#21 Quantifying Wetland Carbon Emissions in Southeastern US

Viraj Acharya, Patrick Duan, Dennis Wu, Sophia Yang, Edwin Yu
Duke University

Data+

Climate+

Introduction

Wetlands are incredibly valuable ecosystems. Despite comprising a small fraction of Earth's surface, they provide ecological services such as shoreline stabilization, water filtration, flood control, and carbon storage.

Wetland soils are capable of storing more carbon per acre than many other ecosystems, and at rapid rates. However, there is increasing evidence that global warming with its increase in temperatures and sea level rise among other influences may dramatically alter the roles of wetlands in the carbon cycle.

We hope to answer the question:

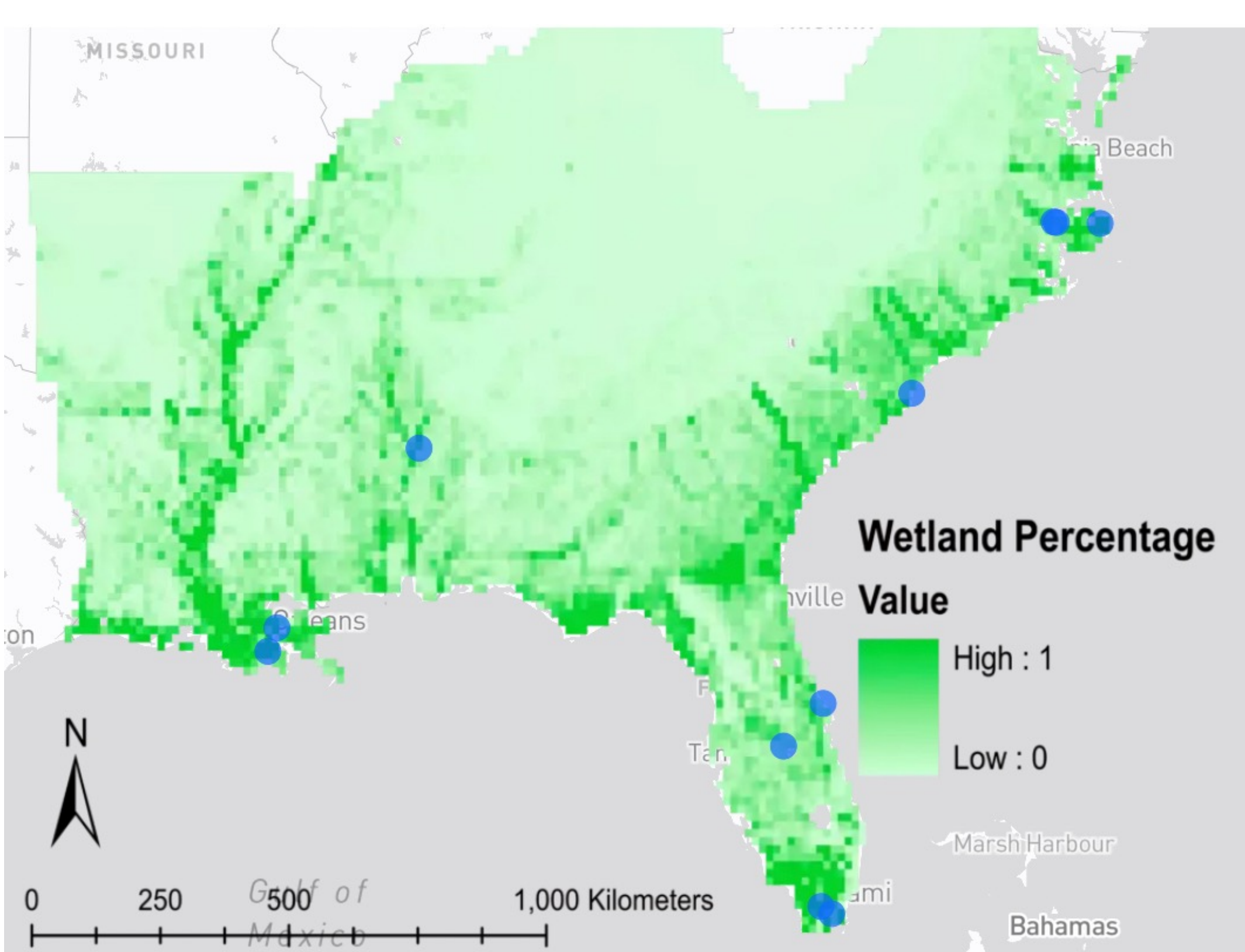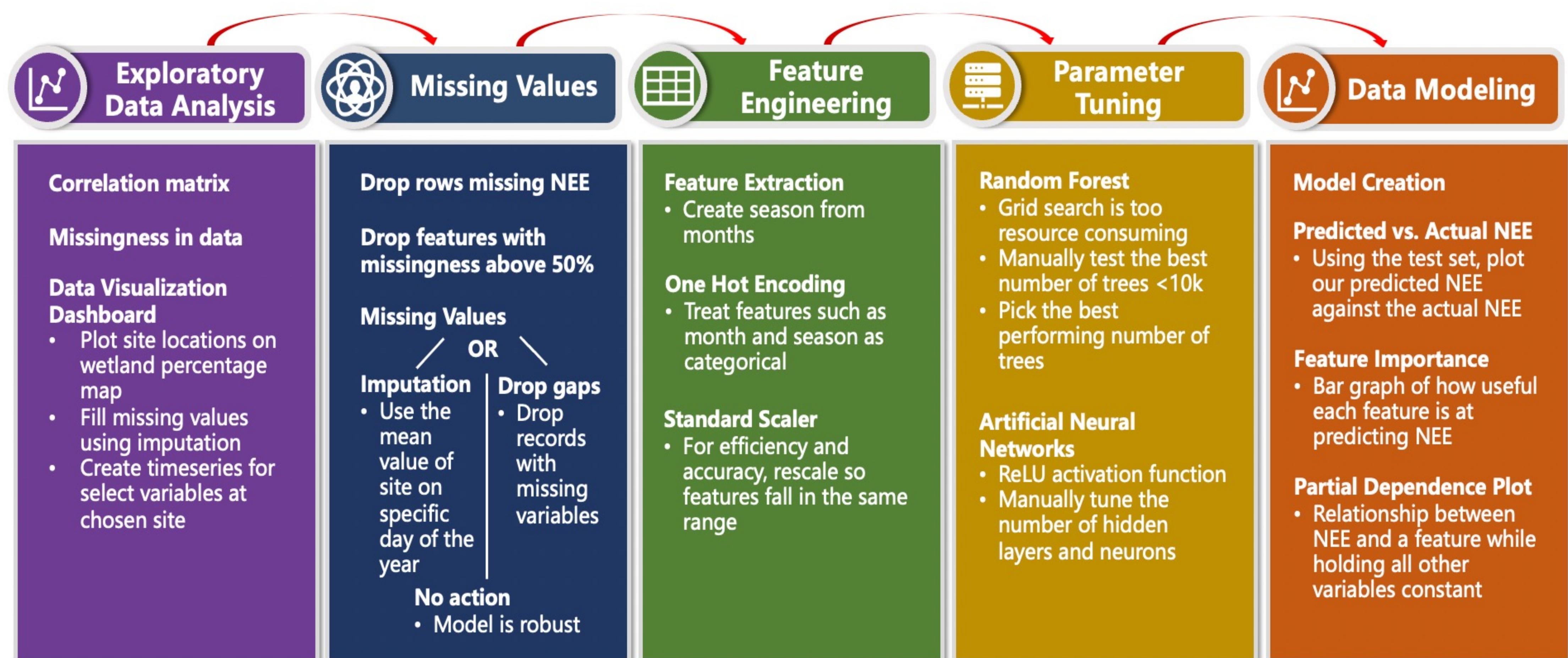- **How do different hydro-climatological variables influence wetland CO_2 emissions?**
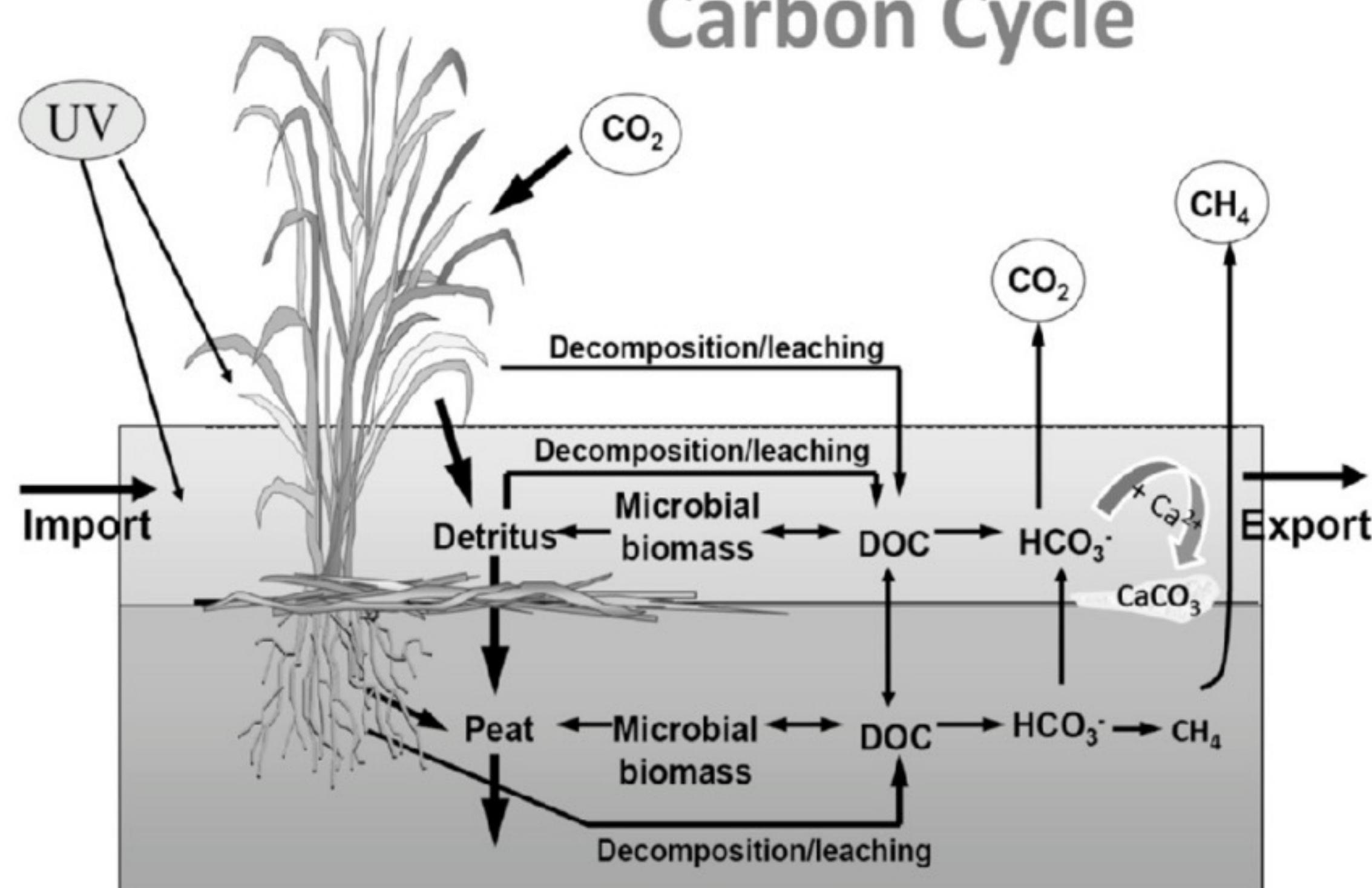
Chambers et al. (2015)

Data

- AmeriFlux network, 11 sites, 20+ variables
- **Independent variable: Net ecosystem exchange (NEE)**, + for carbon absorption from atmosphere and – for carbon release into atmosphere


Methods

Exploratory Data Analysis

Correlation matrix

Missingness in data

Data Visualization Dashboard
- Plot site locations on wetland percentage map
- Fill missing values using imputation
- Create timeseries for select variables at chosen site

Missing Values

Drop rows missing NEE

Drop features with missingness above 50%

Missing Values
OR
- **Imputation**
  - Use the mean value of site on specific day of the year
- **Drop gaps**
  - Drop records with missing variables
- **No action**
  - Model is robust

Feature Engineering

Feature Extraction
- Create season from months

One Hot Encoding
- Treat features such as month and season as categorical

Standard Scaler
- For efficiency and accuracy, rescale so features fall in the same range

Parameter Tuning

Random Forest
- Grid search is too resource consuming
- Manually test the best number of trees <10k
- Pick the best performing number of trees

Artificial Neural Networks
- ReLU activation function
- Manually tune the number of hidden layers and neurons

Data Modeling

Model Creation

Predicted vs. Actual NEE
- Using the test set, plot our predicted NEE against the actual NEE

Feature Importance
- Bar graph of how useful each feature is at predicting NEE

Partial Dependence Plot
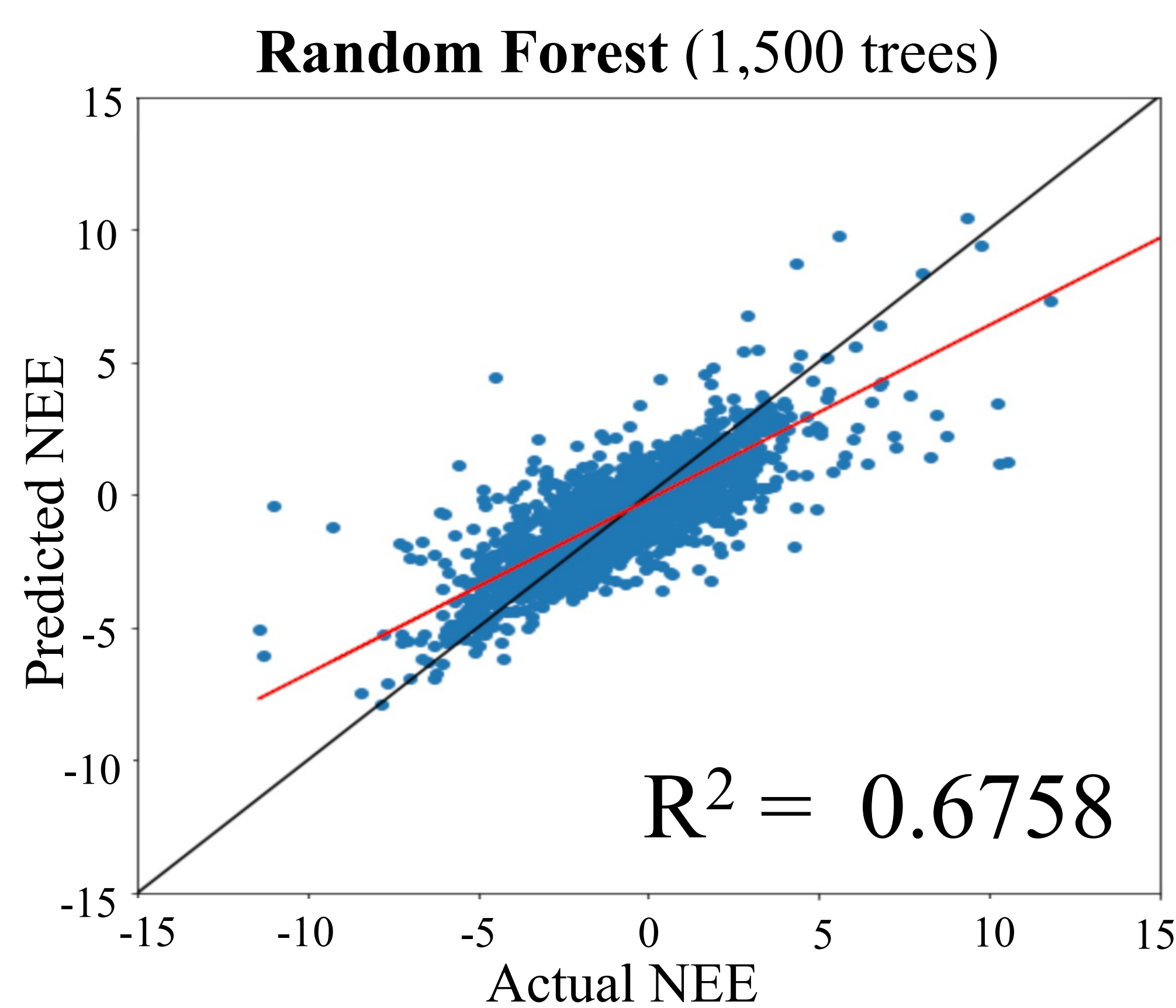- Relationship between NEE and a feature while holding all other variables constant

Results

#1 Random forest has better performance than artificial neural networks:

Predicted vs Actual NEE Plots
- Black represents ideal model
- Red represents our model

Random Forest (1,500 trees)

$R^2 = 0.6758$

Artificial Neural Network
(3 hidden layers, 50 neurons each, ReLU)

$R^2 = 0.5316$
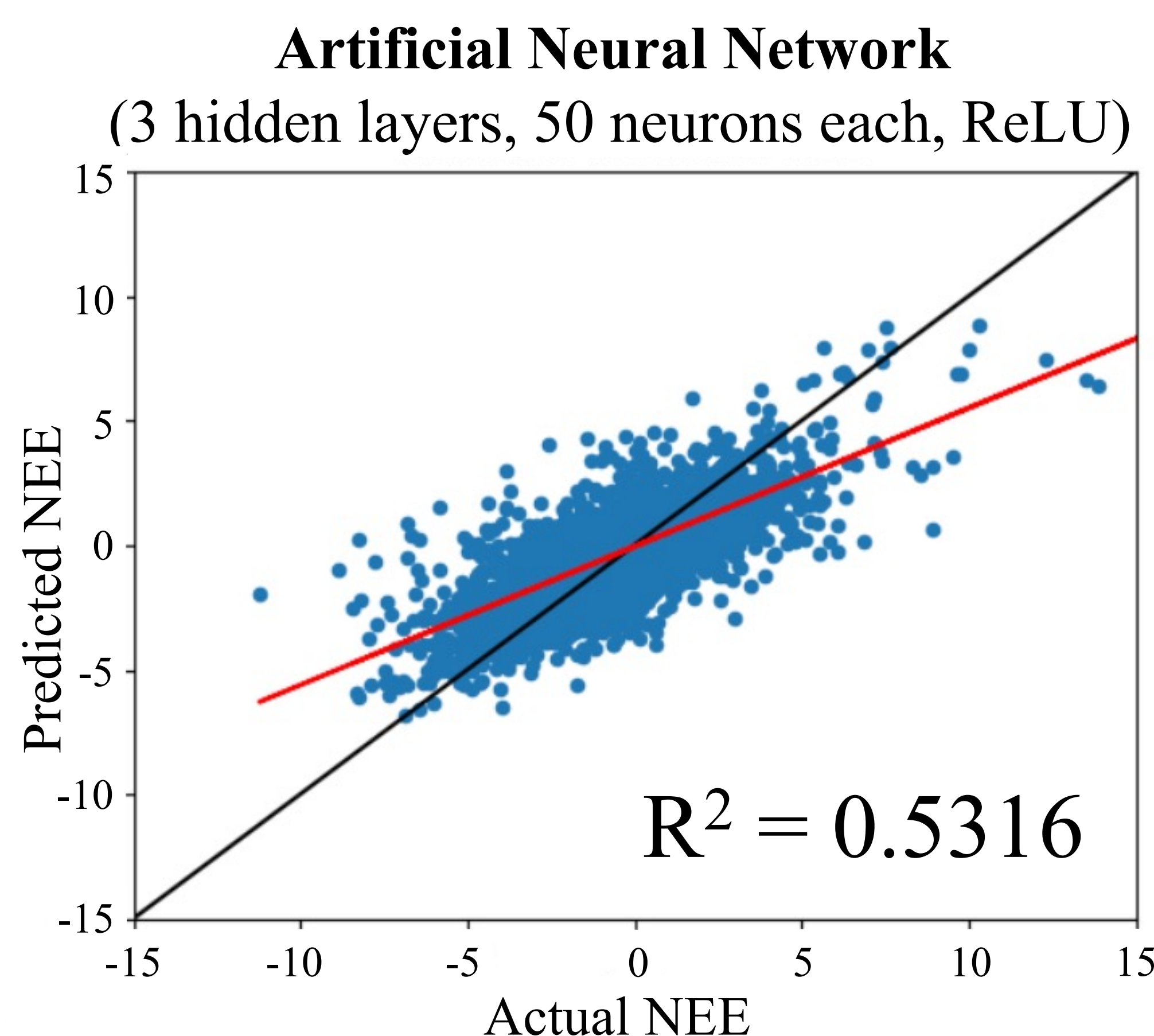
#2 Variables such as month and season seem less important than variables like water table depth (WTD)

Variable Importance
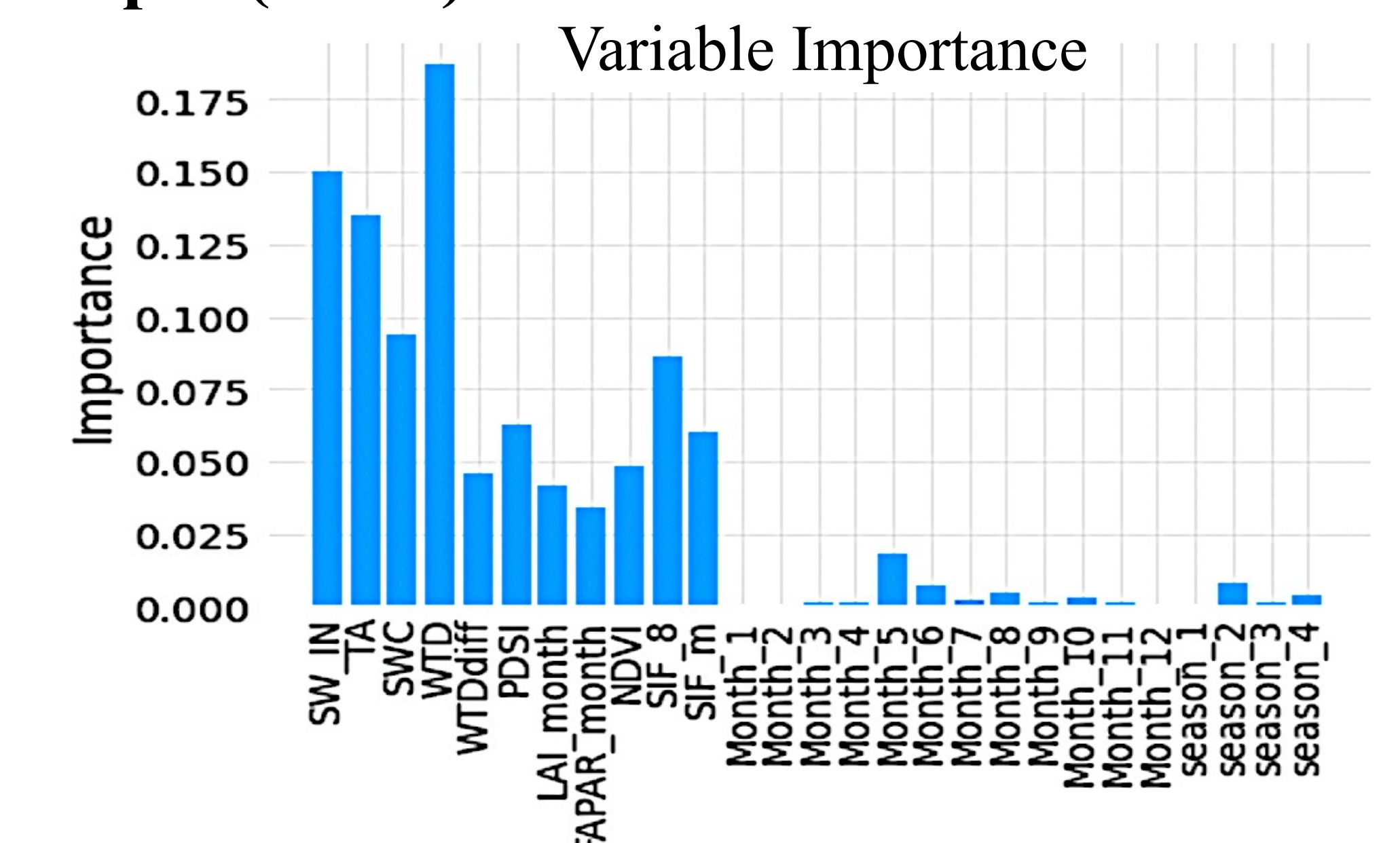
#3 There are varying relationships between hydro-climatological data and NEE

Ex:
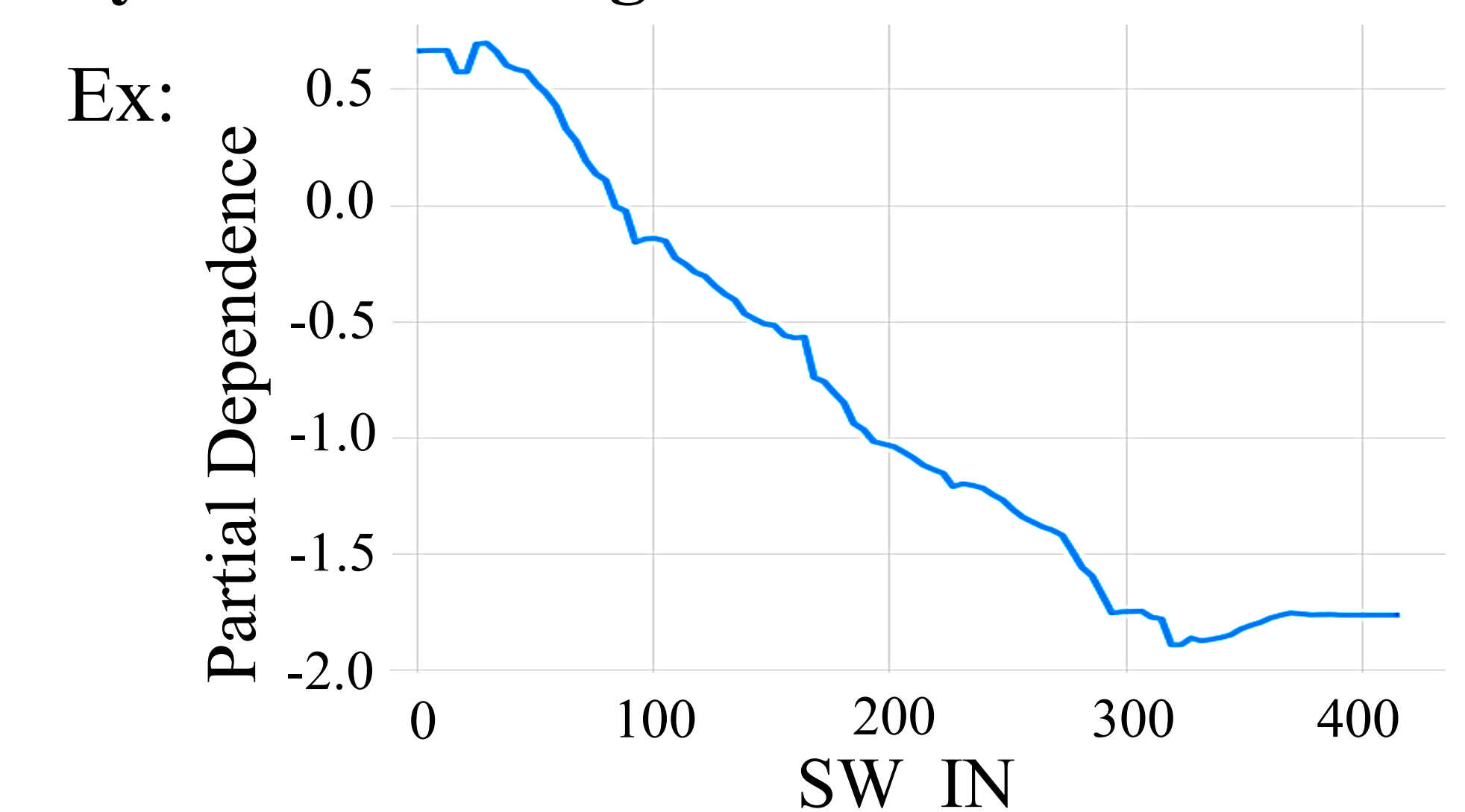
There appears to be a negative, linear correlation between incoming shortwave radiation (SW_IN) and NEE

Future Work

2023-24 Bass Connections
- Model optimization
- Use model on USGS groundwater data and analyze results