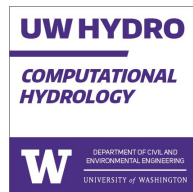


Searching for new physics: Using explainable AI to understand deep learned parameterizations of turbulent heat fluxes

Andrew Bennett & Bart Nijssen

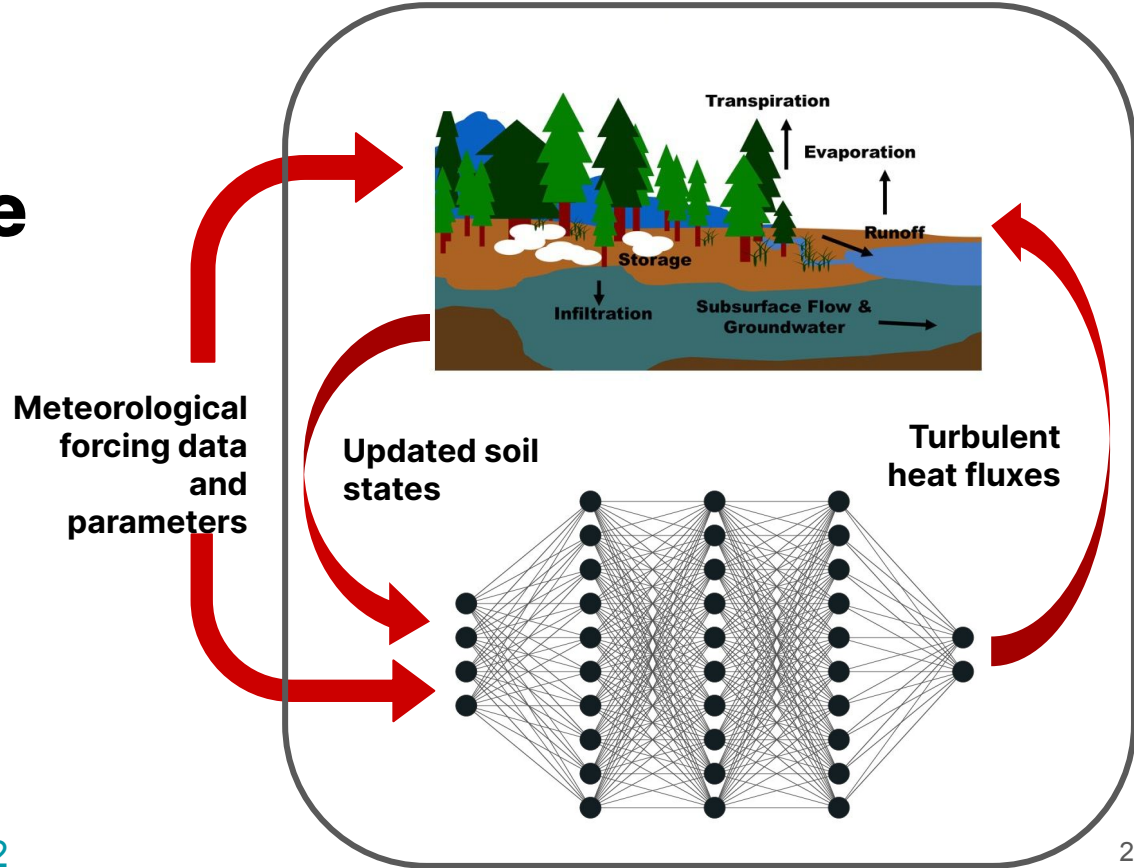
Email: andrbenn@uw.edu

Preprint link!



We previously showed that a coupled DL-PBHM approach can make better predictions of turbulent heat fluxes than a PBHM alone

Process based hydrologic model



Link to our previous work:

<https://doi.org/10.1002/essoar.10505081.2>

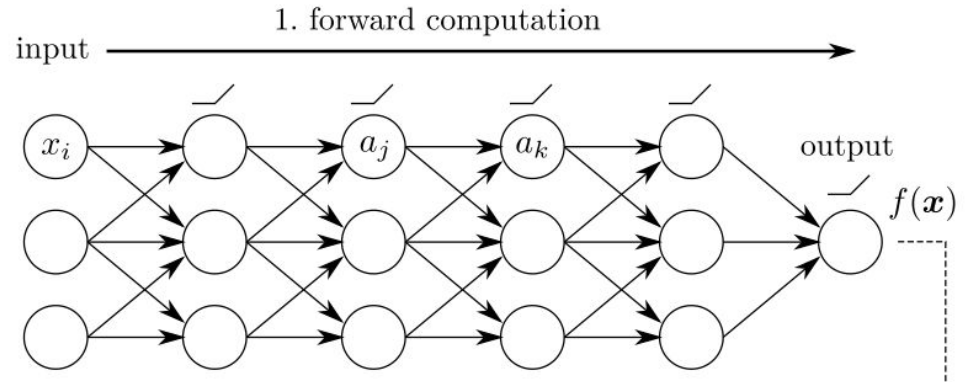
This led to two main questions:

- 1. Did the neural network learn something physically plausible?**

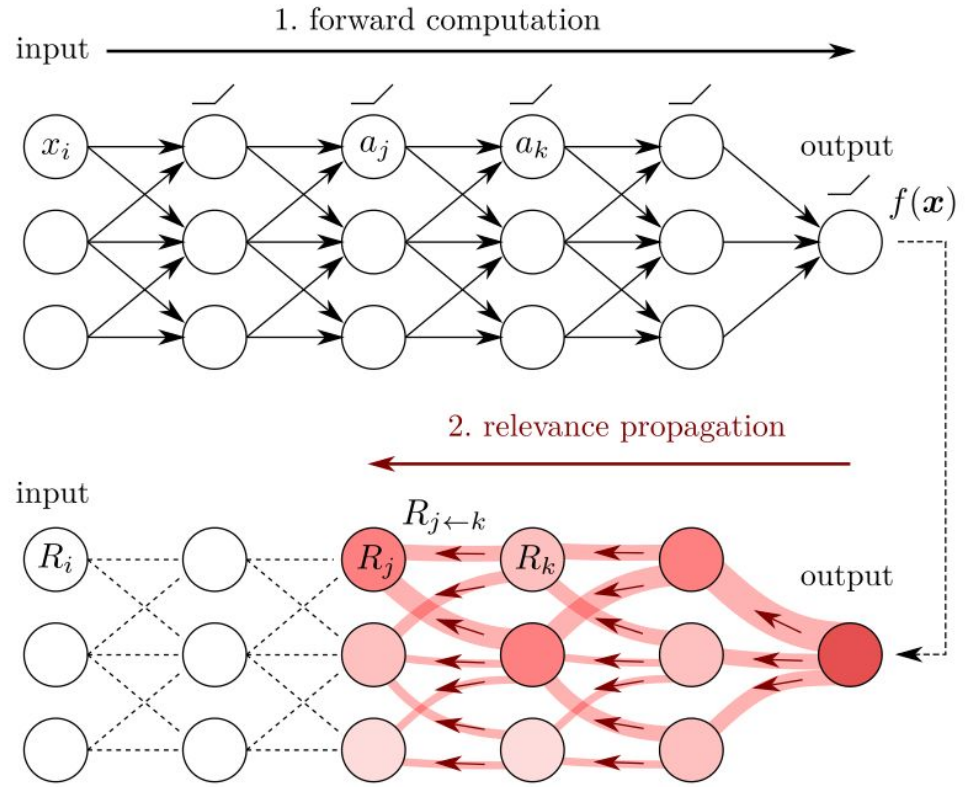
This led to two main questions:

- 1. Did the neural network learn something physically plausible?**
- 2. How did the network learn to generalize between sites?**

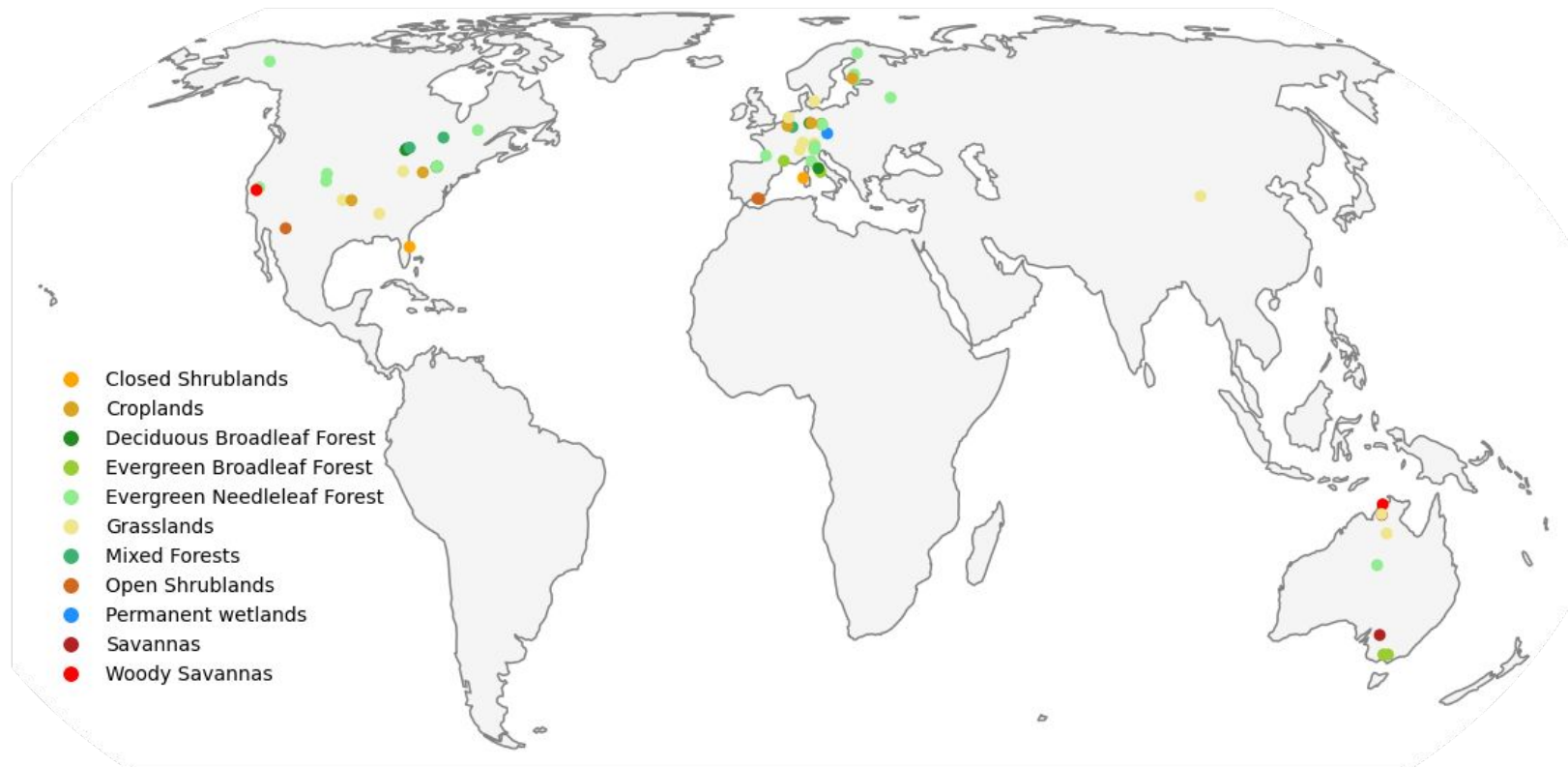
To explore these questions, we used a technique called Layerwise Relevance Propagation, or LRP



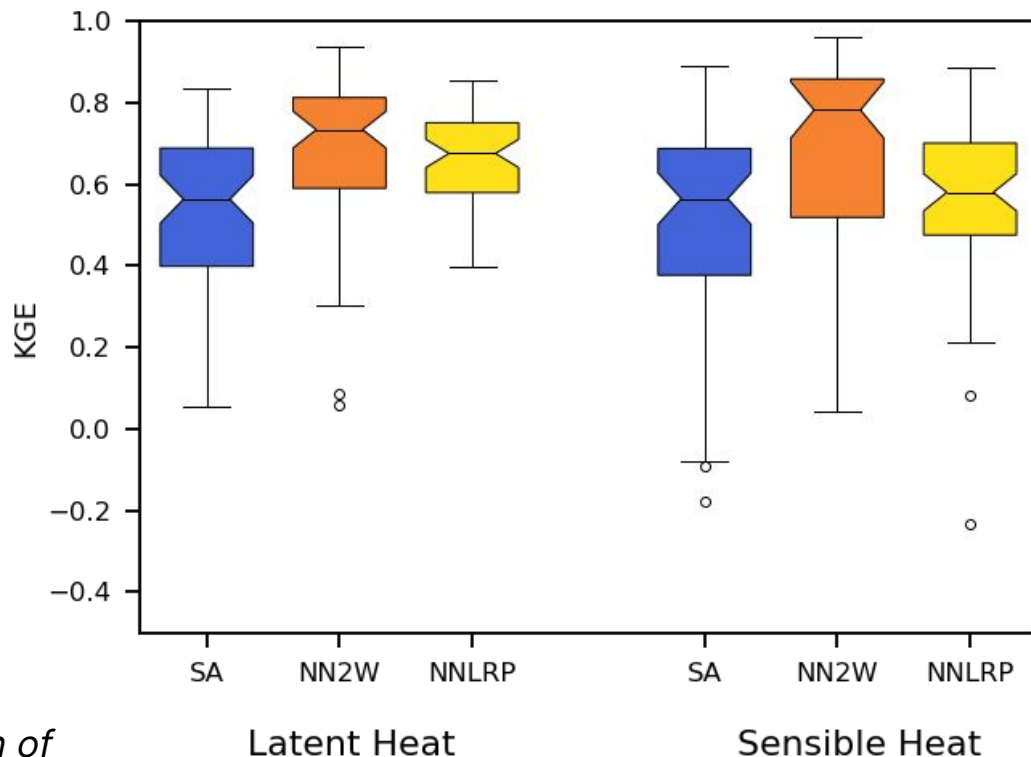
To explore these questions, we used a technique called Layerwise Relevance Propagation, or LRP



We trained on 60 FluxNet sites at 30 minute timescale without cross validation



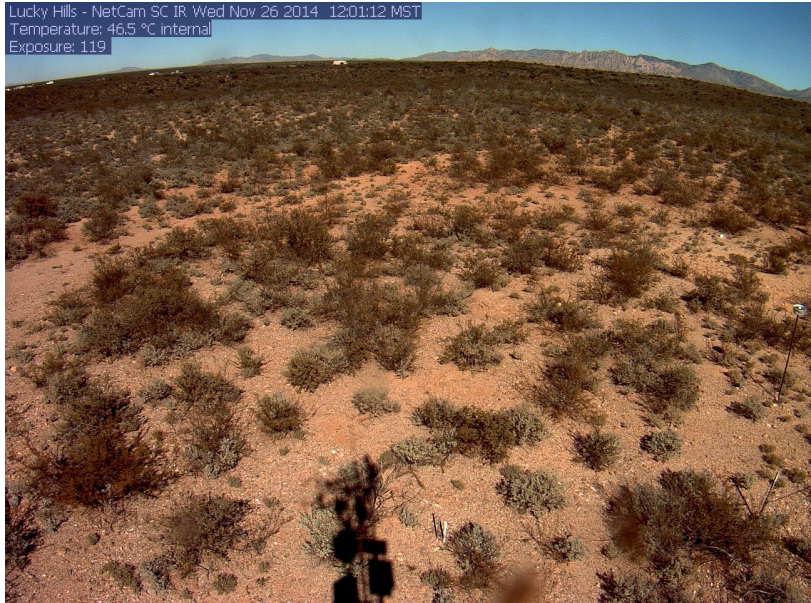
Our simplified model, NNLRP, still performs well*



**Note that this should be taken with a grain of salt because we did not cross-validate, but predictive performance isn't actually our goal*

Looking specifically at two locations

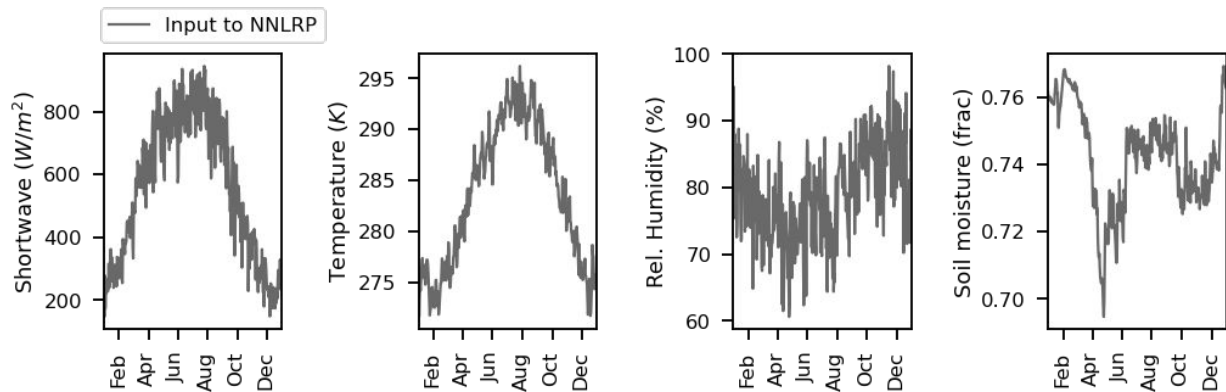
US-Whs



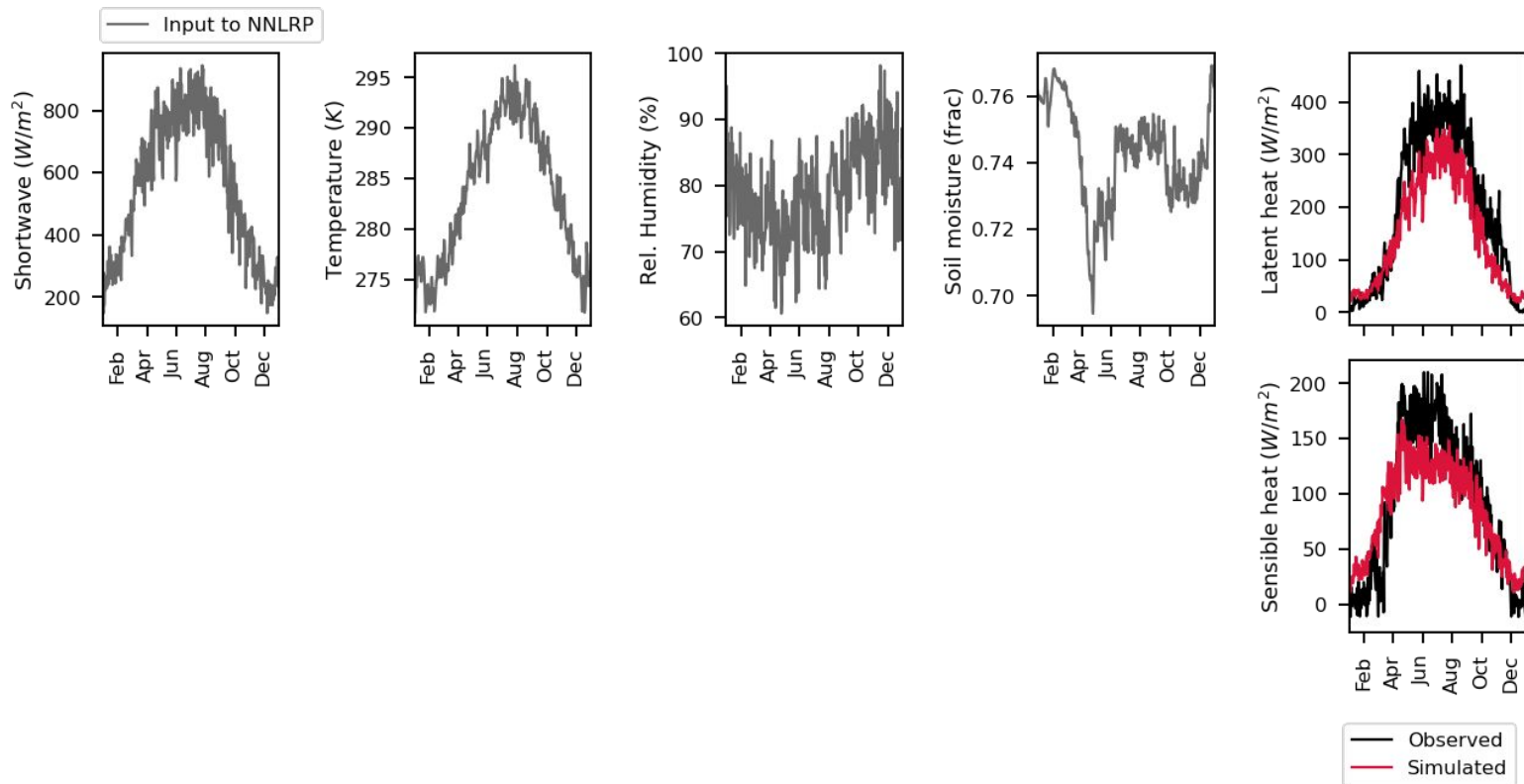
CH-Fru



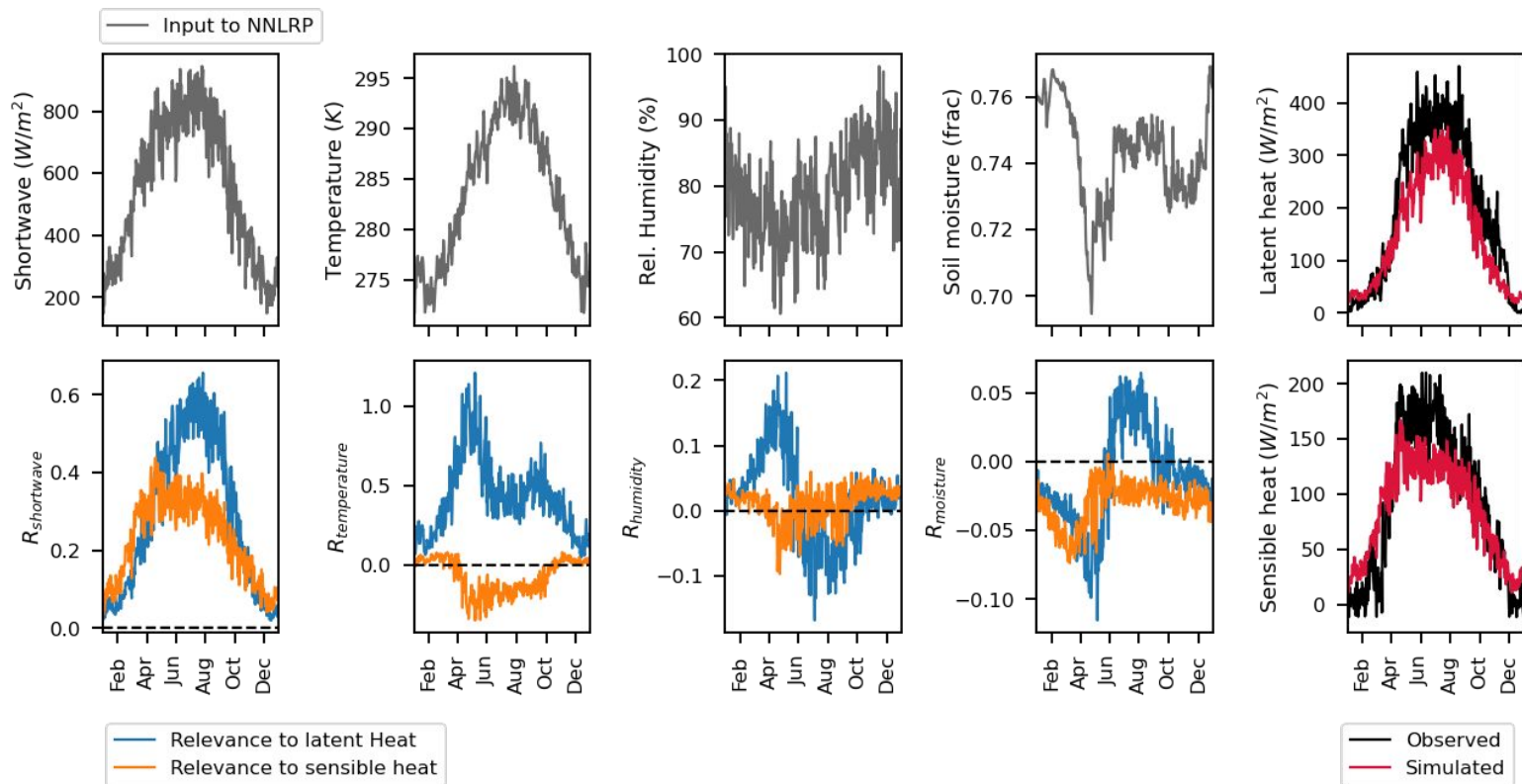
First, CH-Fru, a grasslands site near the Swiss Alps



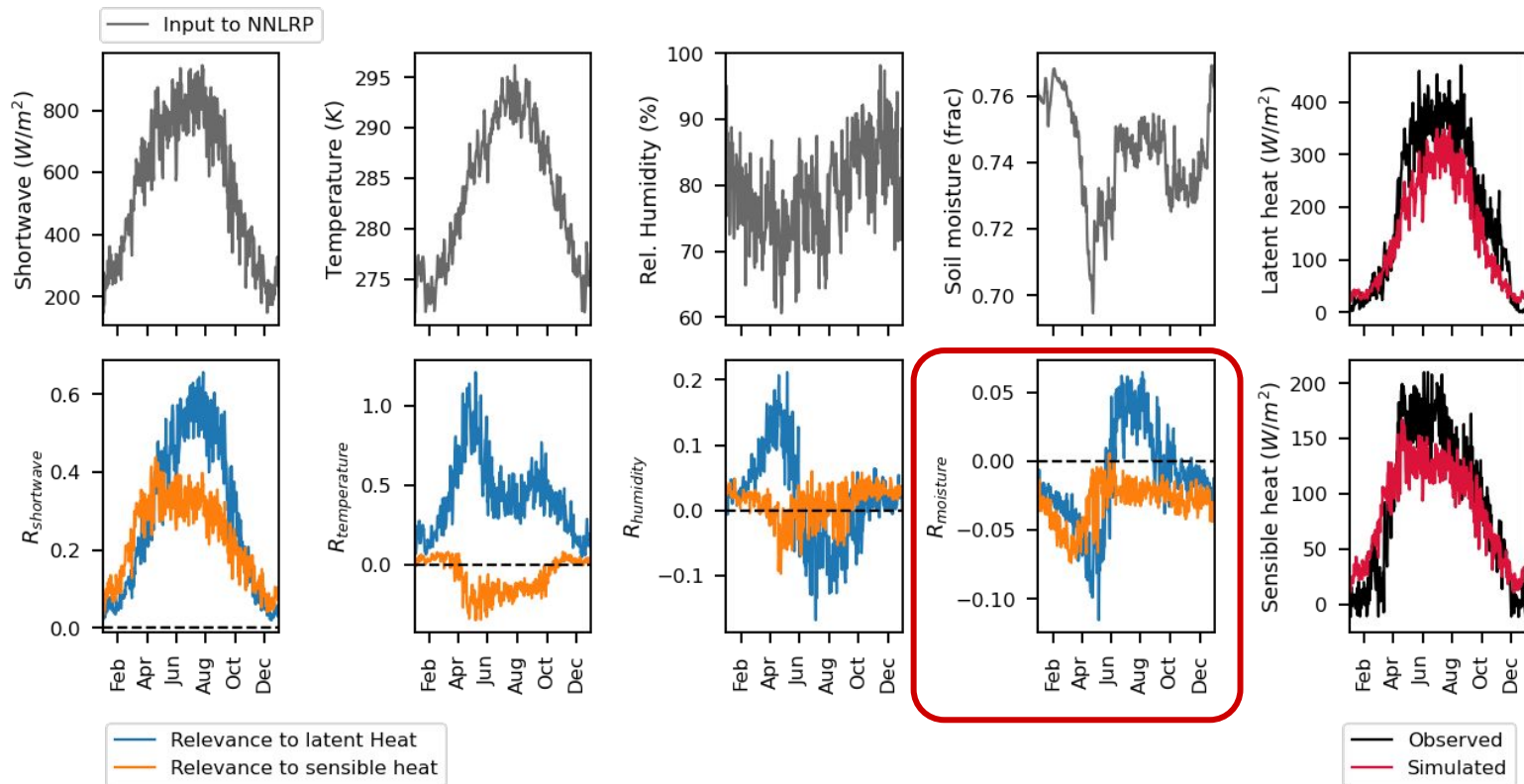
First, CH-Fru, a grasslands site near the Swiss Alps



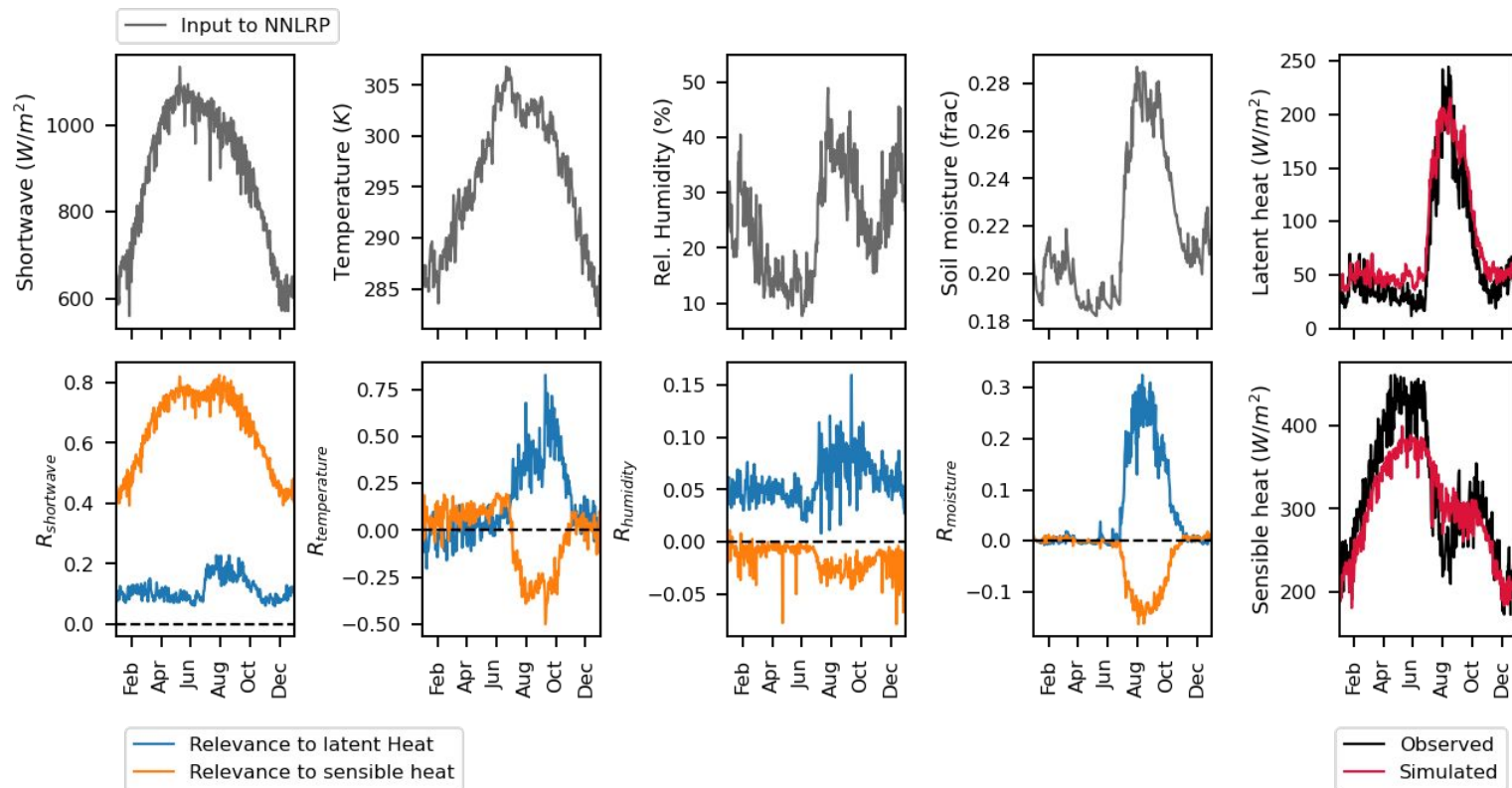
First, CH-Fru, a grasslands site near the Swiss Alps



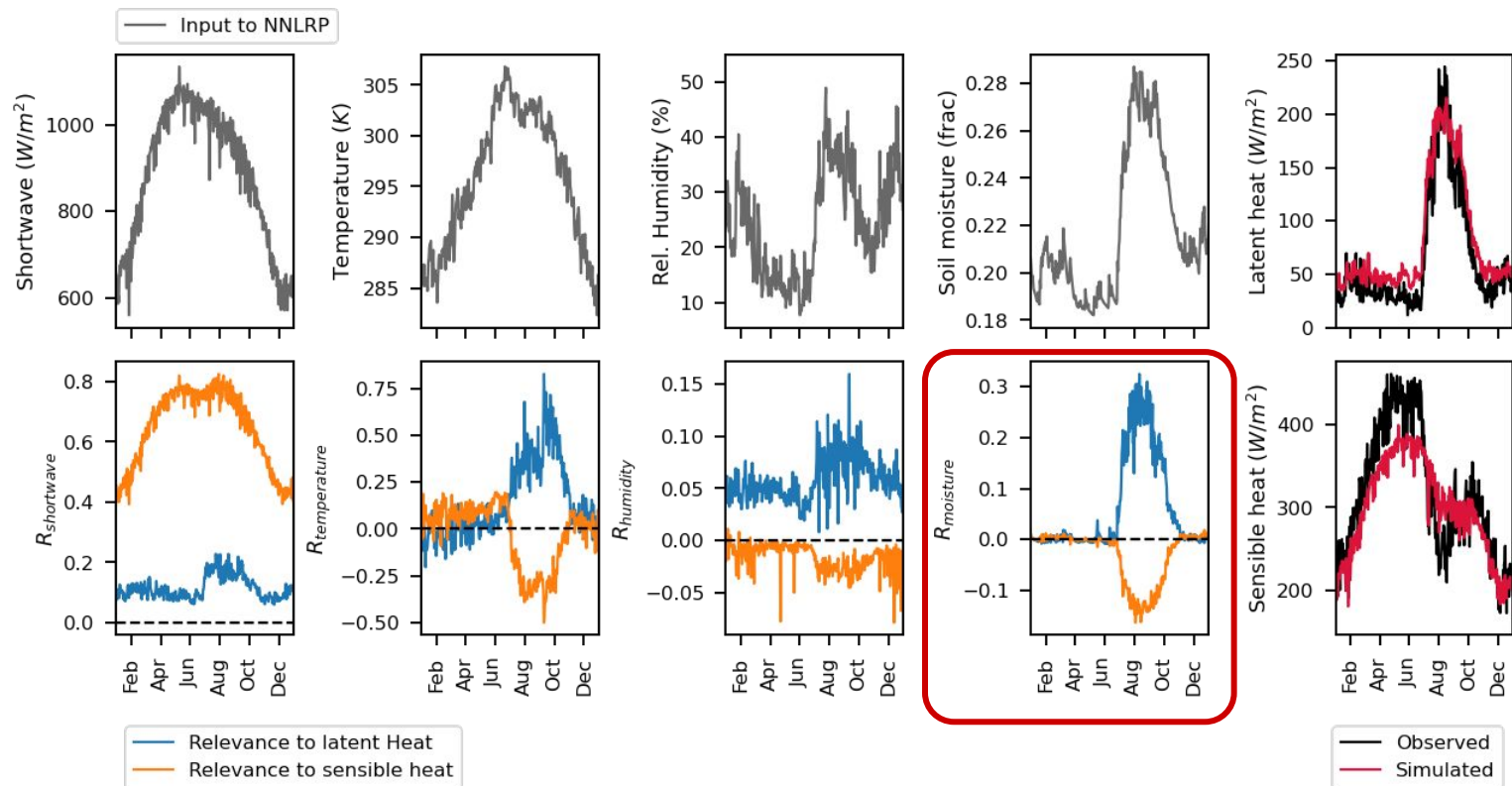
First, CH-Fru, a grasslands site near the Swiss Alps



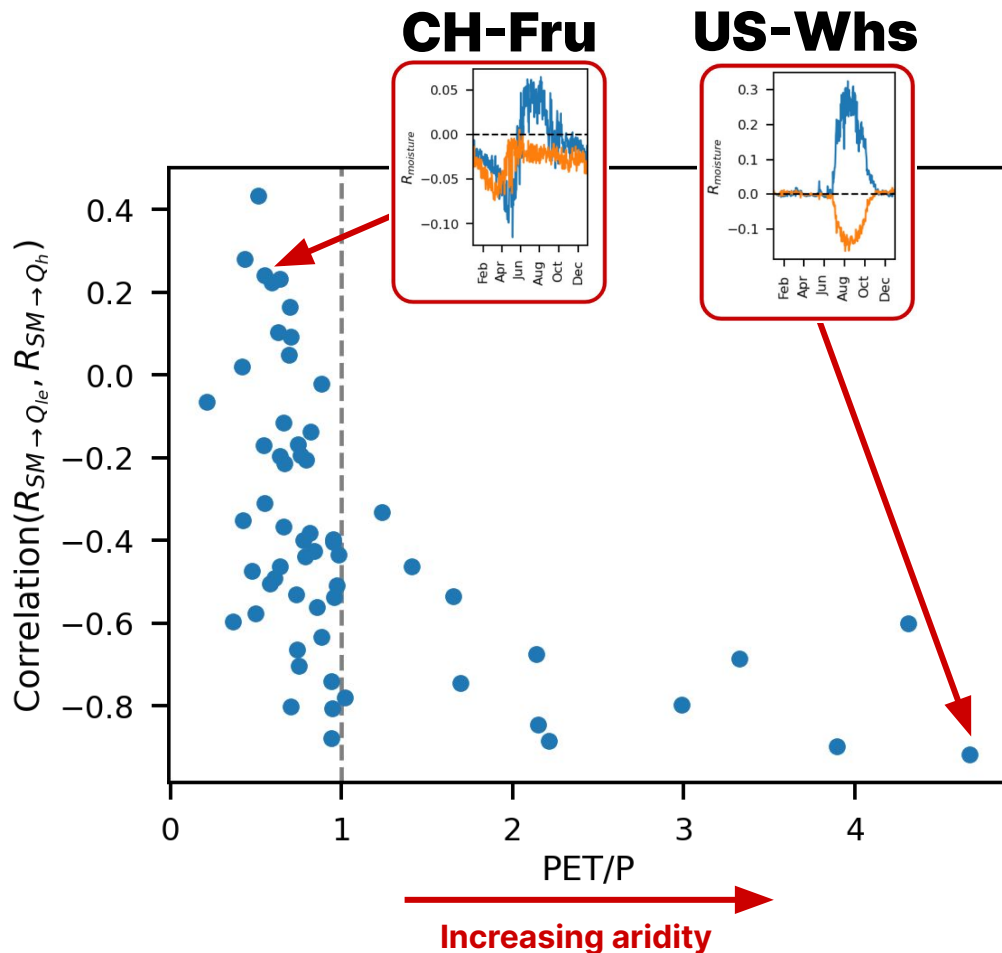
And similarly for US-Whs, a shrubland site in the Chihuahuan desert near Tucson, AZ



And similarly for US-Whs, a shrubland site in the Chihuahuan desert near Tucson, AZ

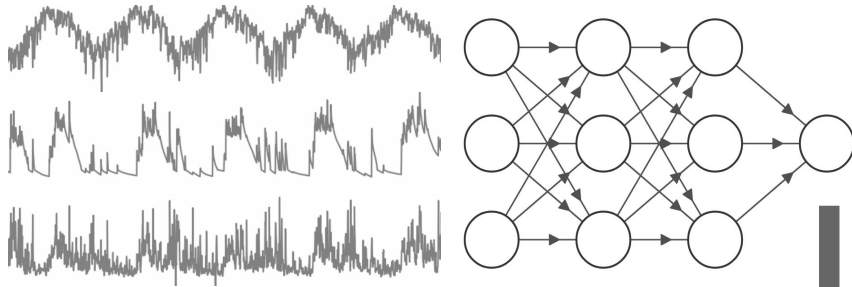


**NNLRP learned
to partition
between latent
and sensible
heat at arid
sites based on
soil moisture**



NNLRP learns a good input-output relationship

Inputs

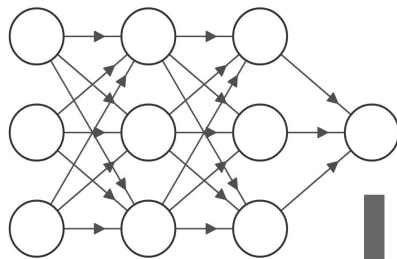
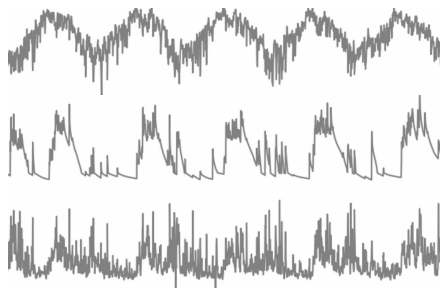


Outputs

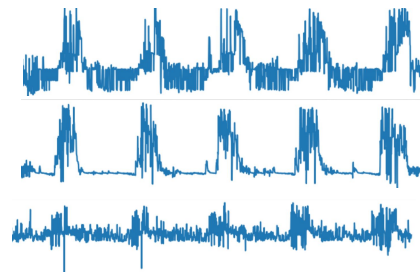
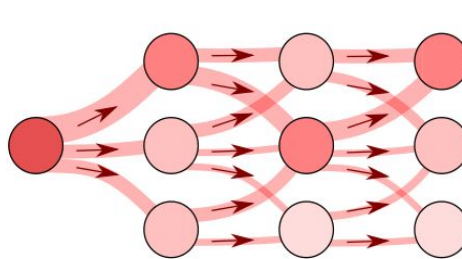


LRP shows us some aspects of *what* NNRLP learned

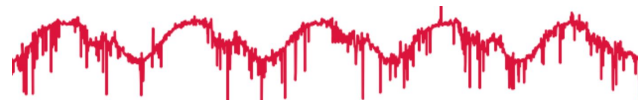
Inputs



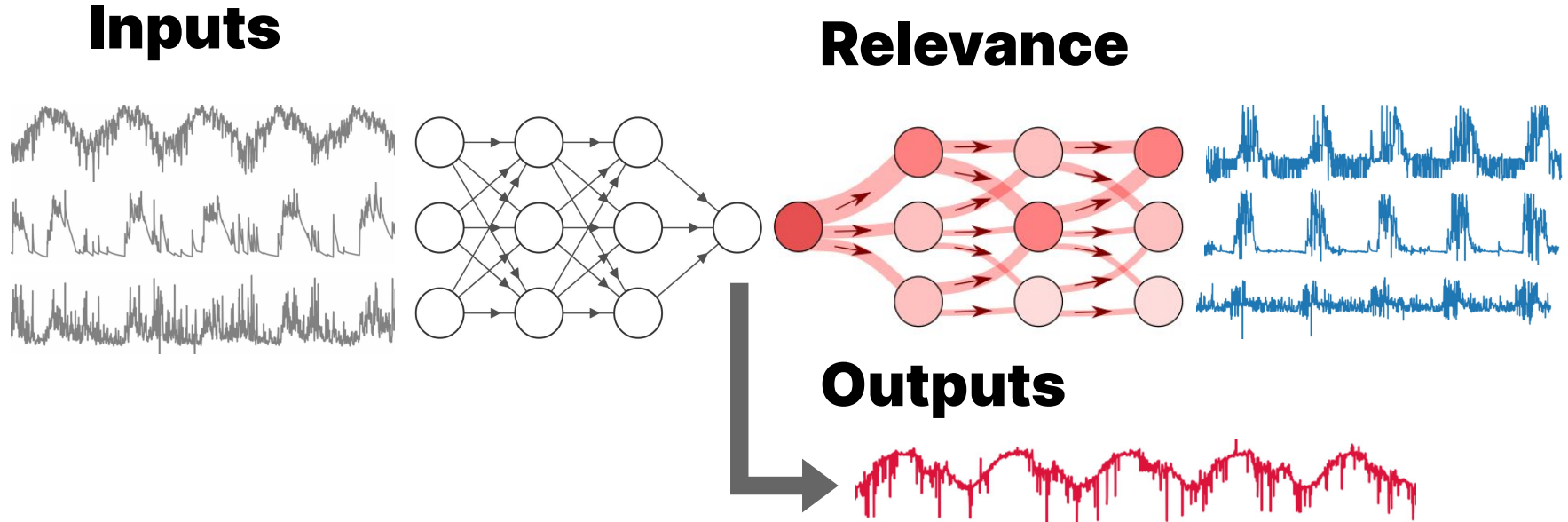
Relevance



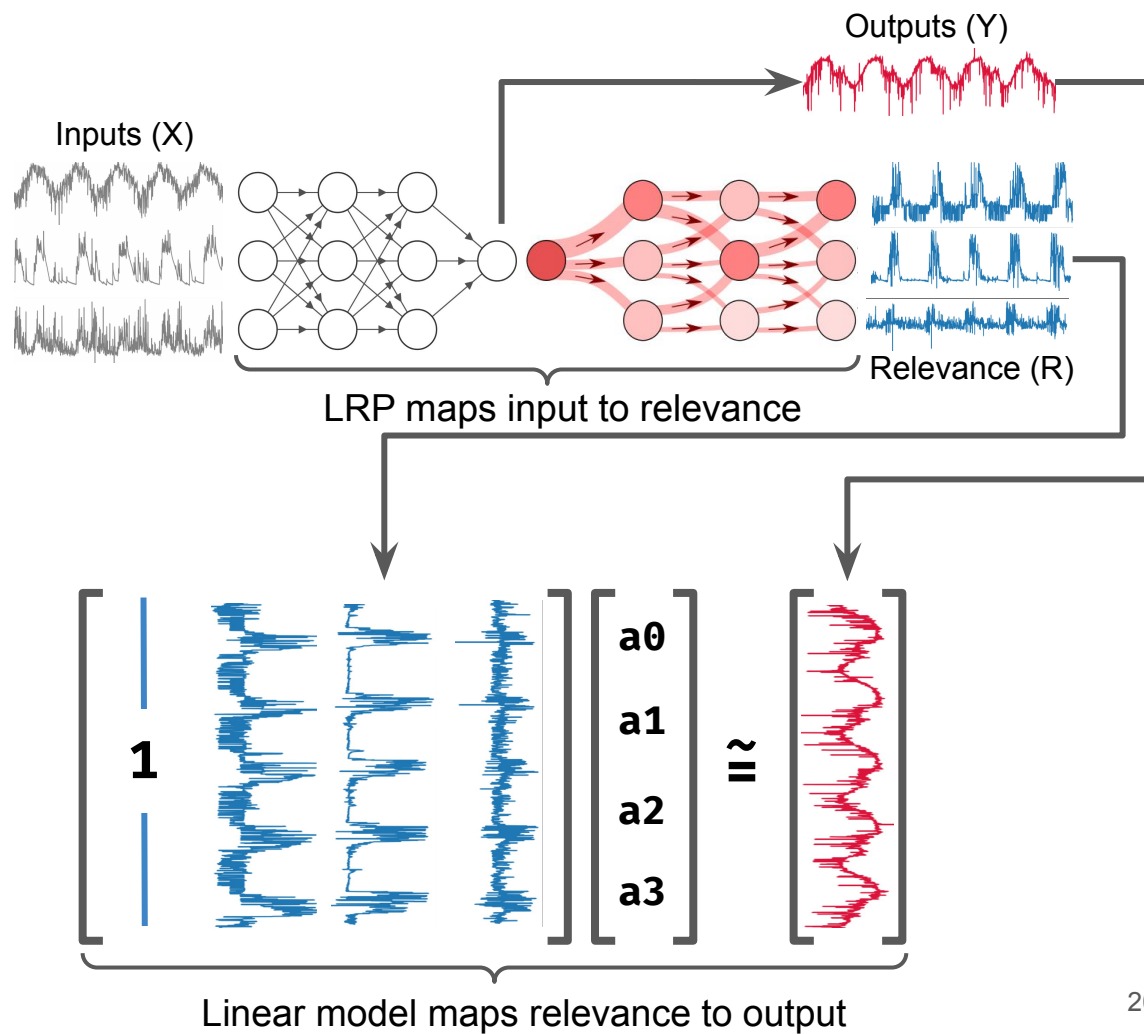
Outputs



Could relevance provide a good way of understanding *how* it learned?

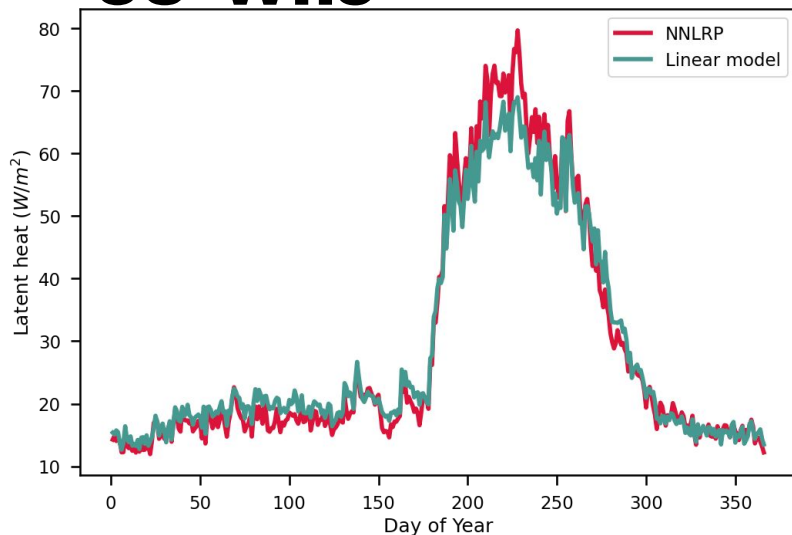


We fit a linear model with the relevance from each site

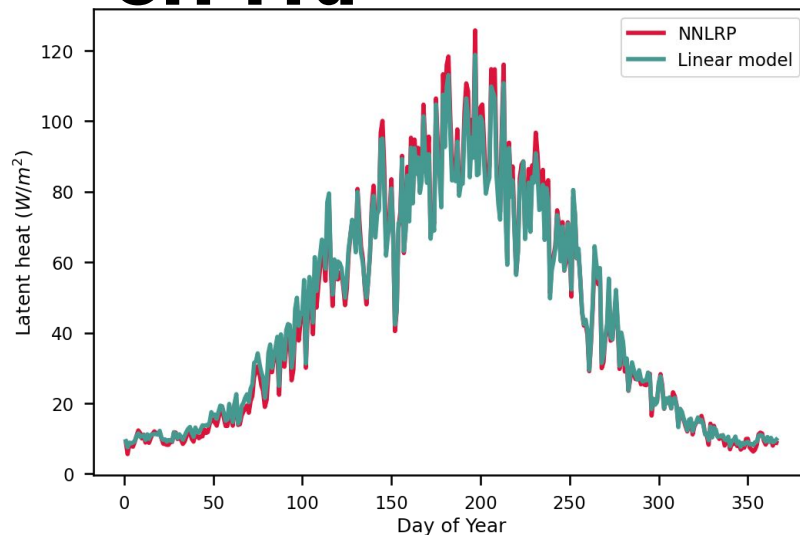


The linear models produce good approximations of NNLRP

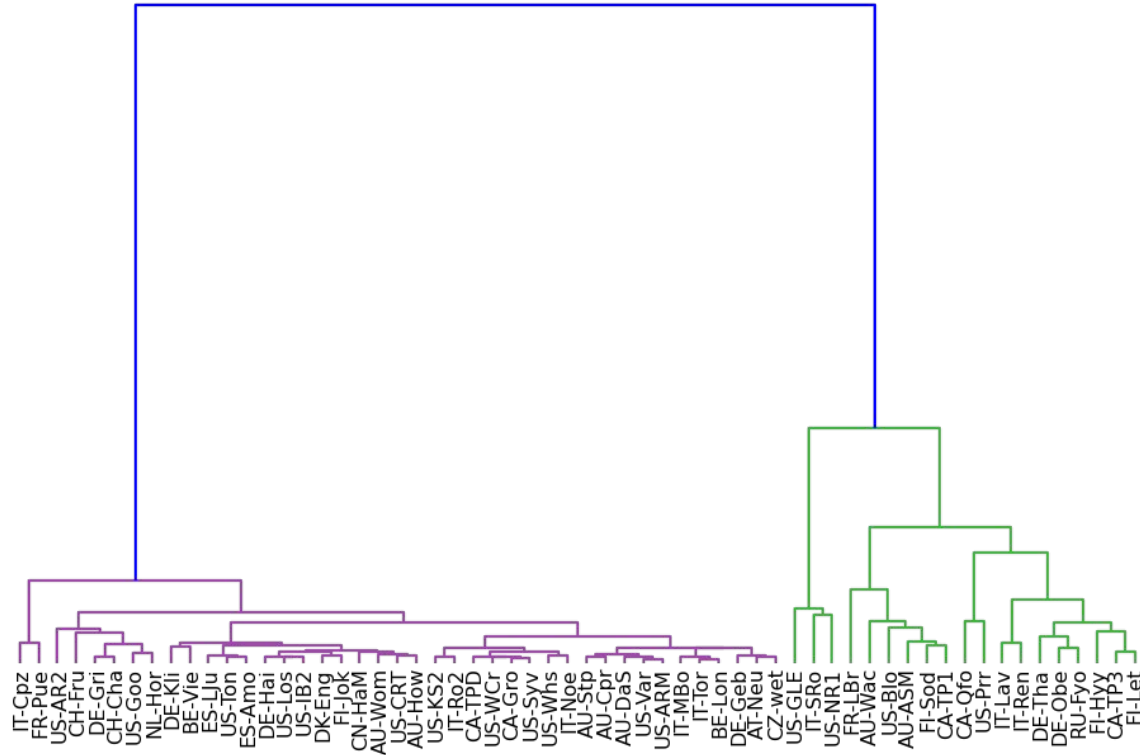
US-Whs



CH-Fru



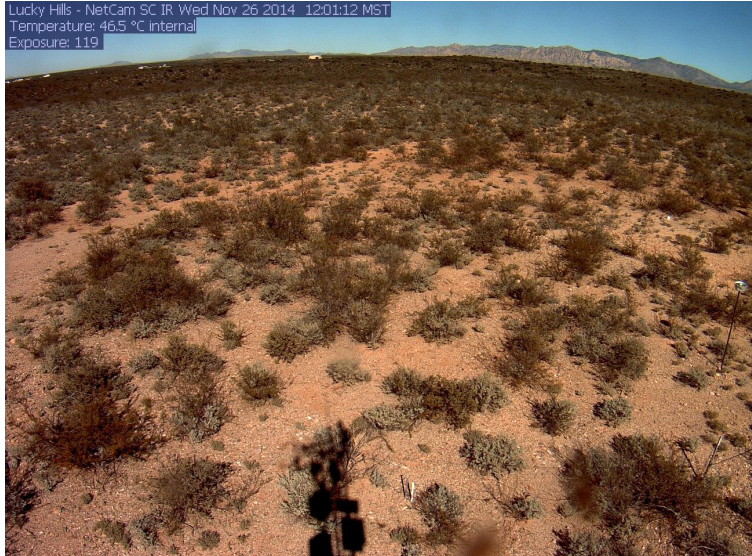
Hierarchical clustering reveals two main groups: **Evergreen needleleaf forest**, and... **everything else**



**What if we fit a linear model
from one site and then
apply it to another site?**

Let's fit a model at US-Whs and apply it at CH-Fru

US-Whs

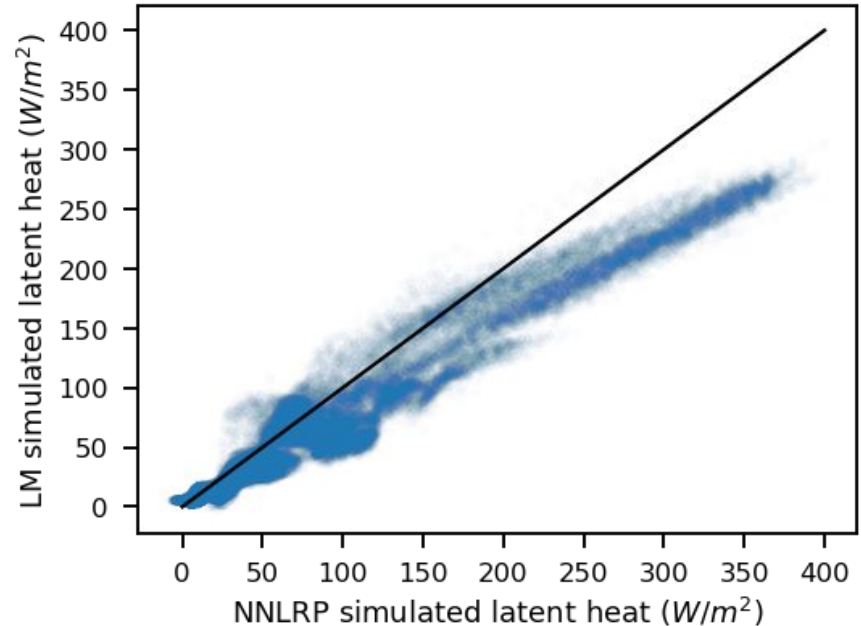


CH-Fru



**Let's fit a model at US-Whs
and apply it at CH-Fru**

It's pretty good!



Now let's fit at FI-Hyy and apply at CH-Fru

FI-Hyy

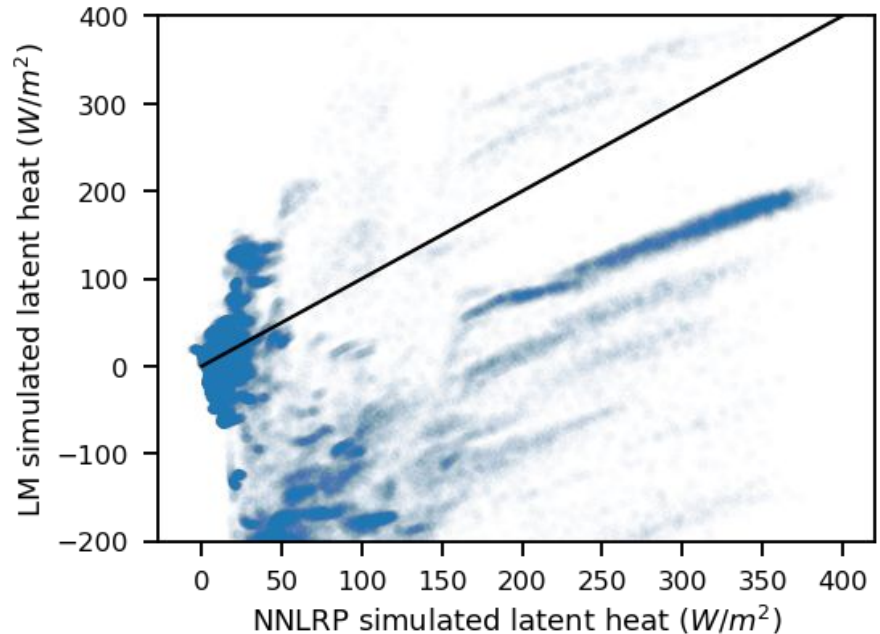


CH-Fru

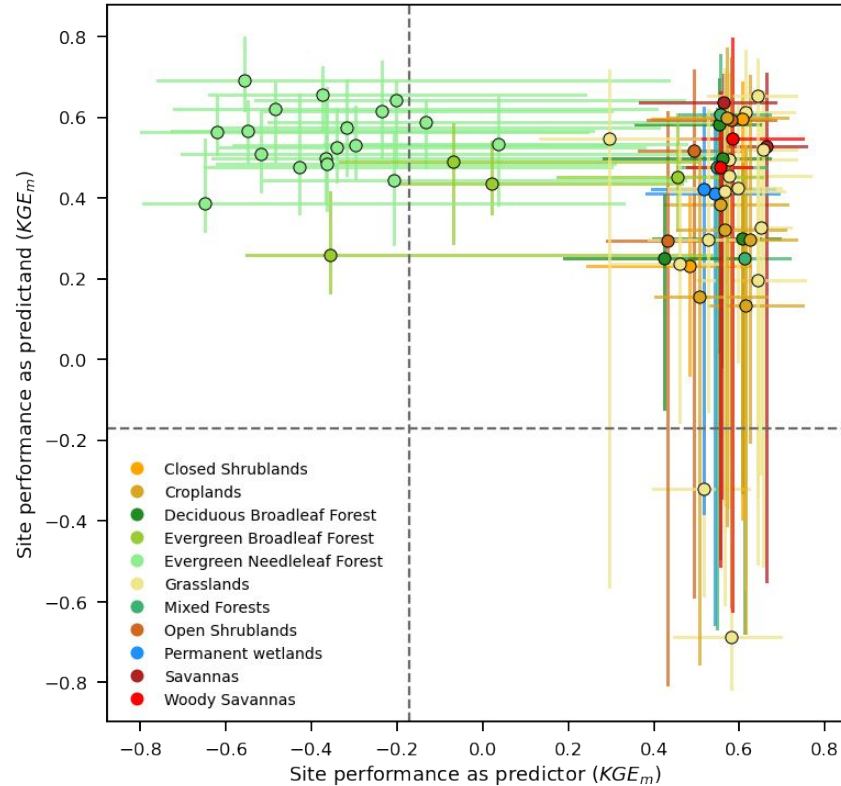


**Now let's fit at FI-Hyy
and apply at CH-Fru**

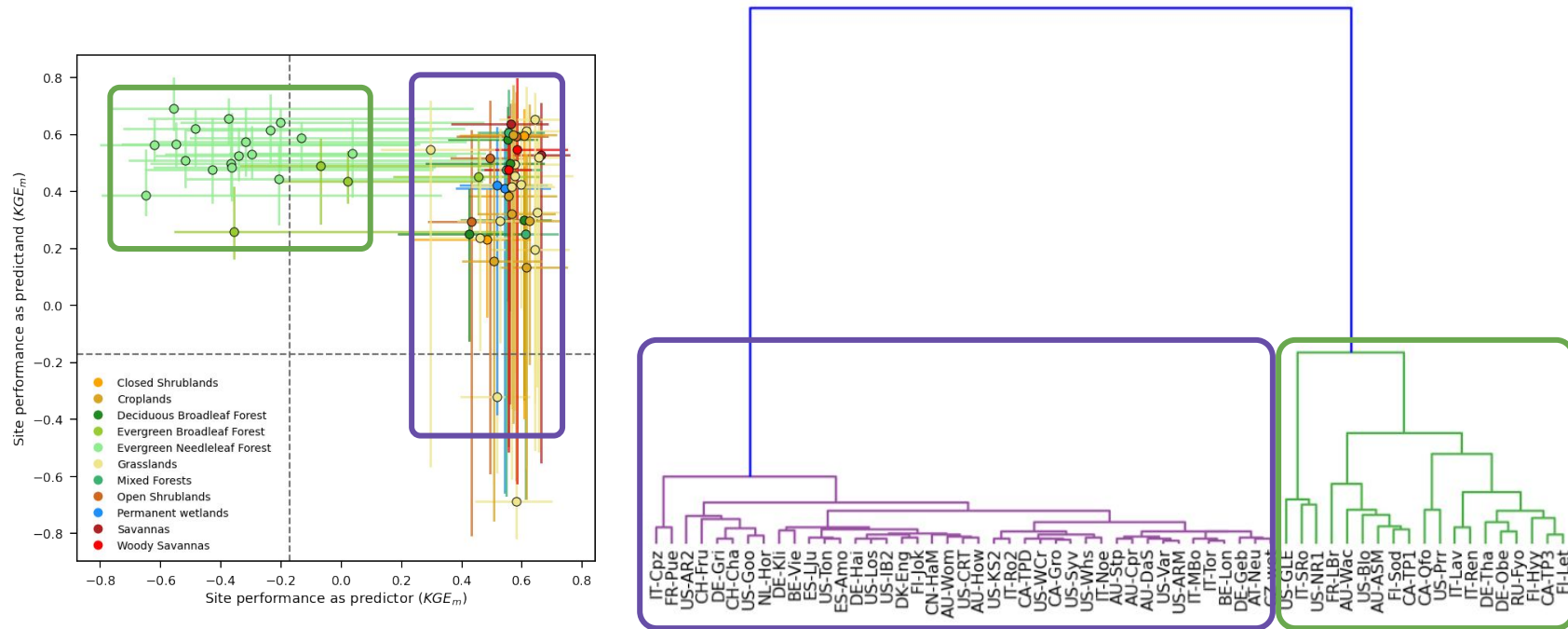
That didn't work...



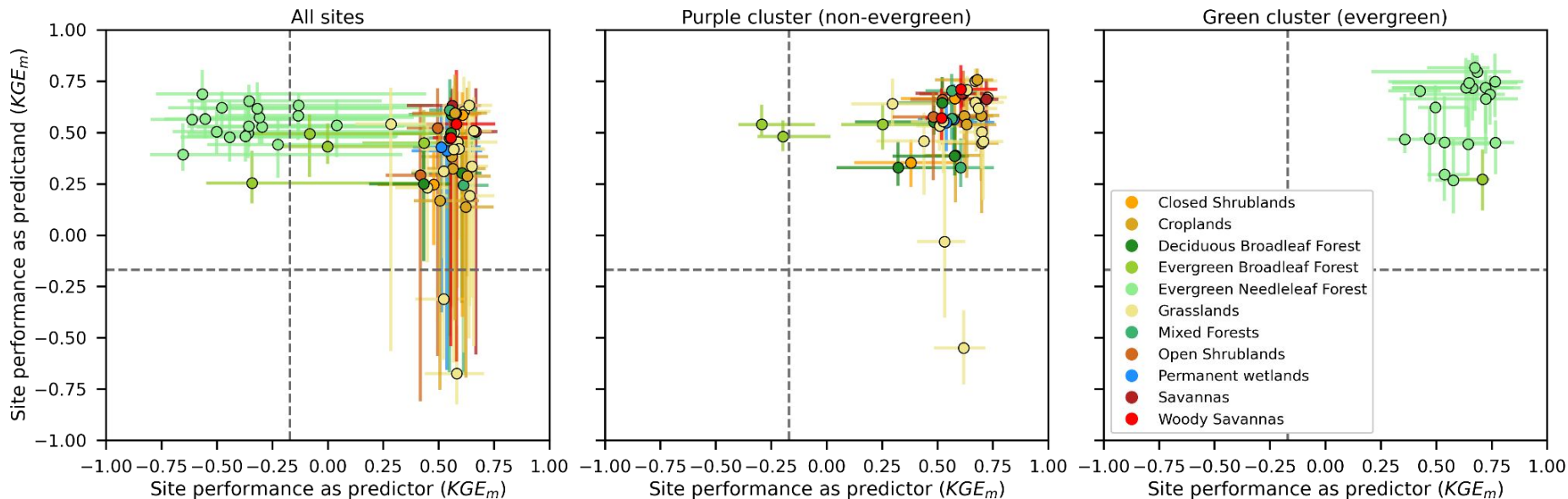
We did this for every pair of sites



Recall the two clusters



Separating the clusters out shows very good performance “in-cluster”



Thanks for listening!

Some conclusions/where do we go from here?

Coupling ML and physics for hydrologic modeling works better than either in isolation

Using explainable methods and designing experiments around them can help us to understand how and why ML works/doesn't work

There is huge potential for using ML in innovative ways that incorporate scientific knowledge

Link to coupled model development preprint!



Link to LRP analysis preprint!

