A Machine Learning Surface Layer Parameterization for WRF

David John Gagne

NCAR CISL/RAL

Collaborators: Tyler McCandless, Branko Kosovic, Amy DeCastro, Thomas Brummet, Sue Ellen Haupt, Rich Loft, Bai Yang

NCAR

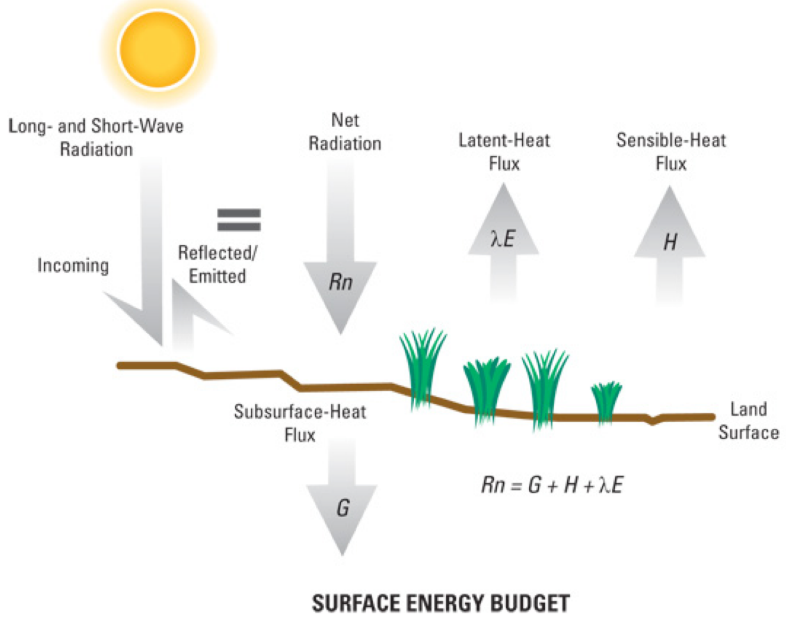
JUNE 11, 2019

NSF

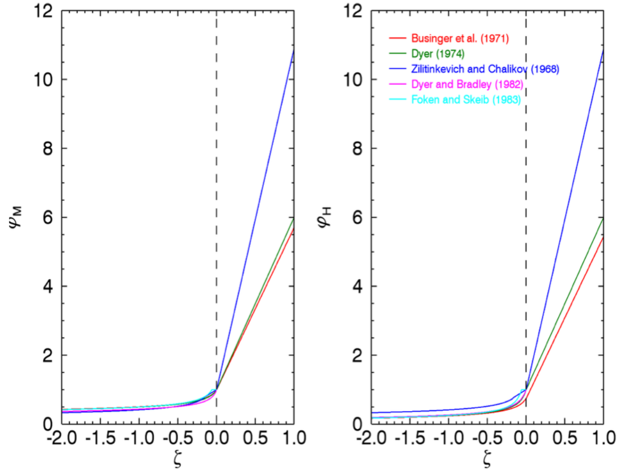
This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

Motivation: Surface Layer Parameterization

- The energy transfer (flux) from the atmosphere to the land surface is modeled by the surface layer parameterization.
- The flux depends on gradients in wind speed, temperature and moisture between the lower-atmosphere and air just above the surface.
- Monin-Obukhov similarity theory is used to determine model surface fluxes and stresses.
- Stability functions Φ_M (momentum) and Φ_H (heat) are determined experimentally.
- Stability functions come from field studies under nearly ideal atmospheric flow conditions characterized by horizontally homogeneous, flat terrain and stationarity. **However, the stability functions show a large amount of variation.**



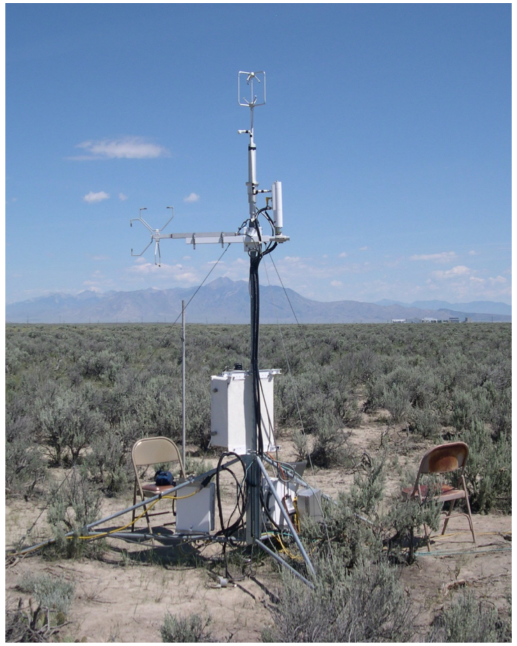
https://nevada.usgs.gov/et/measured.htm



Motivation: Surface Layer Methods

- Regression is commonly used to flux estimate stability functions
- Instead, we use machine learning algorithms to develop models relating surface stresses and fluxes to wind and temperature profiles.
- Most of the previous field studies used to determine stability functions were only a few months in length.
- To develop robust machine learning models, we need long observational records.
- We have therefore selected two data sets that provide multiyear records
- Fit random forests to each site to predict friction velocity, sensible heat flux, and latent heat flux

Cabauw, Netherlands
KNMI Mast
213 m tower
Data from 2003-2017



Scoville, Idaho, USA
FDR Tower
Flux tower
Data from 2015-2017

Input and Output Variables

Common Variables	Heights (Idaho/Cabauw)
Virtual Potential Temperature (K)	10 m, 15 m/20 m
Mixing Ratio (g/kg)	10 m, 20 m
Solar Radiation (w m-2)	Surface
Wind Speed (m/s)	10 m, 15 m/20 m
Bulk Richardson number	10 m- 0 m
Pressure (hPa)	Surface
Saturation Mixing Ratio (%)	10 m
Moisture Availability (%)	5 cm/3 cm
Skin Temperature (K)	0 m
Skin Saturation Mixing Ratio (g/kg)	0 m
Solar Zenith Angle (degrees)	0 m

Output equations

$$\tau = \rho u_*^2$$

$$H = -\rho c_p u_* \theta *$$

$$LH = L_e \rho u_* q_*$$

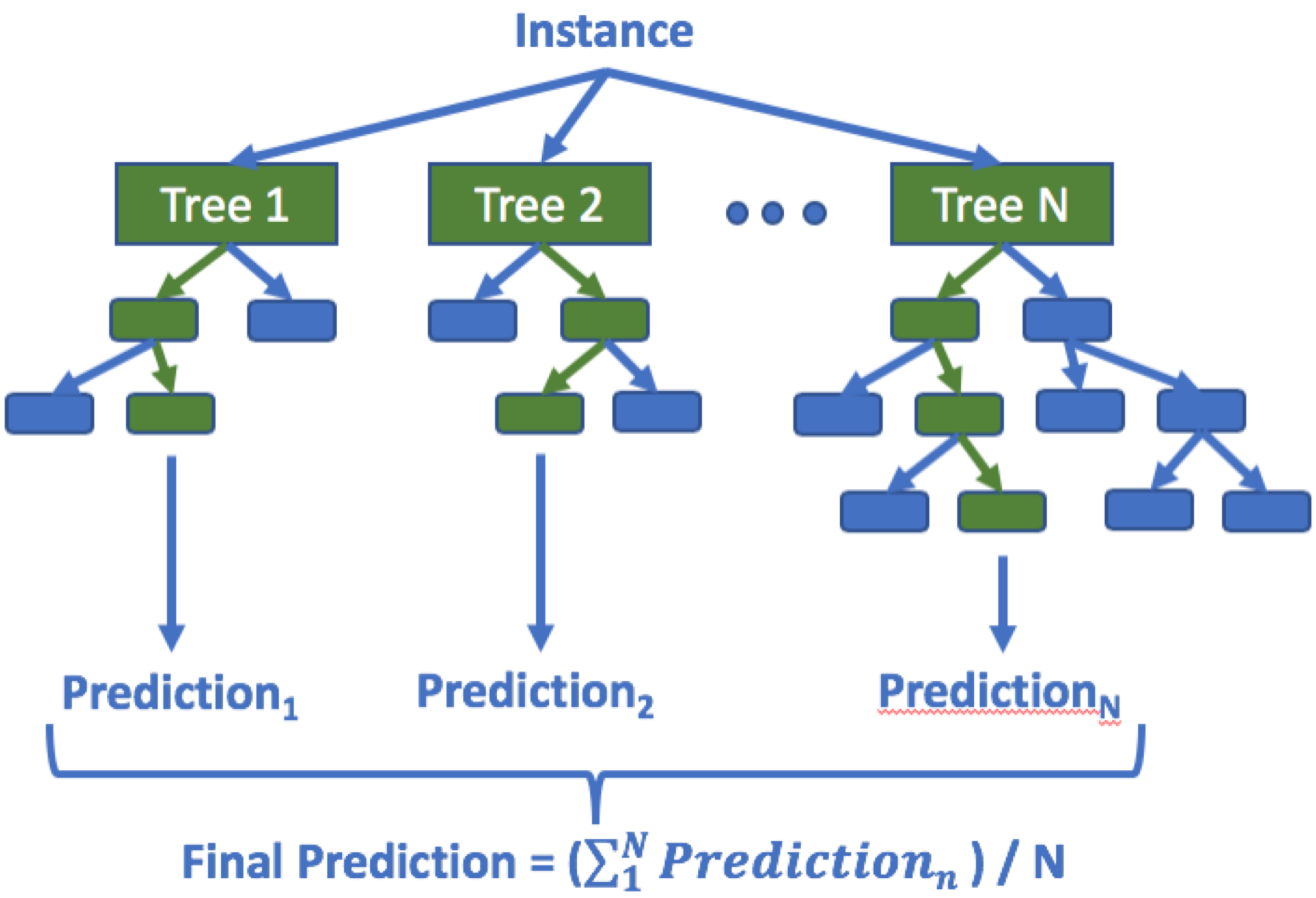
Predictands

u*=Friction velocity

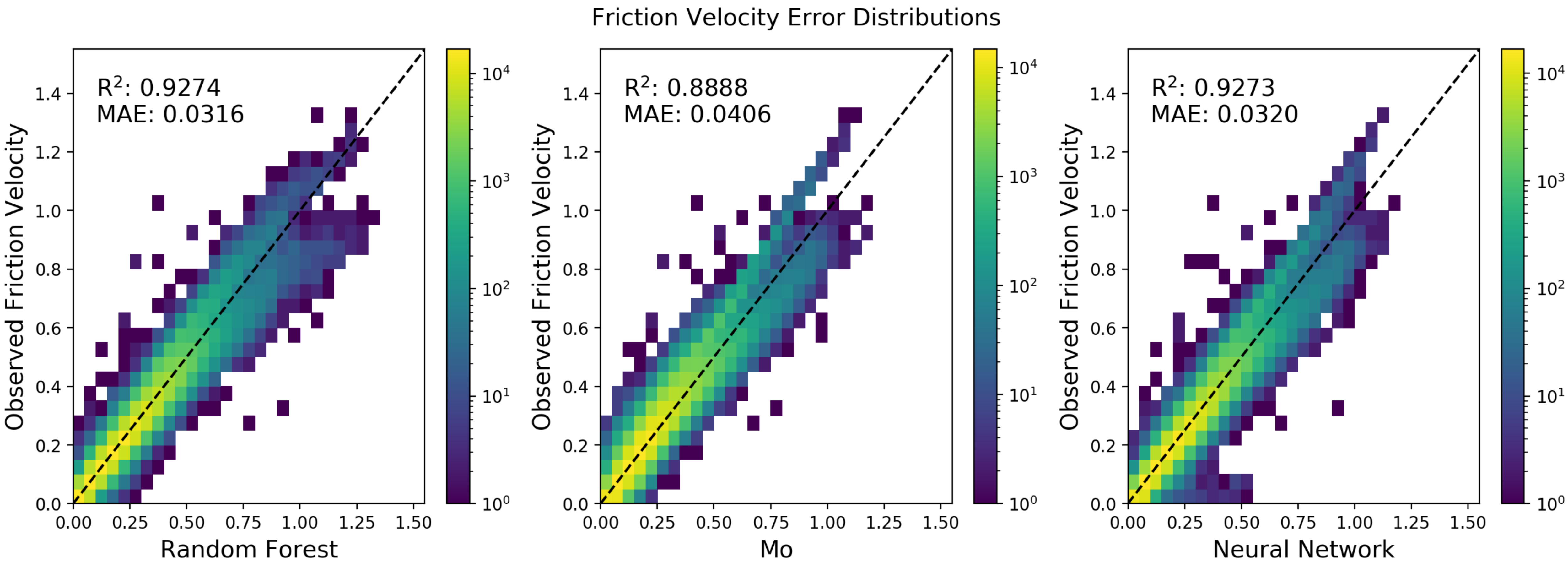
θ*=Temperature scale

q*=Moisture scale

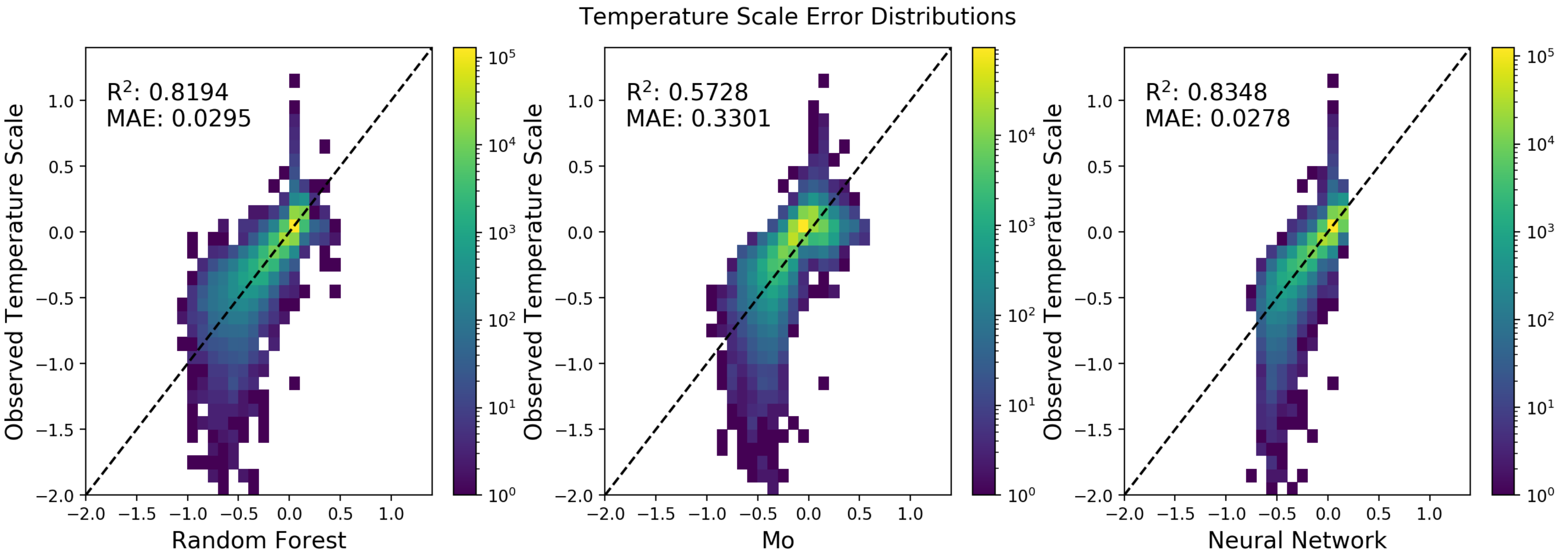
Random Forest



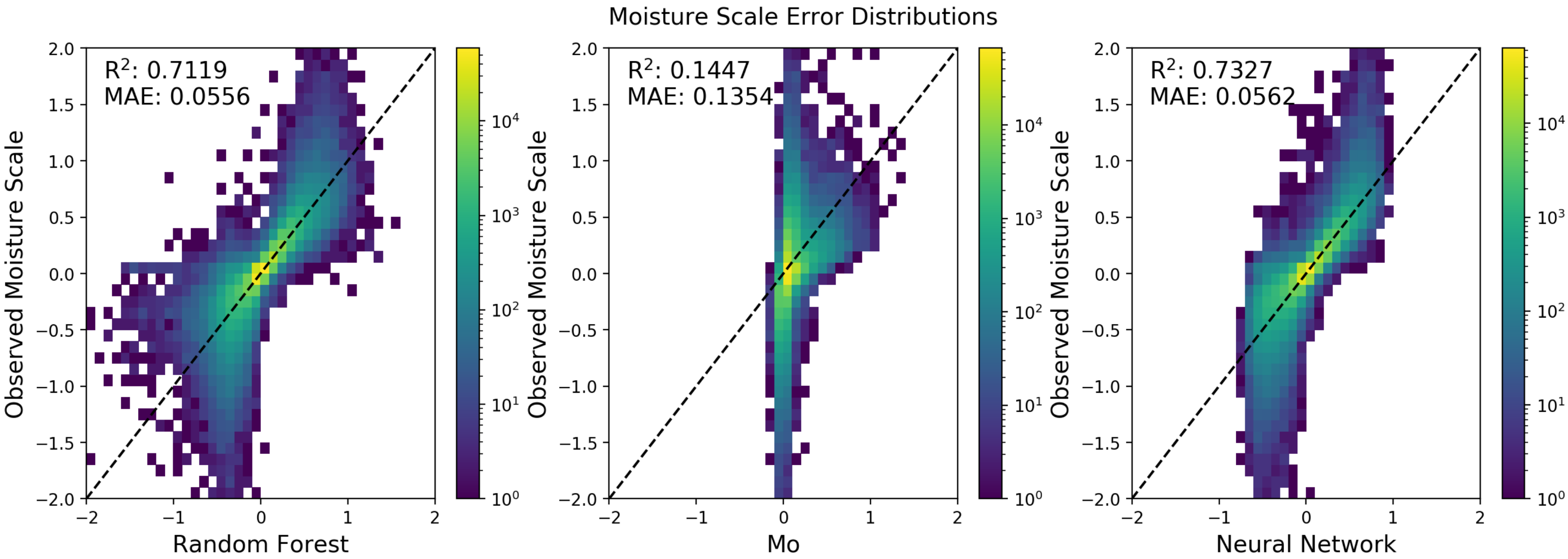
Offline Results: Friction Velocity



Offline Results: Temperature Scale



Offline Results: Moisture Scale



Cross-Testing ML Models

Idaho Test Dataset	R^2 Friction Velocity	R^2 Temperature Scale	R^2 Moisture Scale	MAE Friction Velocity	MAE Temperature Scale	MAE Moisture Scale
MO Similarity	0.85	0.42		0.077	0.203	
RF Trained on Idaho	0.91	0.80	0.41	0.047	0.079	0.023
RF Trained on Cabauw	0.88	0.76	0.22	0.094	0.139	0.284

Cabauw Test Dataset	R^2 Friction Velocity	R^2 Temperature Scale	R^2 Moisture Scale	MAE Friction Velocity	MAE Temperature Scale	MAE Moisture Scale
MO Similarity	0.90	0.44	0.14	0.115	0.062	0.135
RF Trained on Cabauw	0.93	0.82	0.73	0.031	0.030	0.055
RF Trained on Idaho	0.90	0.77	0.49	0.074	0.049	0.112

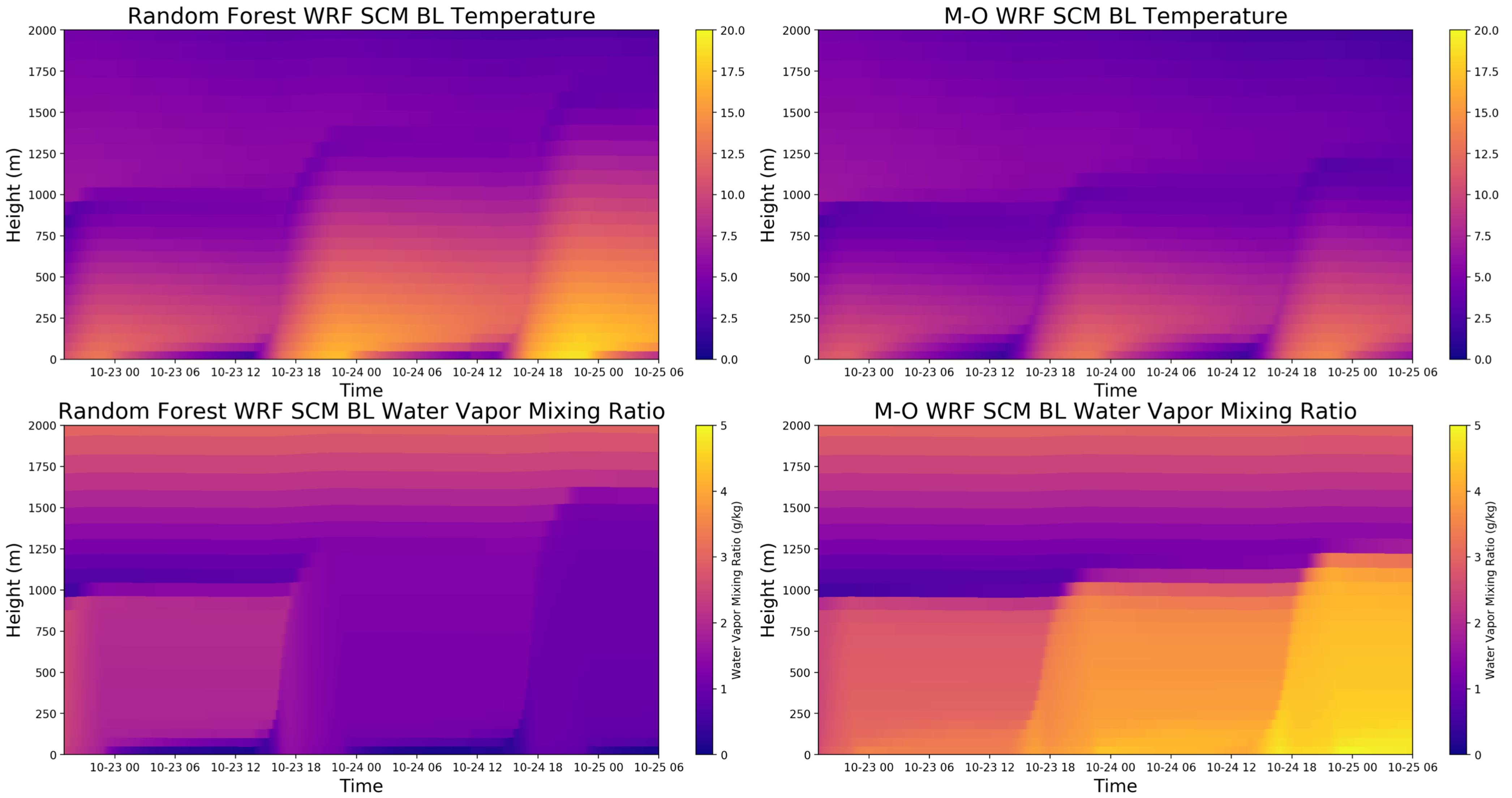
Results Courtesy Tyler McCandless

Random Forest Incorporation into WRF

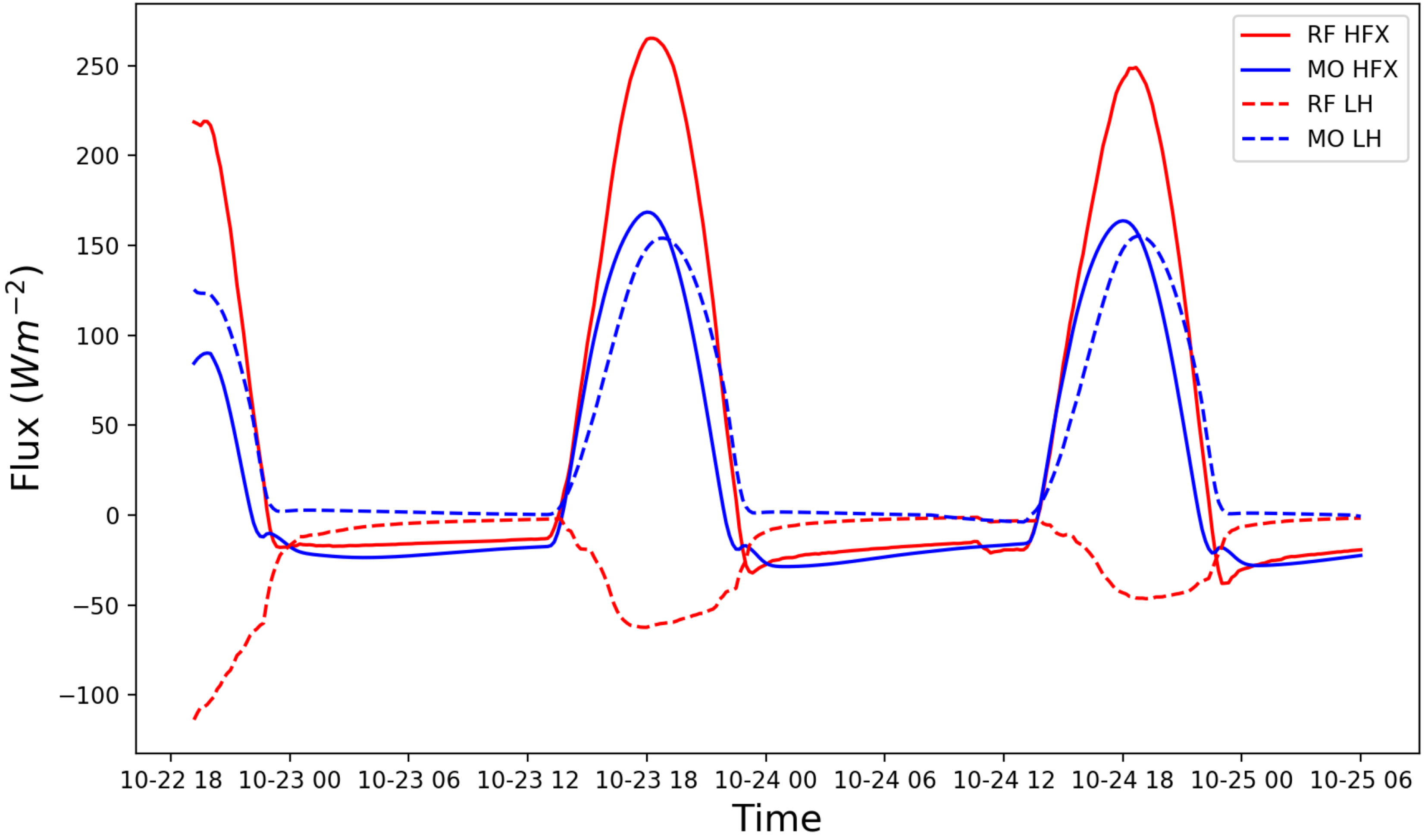
- Save scikit-learn decision trees from random forest to csv files
- Read csv files into Fortran array of decision tree derived types
- Random forest surface layer parameterization
  - Calculate derived input variables for ML models
  - Feed vectors of inputs to random forests for friction velocity, temperature scale, moisture scale
  - Calculate fluxes, exchange coefficients and surface variables
- Test with WRF SCM on idealized case study
  - Using GABLS II constant forcing
  - YSU Boundary Layer

```fortran
type decision_tree
  integer :: nodes
  integer, allocatable :: node(:)
  integer, allocatable :: feature(:)
  real(kind=8), allocatable :: threshold(:)
  real(kind=8), allocatable :: tvalue(:)
  integer, allocatable :: children_left(:)
  integer, allocatable :: children_right(:)
  real(kind=8), allocatable :: impurity(:)
end type decision_tree
```

PRELIMINARY: WRF Idealized Single Column Model Comparison



PRELIMINARY Flux Comparisons



Summary

- Random forests and artificial neural networks generally perform better than MO Similarity theory in estimating the moisture scale, temperature scale and friction velocity
- Random forests trained in one climate can be successfully used to predict in another climate
- Integrated ML into WRF and still in the initial evaluation process
  - Real-world WRF SCM runs planned for Cabauw and other sites