ECOS[s] Center for Ecosystem Science and Society at Northern Arizona University

Physics-Informed Neural Networks (PINNs) to improve biological reality of machine-learning-based models of the soil CO_2 flux in subalpine forests

Yujie Liu

Center for Ecosystem Science and Society
Northern Arizona University

Rocky Mountain Biological Laboratory

NOAA — National Oceanic and Atmospheric Administration, U.S. Department of Commerce

NSF

ESS Environmental System Science

U.S. Department of Energy | Office of Science

SNODDY Project

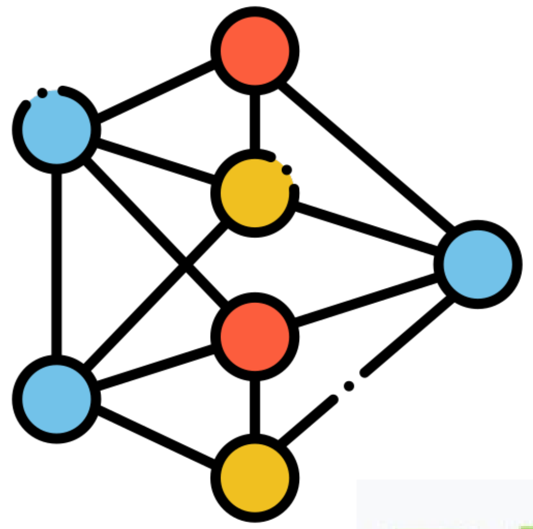
SNODDY Project

DOE ESS Grant

- PI - Mariah Carbone
- Co-PIs: Andrew Richardson, Ben Lucas, Will Wieder, Adrianna Foster
- Collaborators: Austin Simonpietri, Bruce Hungate, Egbert Schwartz, Gena Blumencwejg, Kenneth Williams (LBNL)

qSIP

^{14}C

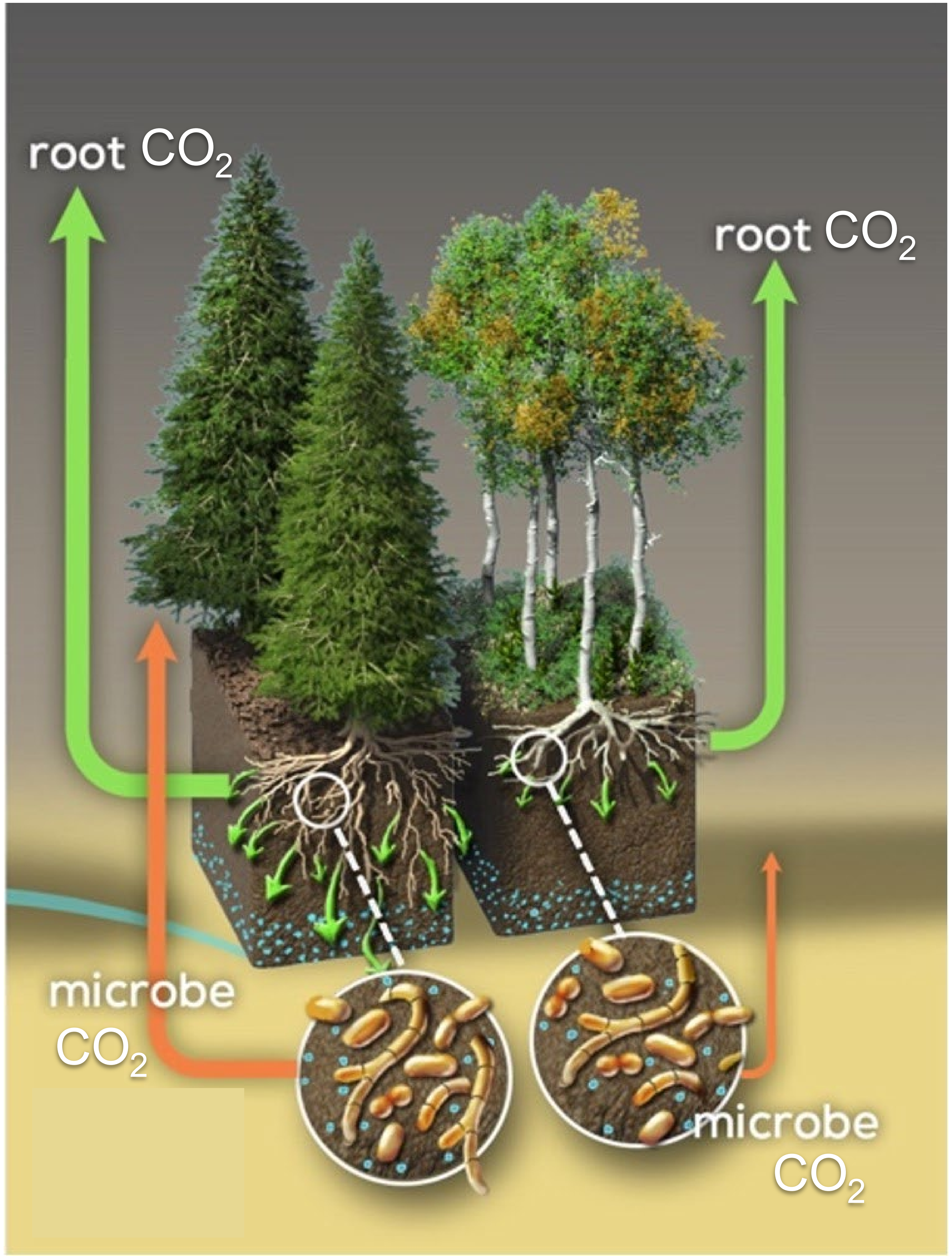


CLM

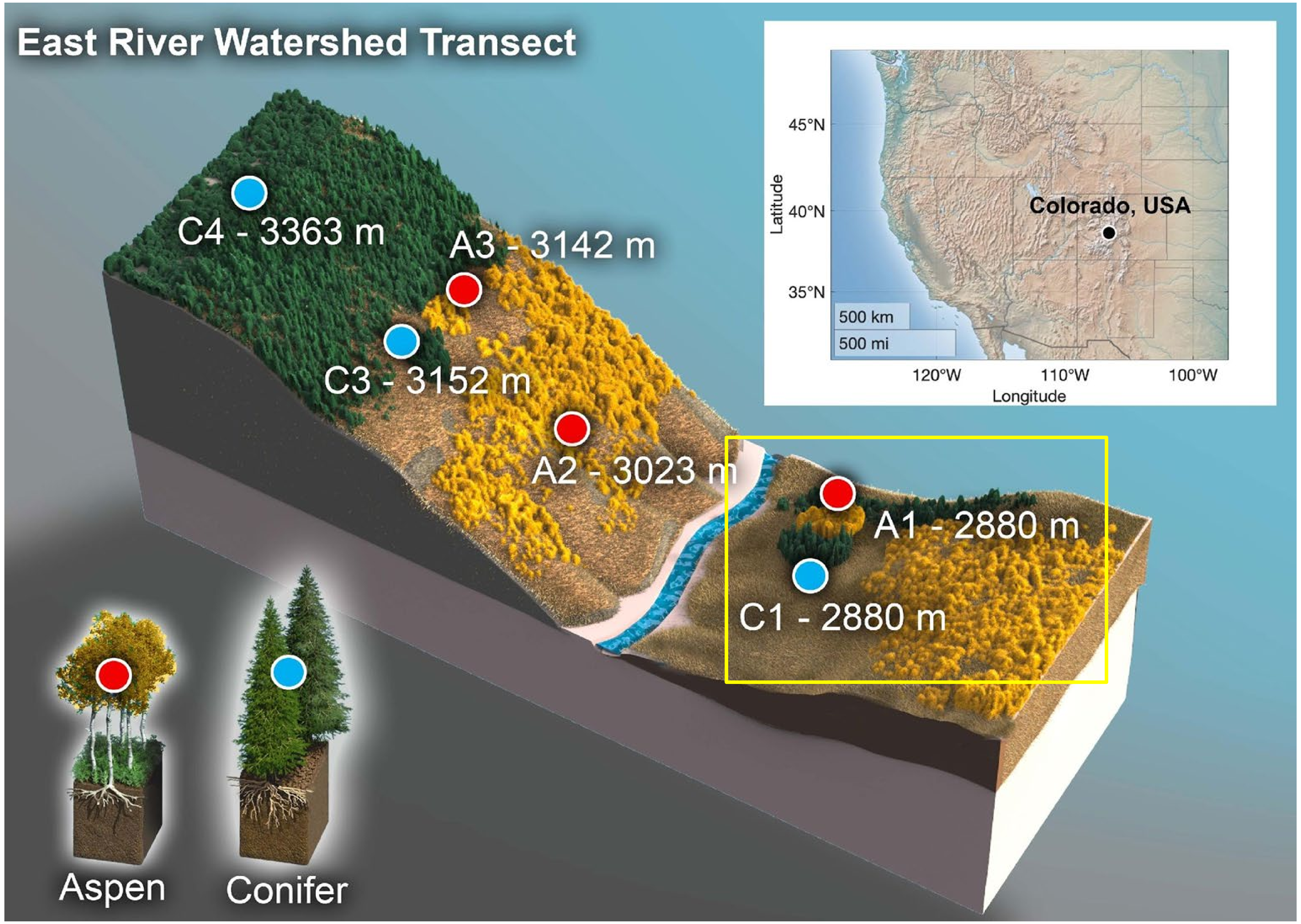
PhenoCam

Why the soil CO_2 flux?

- 2nd largest flux
- Two biological sources:
- Microbial decomposition
- Root respiration
- Difficult to tease apart



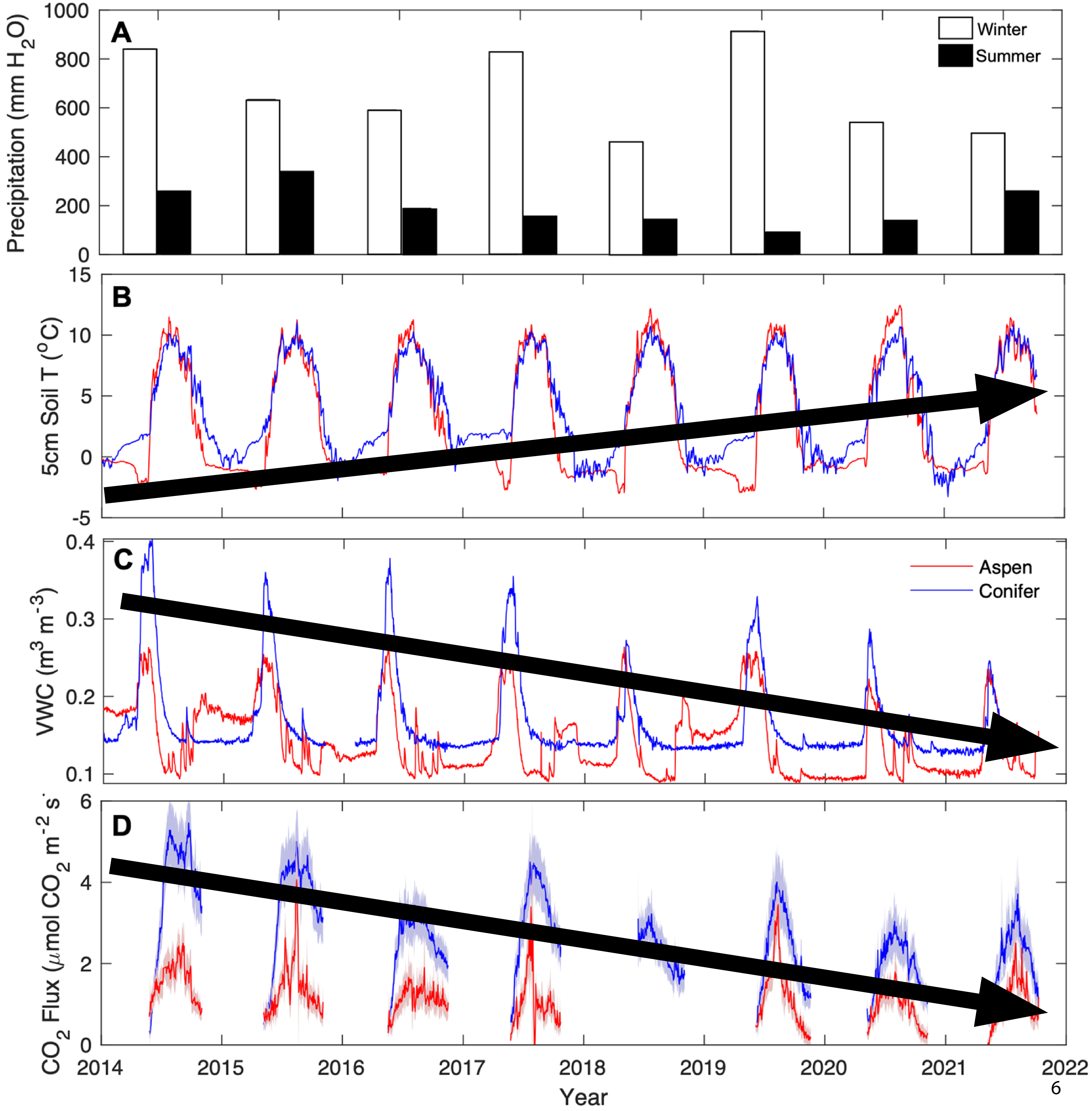
Field sites



- Snodgrass mountain, in the East River watershed, CO
- 2880 m elevation

Dominant forest types:

- Aspen and mixed conifer spruce/fir stands

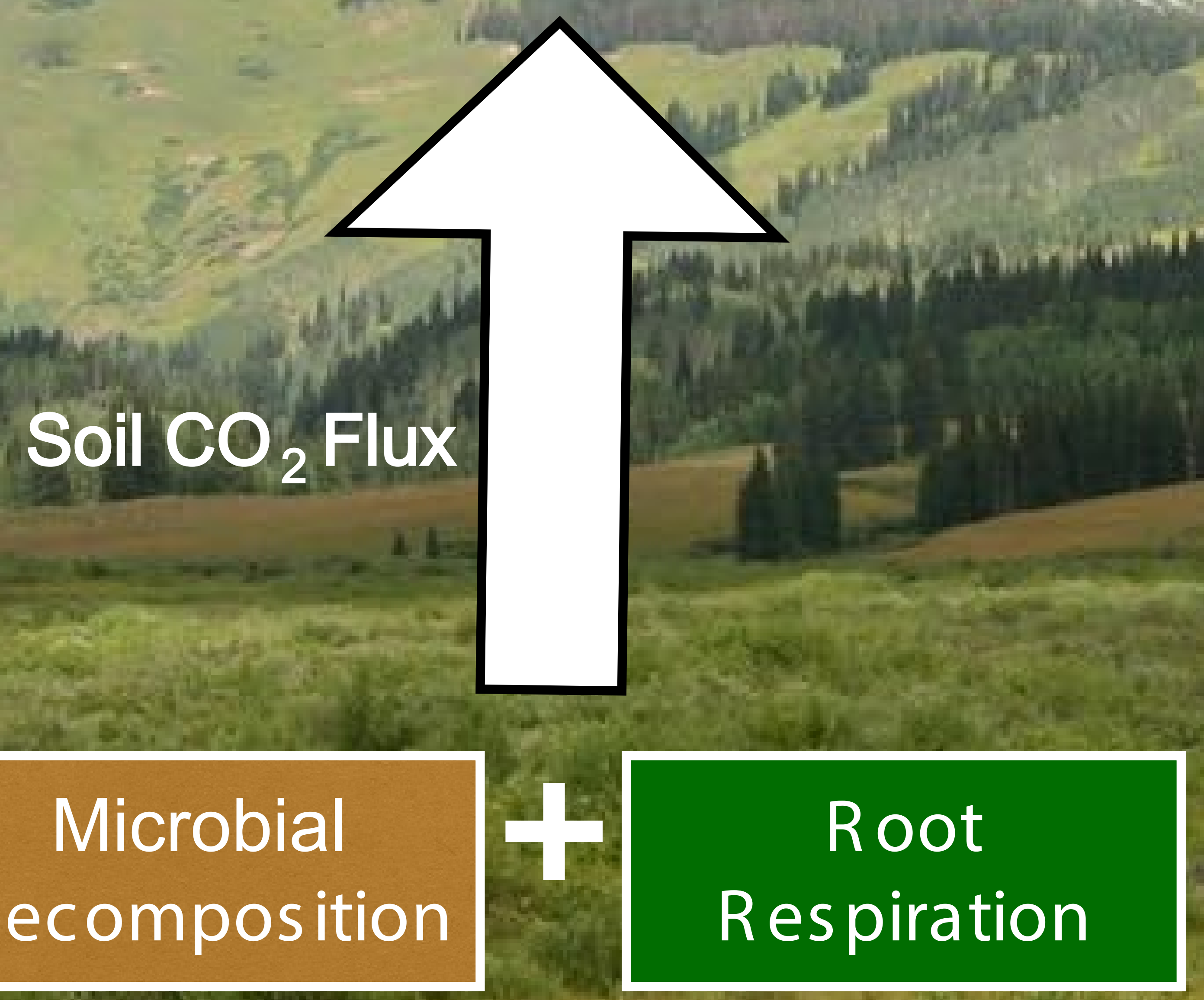


- Variable snow/rain inputs
- Increasing soil and air temperature
- Decreasing soil moisture
- Decreasing soil CO_2 flux
- Surface soils responsible for declining trend

Carbone et al. (2023) ERL

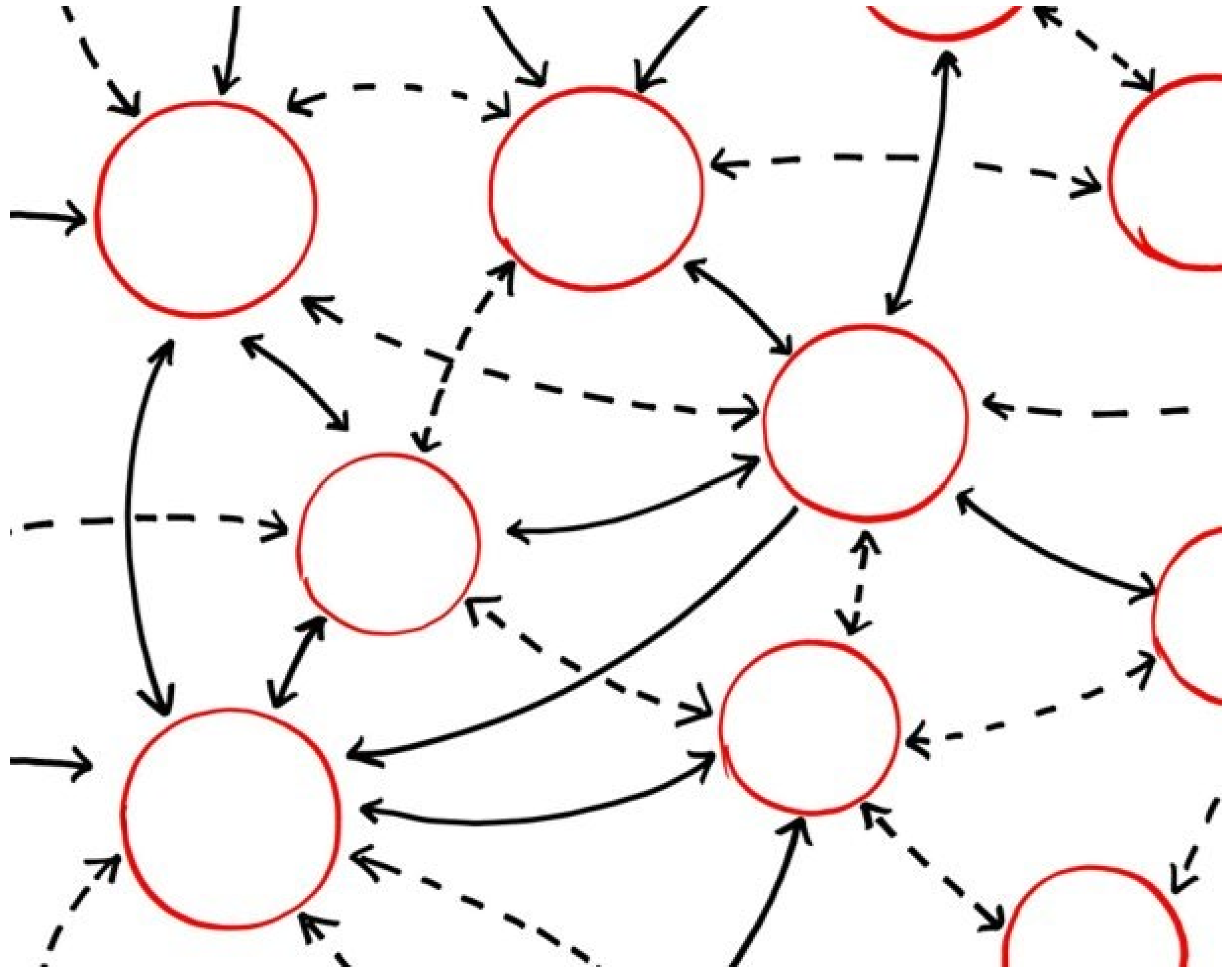
Motivation

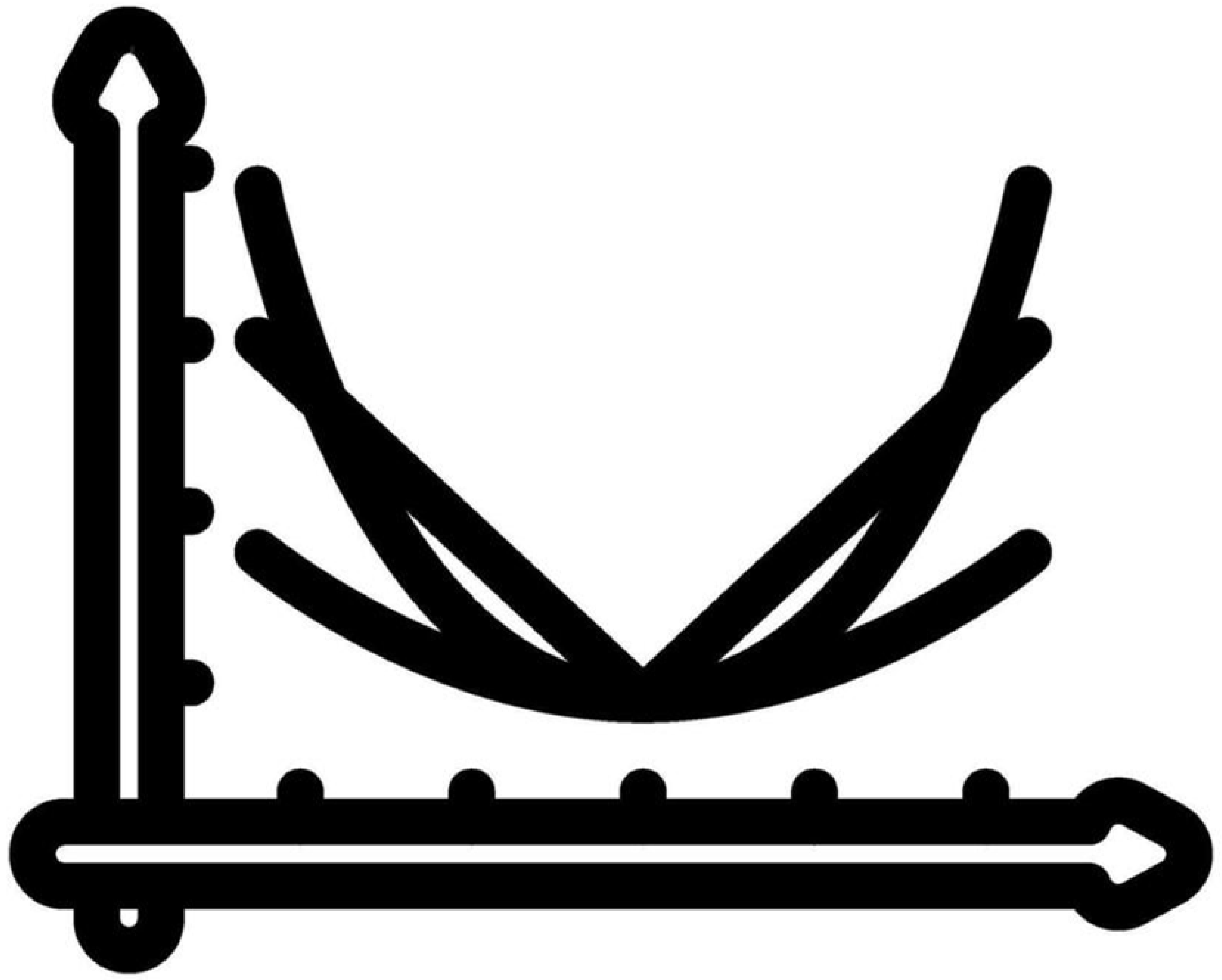
Toward more accurate and ecologically realistic modeling of soil CO_2 flux and its microbial and root contributions



Machine learning goal

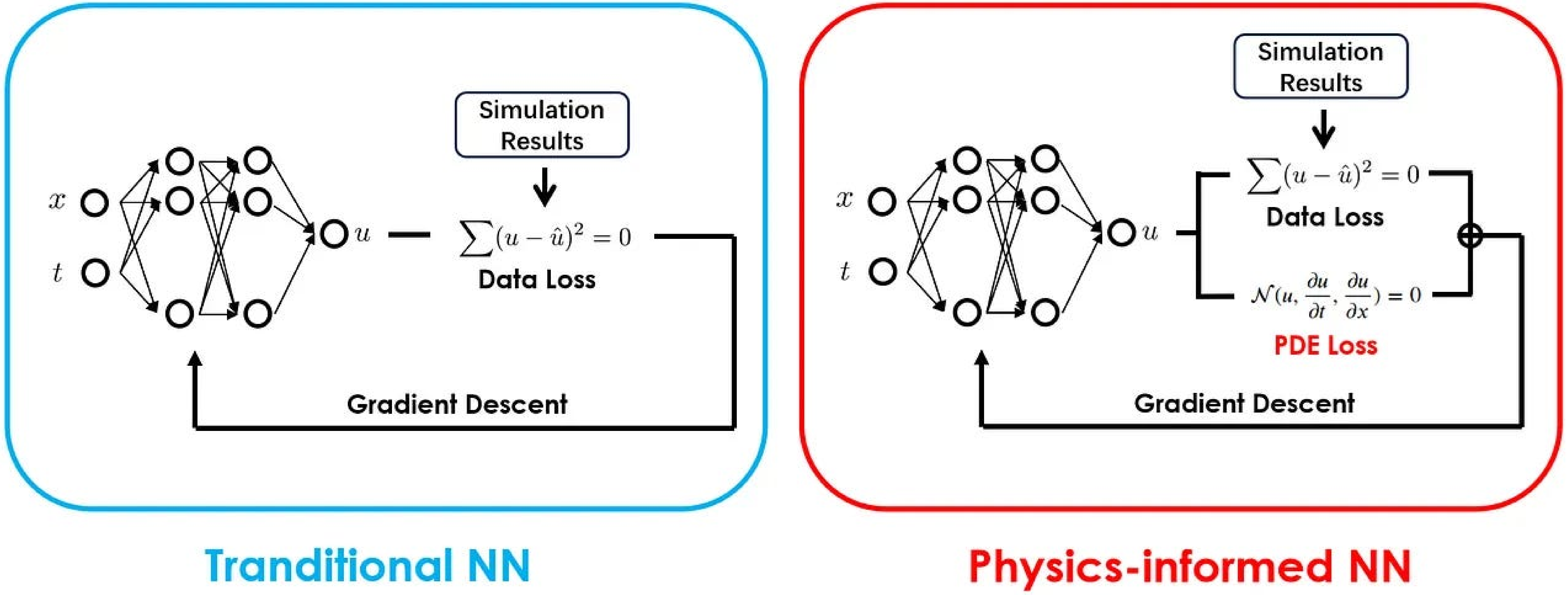
- Goal of classic NN: minimize loss
- Loss function: Mean squared error (observed – predicted)
- Limitation: overfit data, leading to biological implausibility.
- Temperature, moisture, carbon supply





Physics -Informed Neural Network (PINNs)

Modify the loss

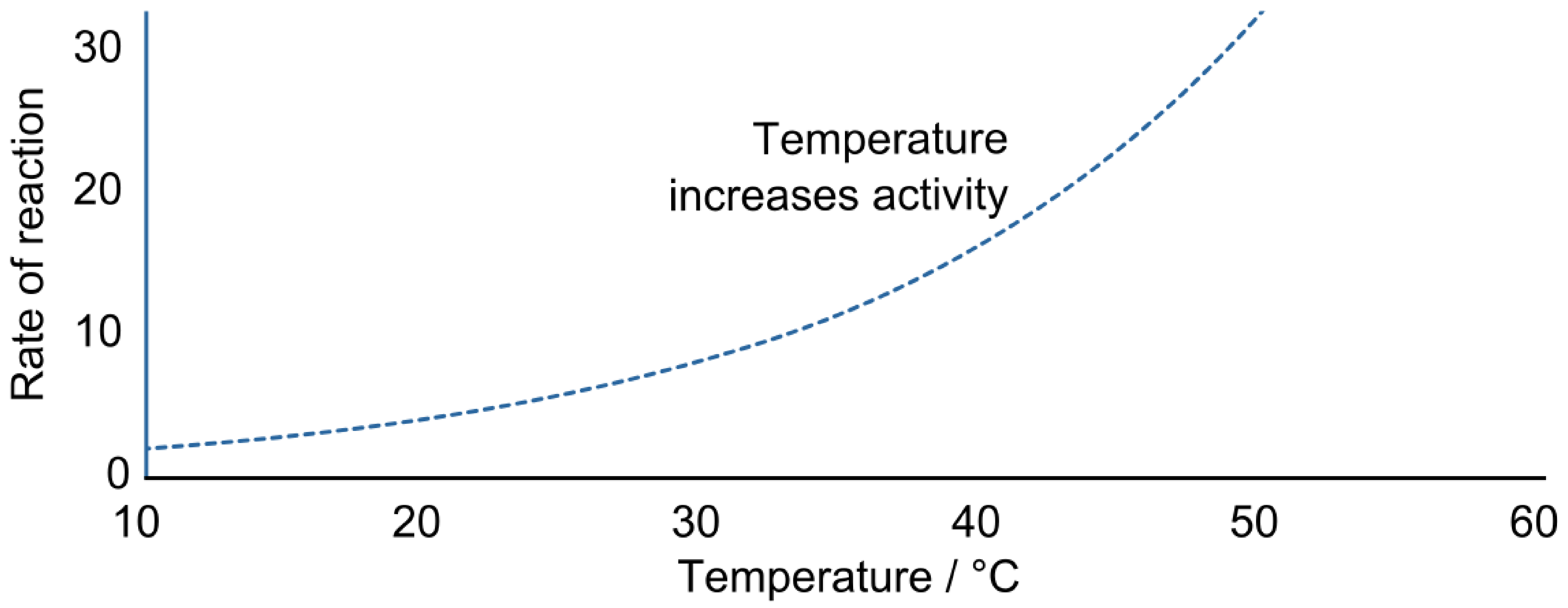


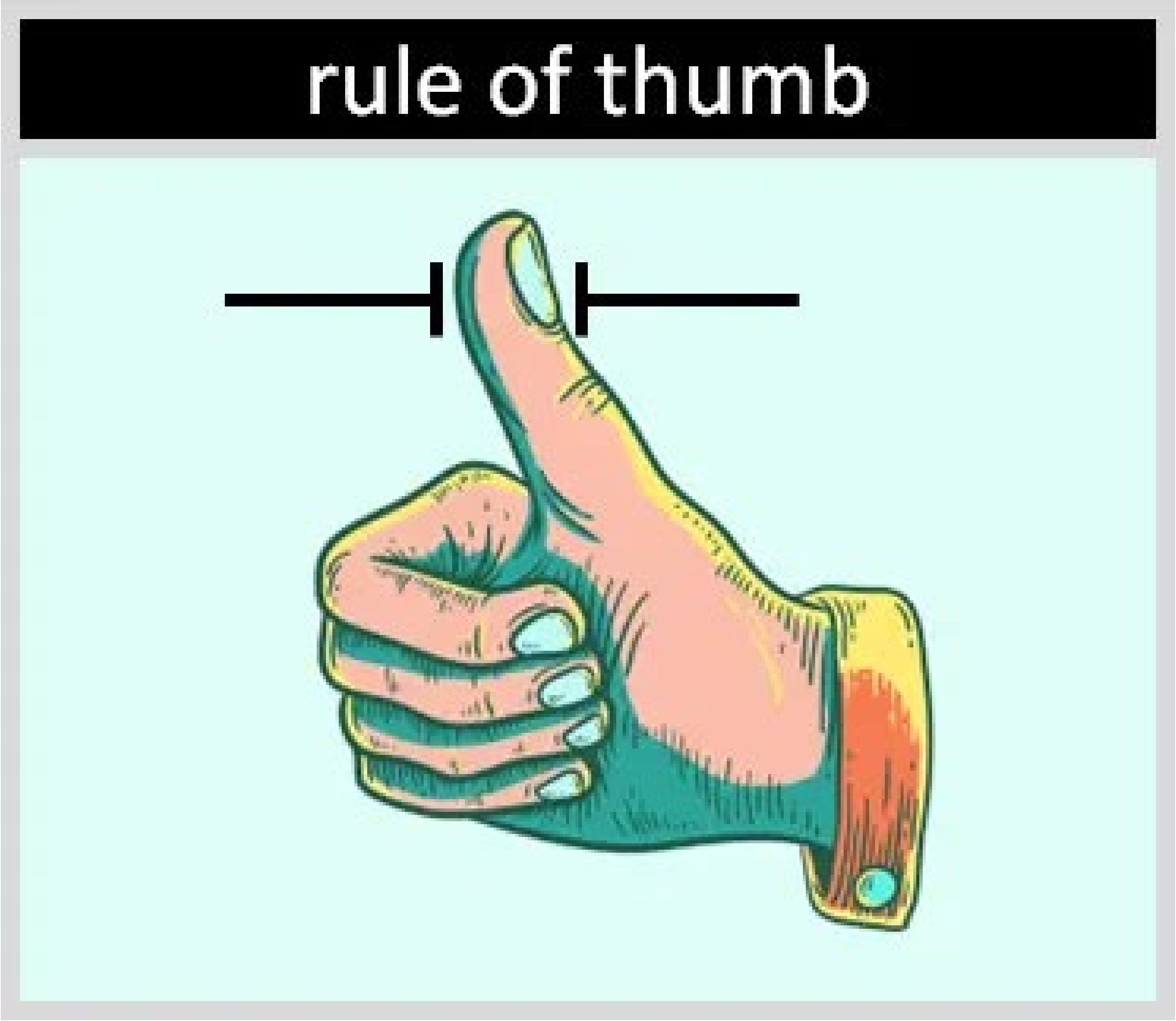
- Idea: Modify data loss with process penalty

Process penalty: Q10

A few notes

- Q10 quantifies how biological reaction rates change with a 10 °C increase in temperature;
- Q10 of 2 is often used in process-based models to represent the temperature response of a plant or ecosystem;

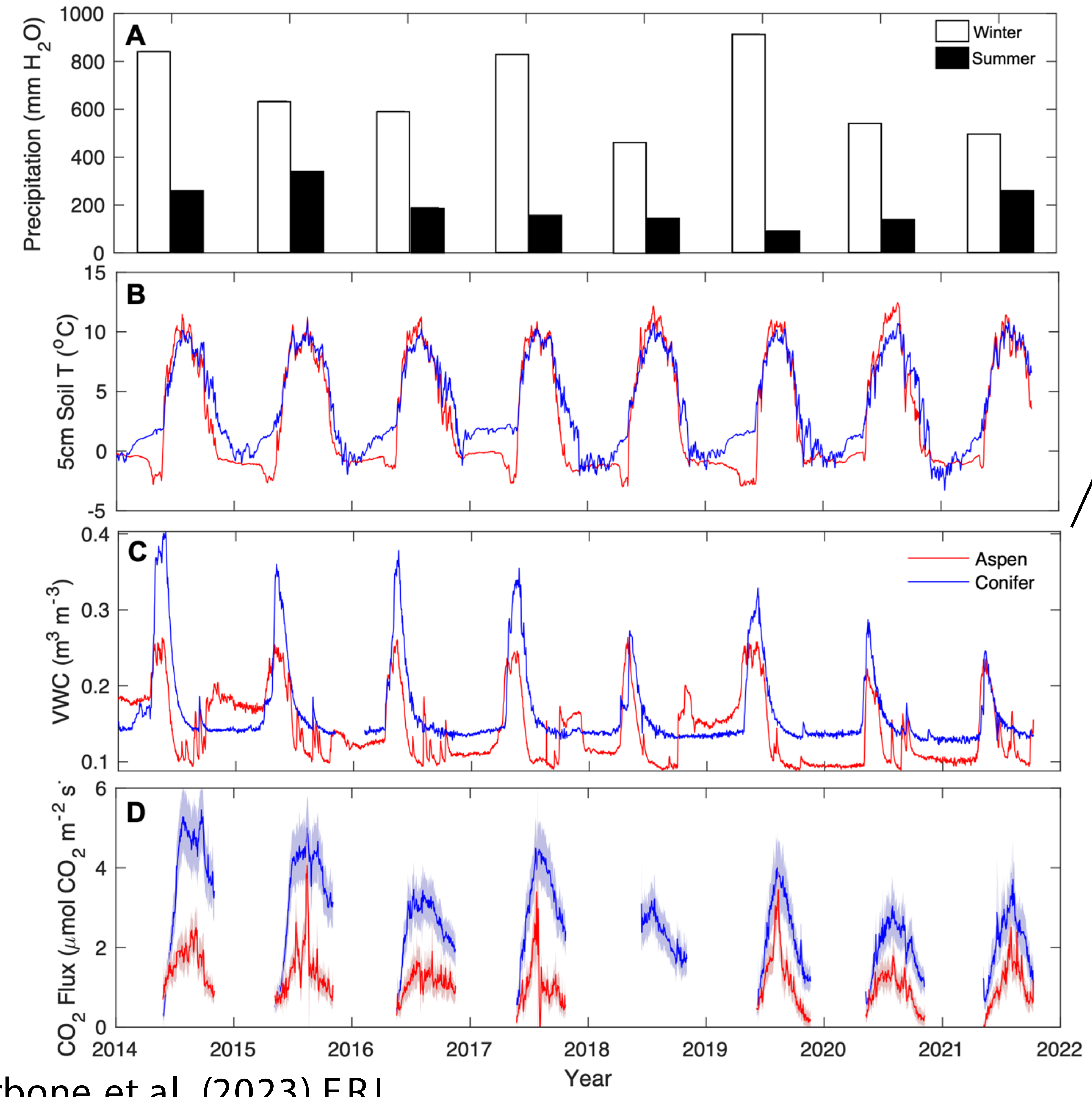




(Potter et al., 1993)

Methods

Classic NN vs Q10 constrained NN



Carbone et al. (2023) ERL

Load data

Daily soil CO_2 efflux (R), Tsoil, VWC, snow/precipitation, stand

Neutral Network architecture

- MLP with tanh activation & dropout
- Target variable: log(R)
- Adam optimizer

Loss function

Classic NN (data fit):

L = MSE(exp(logR_pred), R_obs)

Q10-constrained PINN

L = MSE + λ · L_Q10

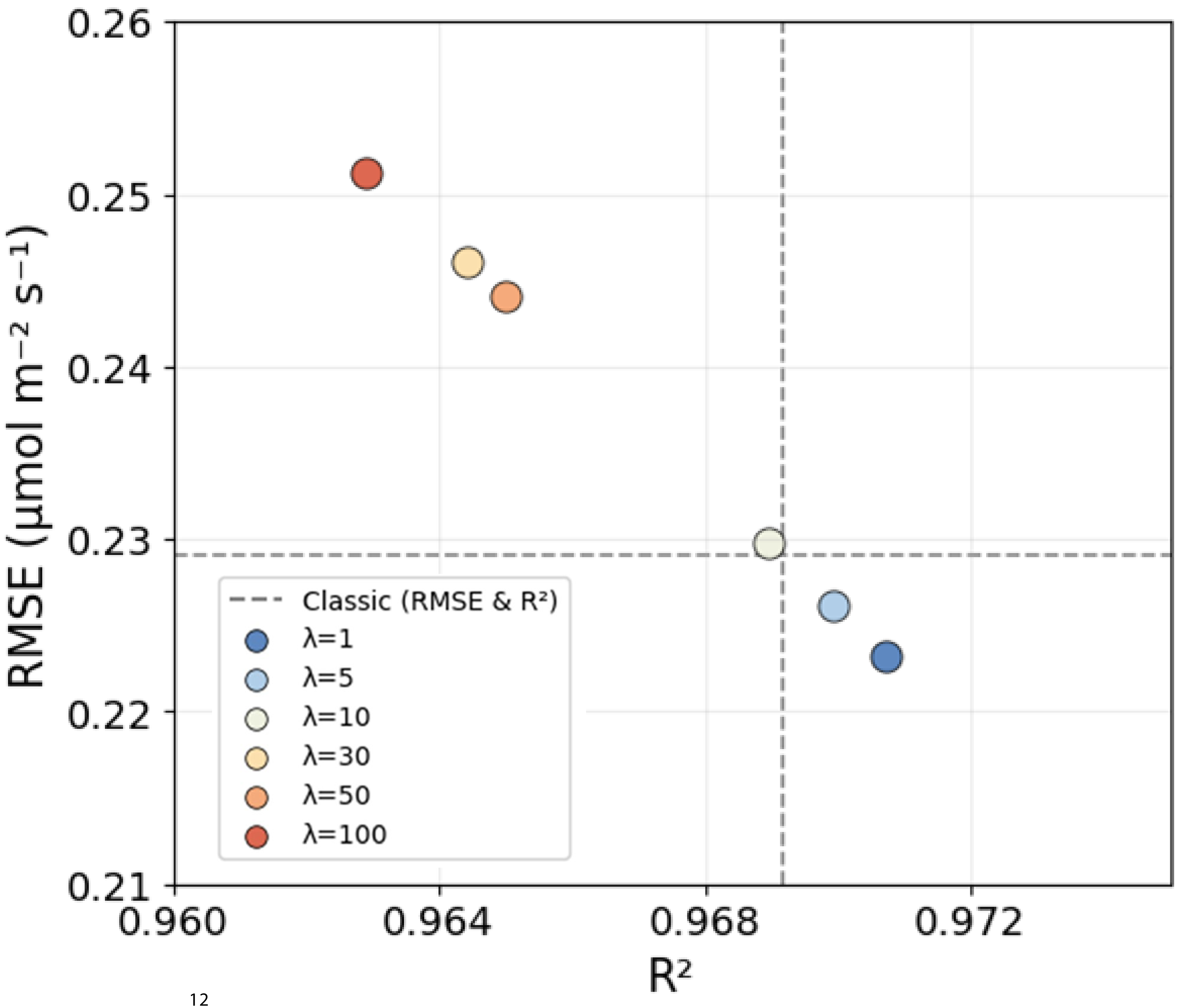
Test for different values λ

Compute local Q10

Results

Predictive skill

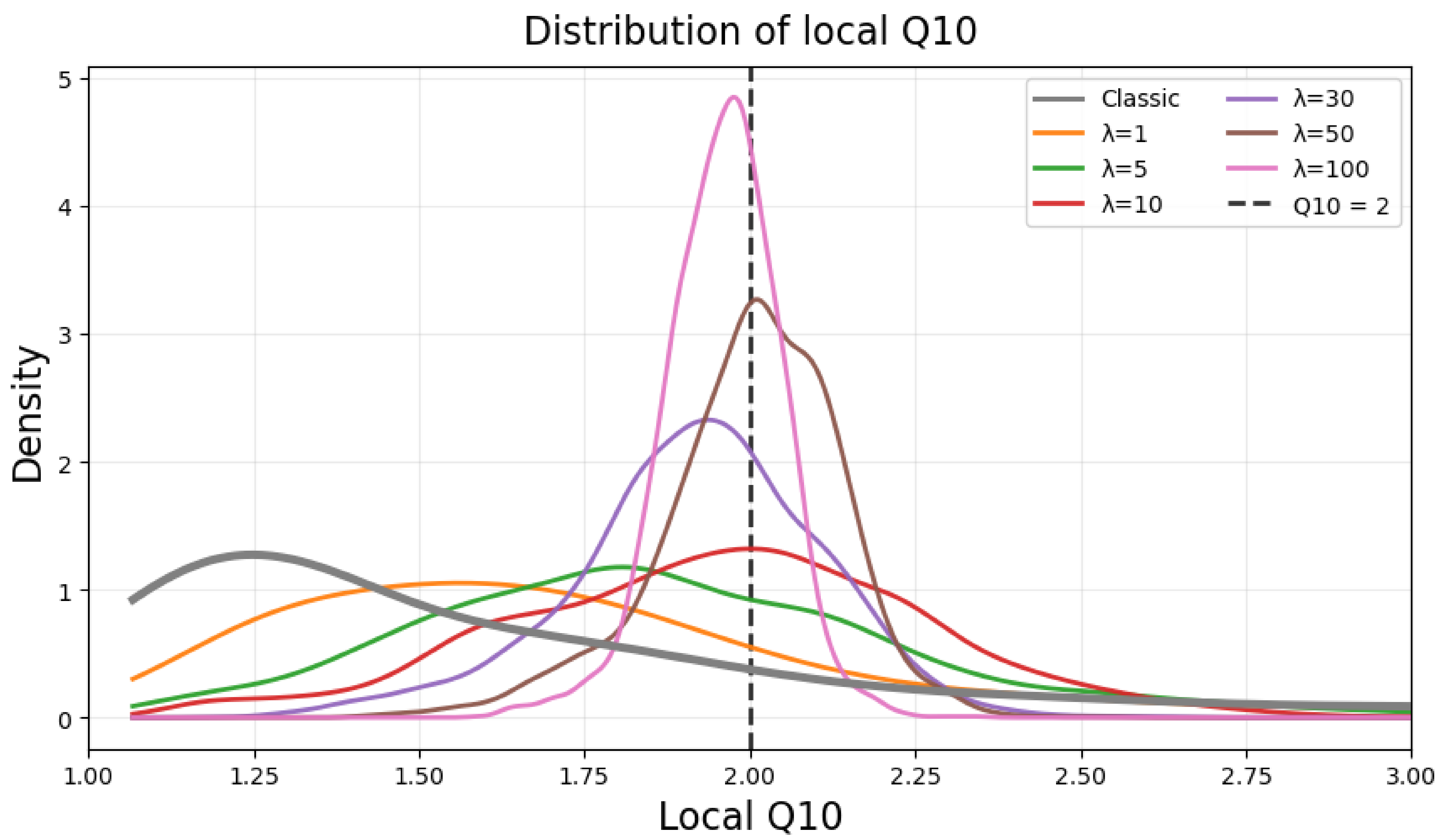
- Grey dash lines: classic NN
- Each point: a Q10 constrained PINN with different λ
- Small λ values (1,5): Q10 constrained PINN outperforms classic NN;
- Q10 constraints are keeping the model from over fitting data;



Results

Q10 distribution

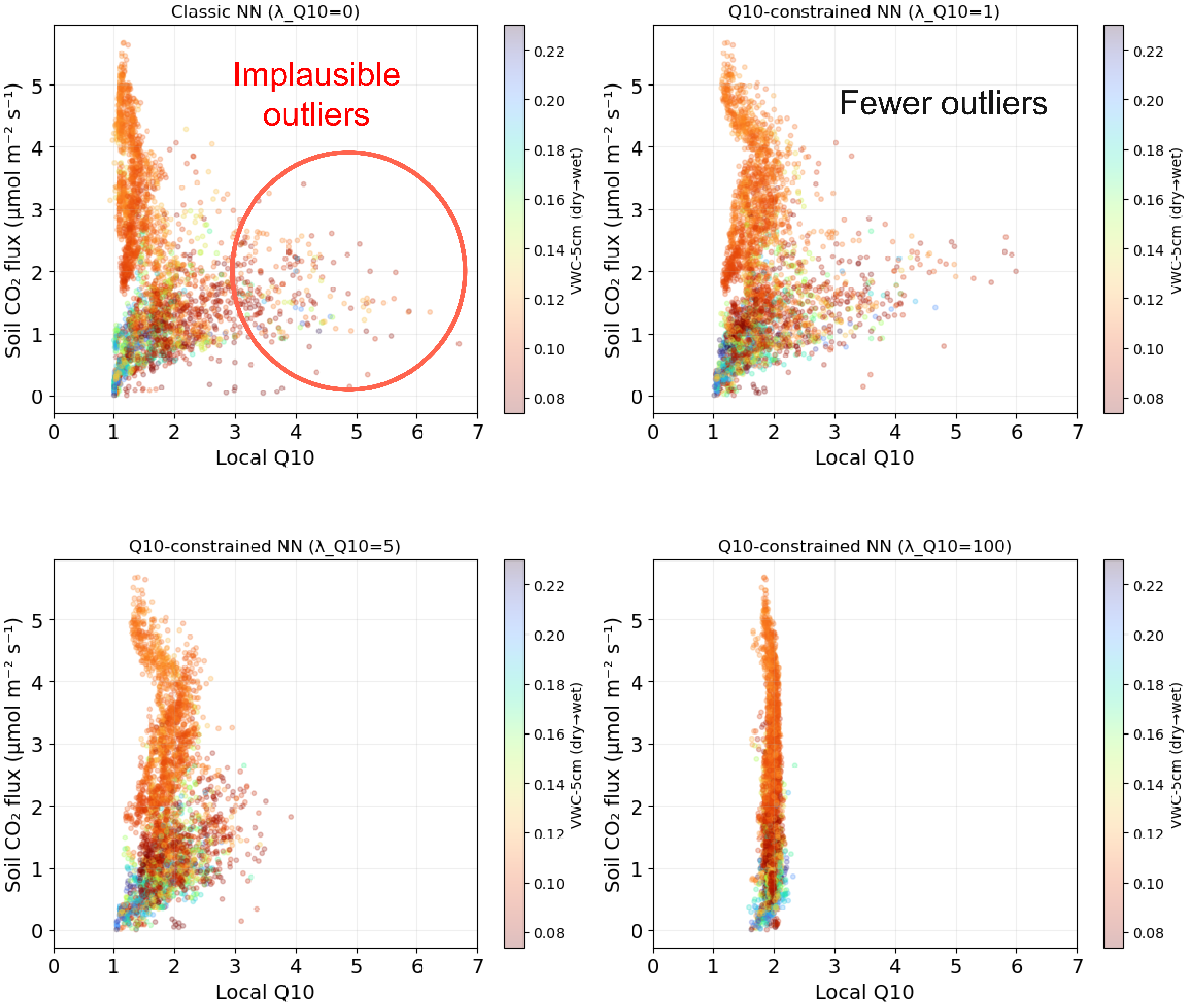
- Grey line: classic NN;
- Classic NN: highly scattered Q10 with high variability;
- PINN narrows Q10 estimates to a biologically plausible band.
- As λ increases, the Q_{10} values become narrower and are pulled closer to the target value of 2.



Results

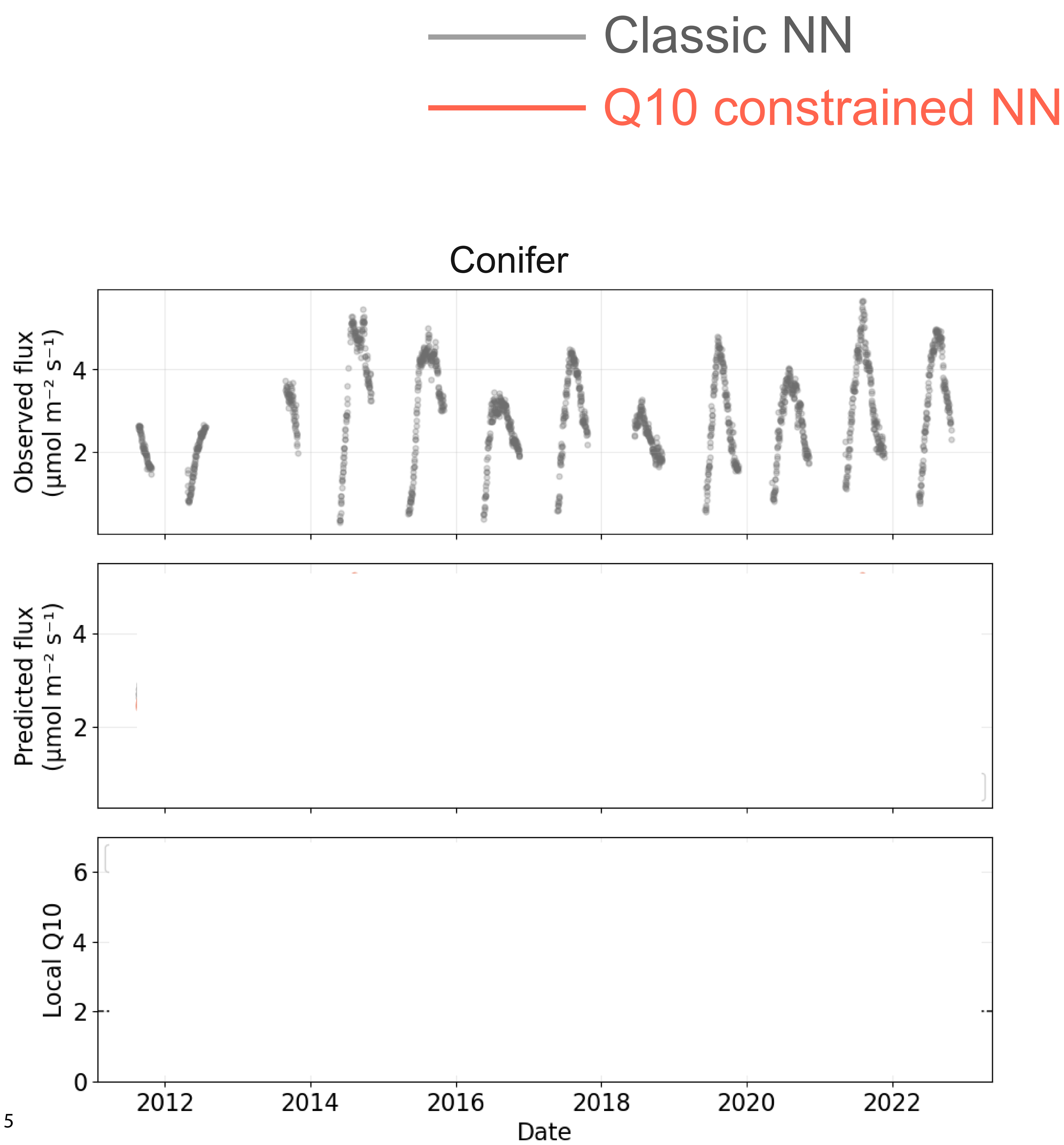
Q10 vs Flux

- Classic NN generates implausible outliers;
- Q10 constrained PINN stabilized the behavior;
- λ = 5, good trade-off between stability and allowing natural variability;



Results

Time series

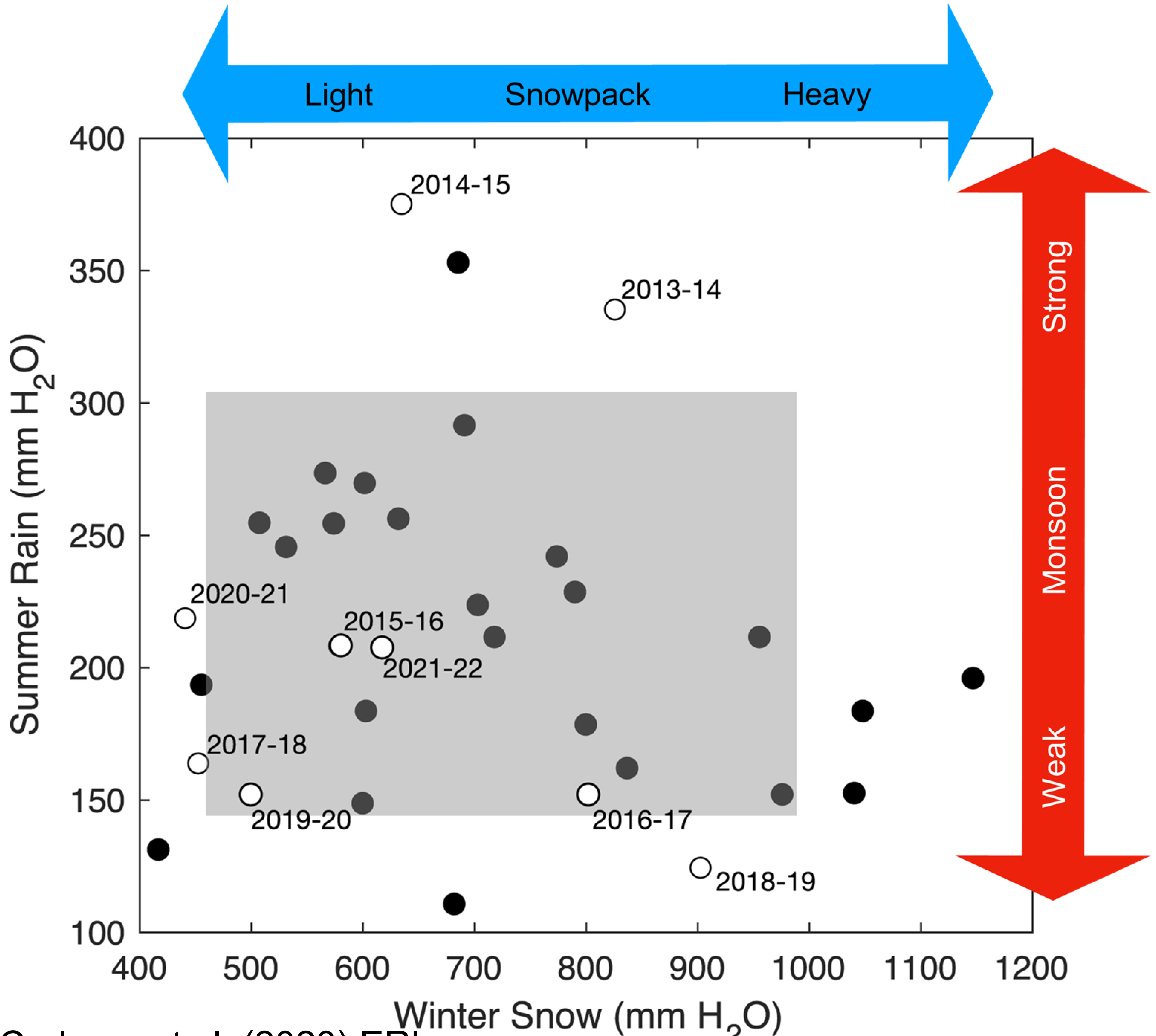


Q10 constrained PINN (λ = 5)

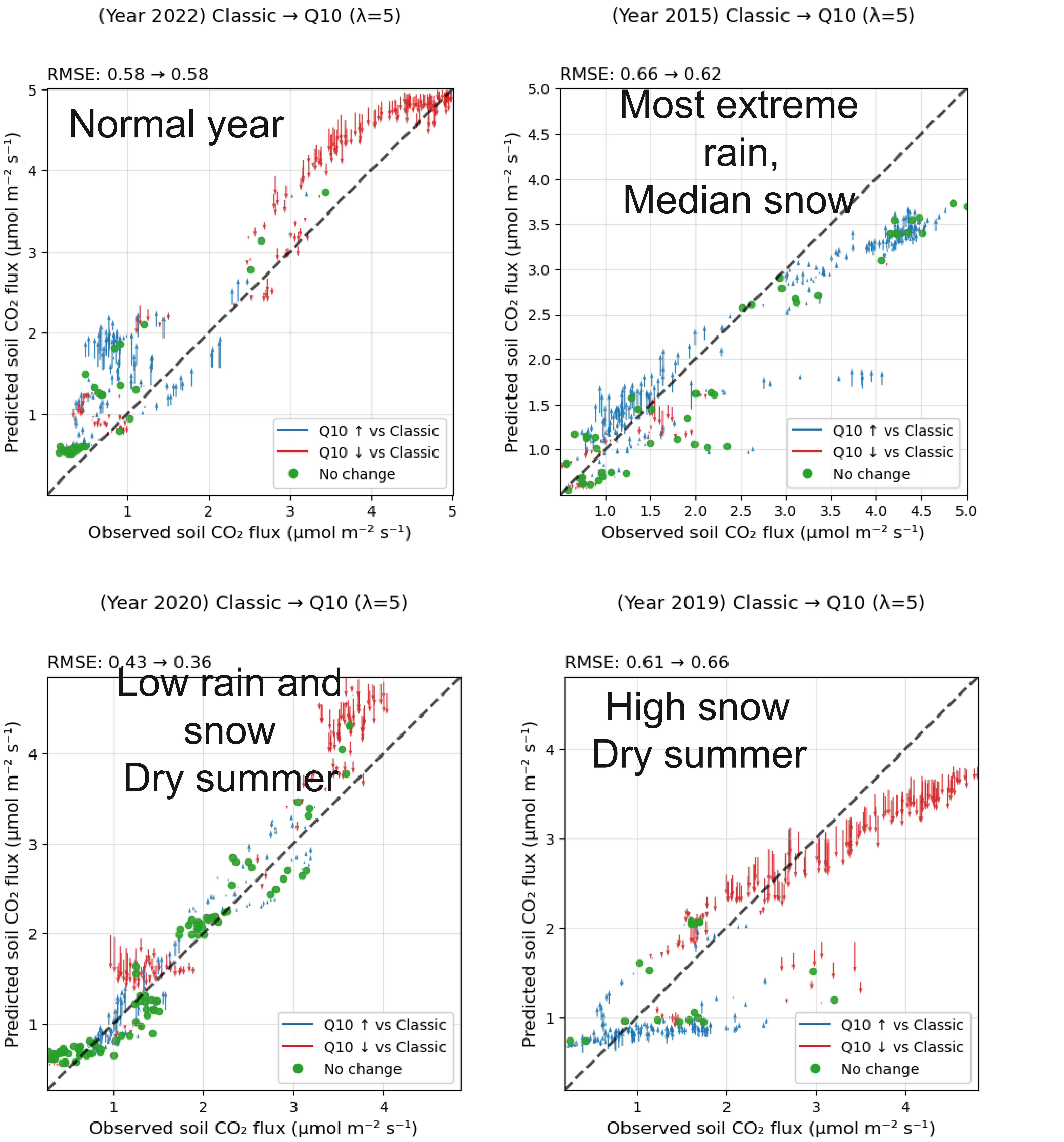
- has slightly higher RMSE and R^2 than the classic NN;
- reproduces the seasonal patterns almost identically to the classic NN;
- smooths out some abrupt jumps that are visible in measurements, but that may NOT be realistic;
- has more realistic seasonal variations in Q10

Results

Leave one year out



Carbone et al. (2023) ERL



Future direction

Fractional contribution of two sources

- Target: microbial decomposition and root respiration;
- Measurement: ^{14}C partitioning;
- Valuable but sparse dataset;

Mass balance: $\Delta^{14}C_{Soil\ CO2} = F \times \Delta^{14}C_{root} + (1\text{-}F) \times \Delta^{14}C_{decomp}$

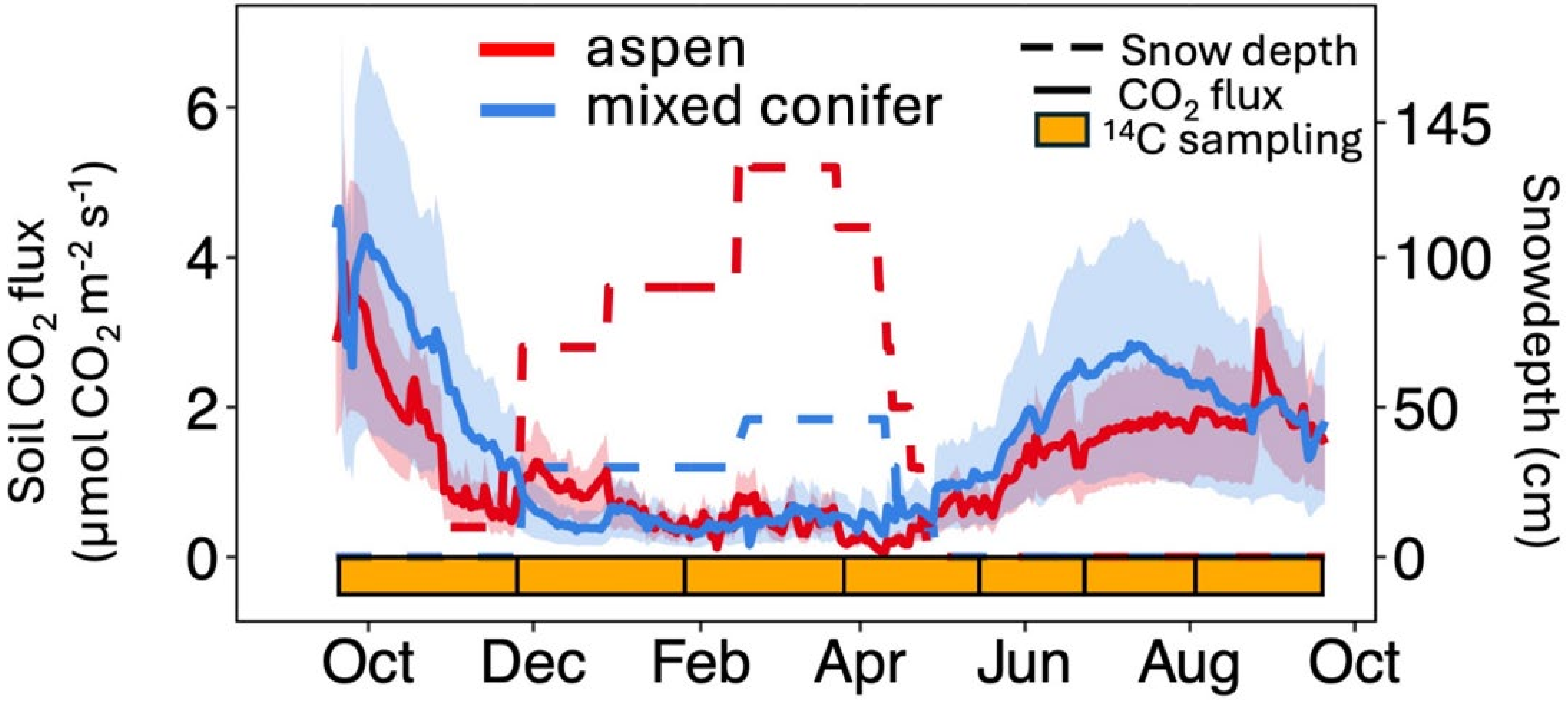
Collect in field

Fraction of respiration from roots

Root incubation

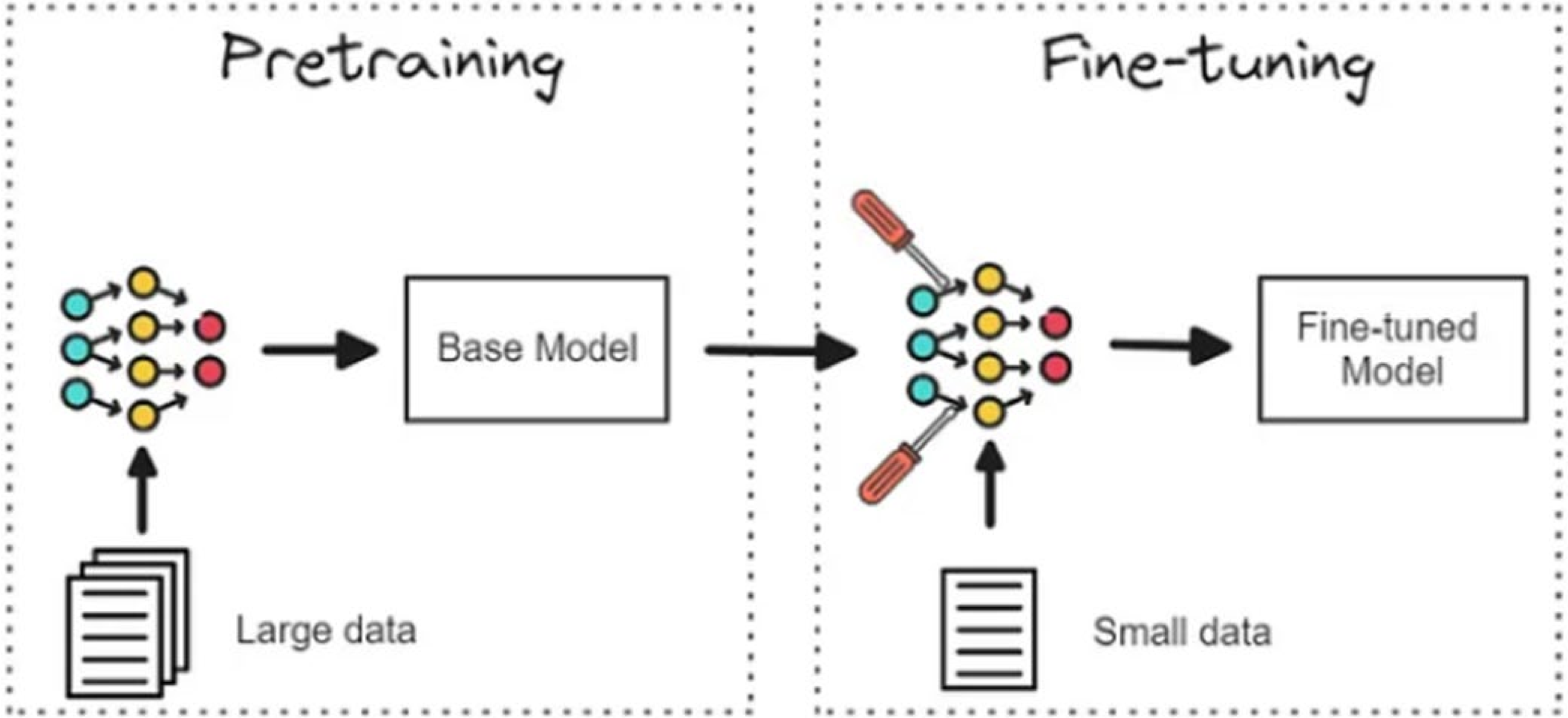
Fraction of respiration from microbes

Soil incubation



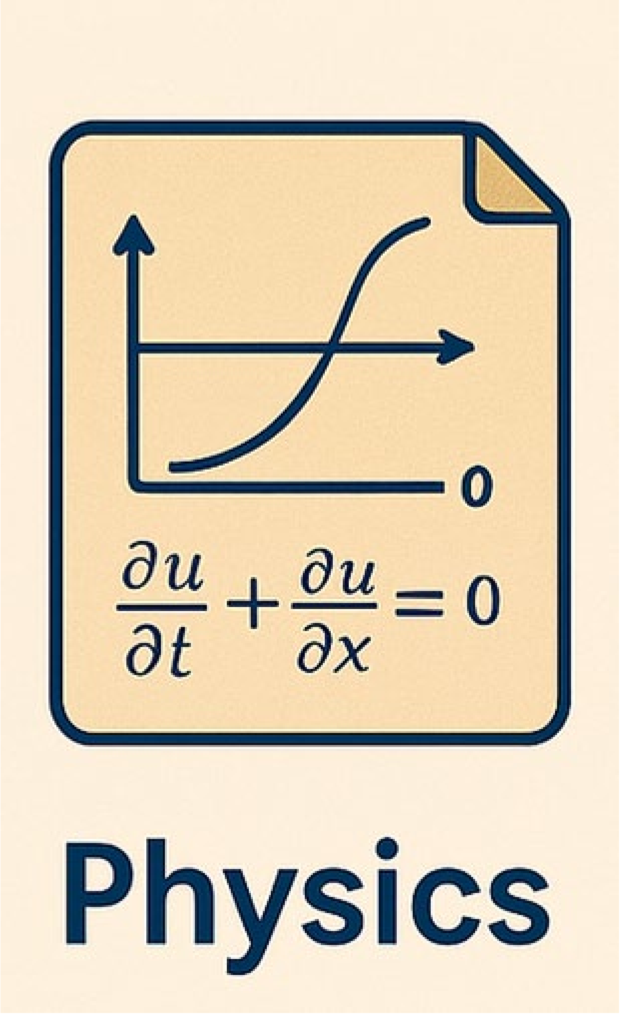
Future direction

Datasets



Using CLM and MIMICS simulations under different scenarios

Isotope-partitioned measurements



- Mass balance
- Different temperature sensitivities
- Substrate supply …

Take home message

- Classic NN: purely data-driven, but may misattribute effects from different drivers.
- PINN: guides temperature sensitivity toward biologically realistic values without sacrificing accuracy.

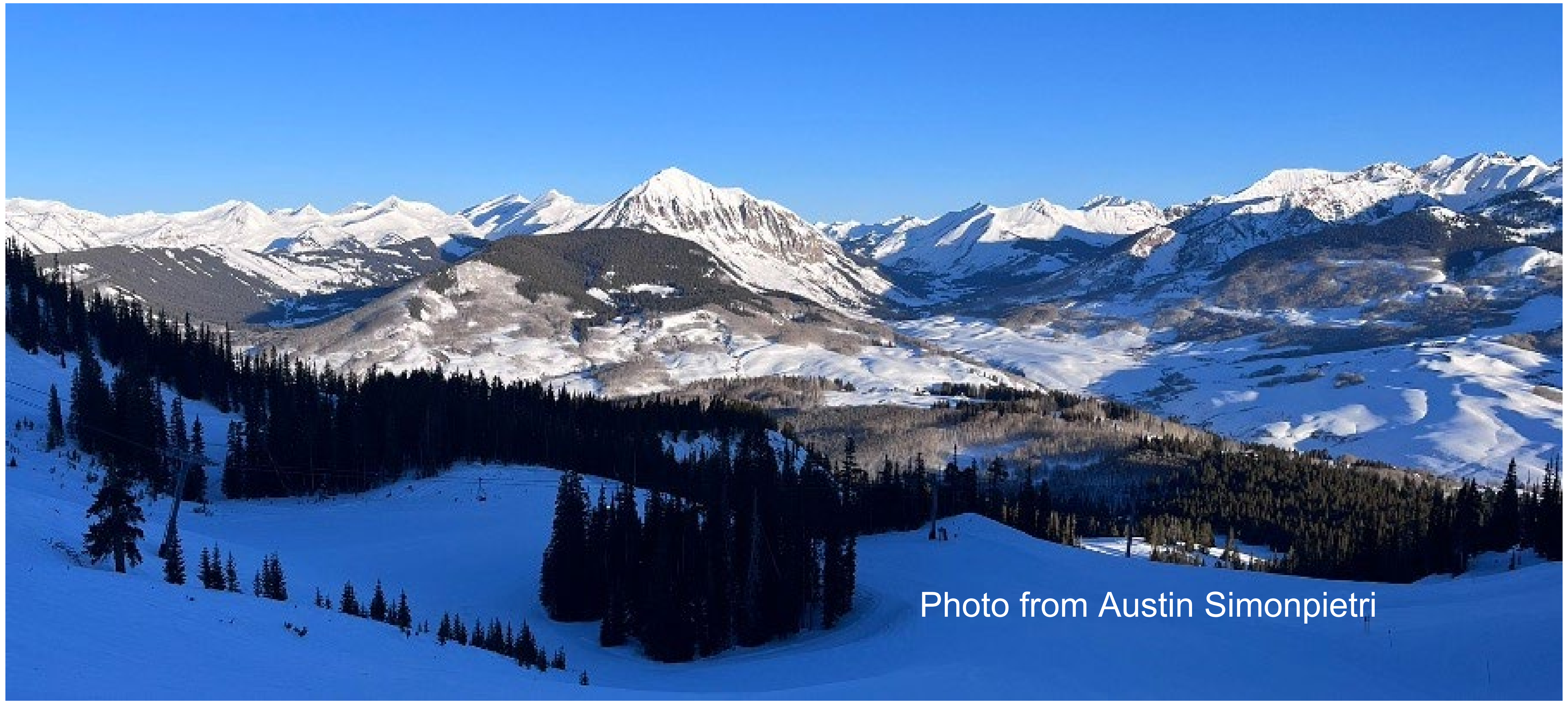


Photo from Austin Simonpietri

yujie.liu@nau.edu

Flux gradient approach

- Flux-gradient approach we are using is very sensitive to the estimated **soil diffusivity** , which is estimated from bulk density and volumetric water content at three depths. So, if heavy rain saturates the upper 1 cm of soil, CO2 may build up at all three depths, the CO2 gradient in the soil would be incorrectly converted to a flux estimate because we aren’t accounting for that saturated layer.