also worked on a different domain)**

Going deeper

If we need non-additive feature interactions, try multi-layer KANs! Example:



INTERMEDIATE = f_1(sw_pot) + f_2(dsw_pot)

Rb = f_3(INTERMEDIATE)

Results

Synthetic labels: ScIReN is by far the best at recovering latent parameters and functional relationships, and extrapolating out-of-distribution (see paper for more)

Method	R^2 (observed, ↑)	R^2 (latent, ↑)	KL, functional relationships (↓)
Pure-NN	0.933 ± 0.015	N/A	N/A
Blackbox-Hybrid, nonlinear constraint	0.996 ± 0.003	0.226 ± 0.800	1.312 ± 0.170
Blackbox-Hybrid, linear constraint	0.995 ± 0.003	0.721 ± 0.226	1.082 ± 0.258
Linear-Hybrid, hardsigmoid	0.973 ± 0.013	0.087 ± 1.014	1.727 ± 0.322
ScIReN, linear constraint (1-layer KAN)	**0.999** ± 0.002	**0.989** ± 0.020	**0.080** ± 0.042

Real labels: With ScIReN, we get **interpretability without sacrificing accuracy**. While we don't have ground-truth for latent params/relationships, they match domain knowledge.

Method	R^2(↑)	MAE (↓)	Pearson correlation (↑)
Pure-NN	0.552 ± 0.173	**4609.3** ± 356.8	**0.780** ± 0.053
Blackbox-Hybrid, nonlinear constraint	0.584 ± 0.082	4726.2 ± 727.3	0.776 ± 0.048
Blackbox-Hybrid, linear constraint	**0.589 ± 0.070**	4849.7 ± 650.3	0.774 ± 0.040
Linear-Hybrid, hardsigmoid	0.552 ± 0.082	4984.8 ± 771.6	0.761 ± 0.046
ScIReN, linear constraint (1-layer KAN)	0.582 ± 0.080	4708.2 ± 673.1	0.769 ± 0.049
ScIReN, linear constraint (2-layer KAN)	0.571 ± 0.094	4707.3 ± 826.3	0.765 ± 0.052

Conclusion

We propose ScIReN, a method that

- **Respects existing scientific knowledge**, provided by any process-based model
- **Reveals new functional relationships** between environmental inputs and **unobserved biogeochemical processes**

The system is trainable end-to-end; every part of the model is **fully transparent**.

Potential future directions:

- Apply ScIReN to new domains
- Experimentally validate the functional relationships learned by the model
- Make ScIReN easier to train
- Understand **uncertainty** of revealed functional relationships
- Improve spatial generalization, e.g. geographic positional embeddings or domain adaptation

Acknowledgements

SCHMIDT FUTURES
Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a program of SCHMIDT FUTURES
Cornell Bowers CIS College of Computing and Information Science
NCAR CISL
Cornell CALS College of Agriculture and Life Sciences
U.S. National Science Foundation
ICS
ECOLAB OF DR. YIQI LUO
USDA National Institute of Food and Agriculture U.S. DEPARTMENT OF AGRICULTURE
AI-LEAF

References

[1] Jobbagy, E. G.; and Jackson, R. B. 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological applications*, 10(2): 423–436.

[2] Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., ... & Zhou, T. (2016). Toward more realistic projections of soil carbon dynamics by Earth system models. *Global Biogeochemical Cycles,* 30(1), 40-56.

[3] Xu, H., Fan, J., Tao, F., Jiang, L., You, F., Houlton, B. Z., ... & Luo, Y. (2025). Biogeochemistry-Informed Neural Network (BINN) for Improving Accuracy of Model Prediction and Scientific Understanding of Soil Organic Carbon. arXiv preprint arXiv:2502.00672. **Submitted to *Geoscientific Model Development*.**

[4] Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., ... & Tegmark, M. (2024). Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756.

[5] Liu, Z.; Ma, P.; Wang, Y.; Matusik, W.; and Tegmark, M. 2024b. Kan 2.0: Kolmogorov-arnold networks meet science. arXiv preprint arXiv:2408.10205.