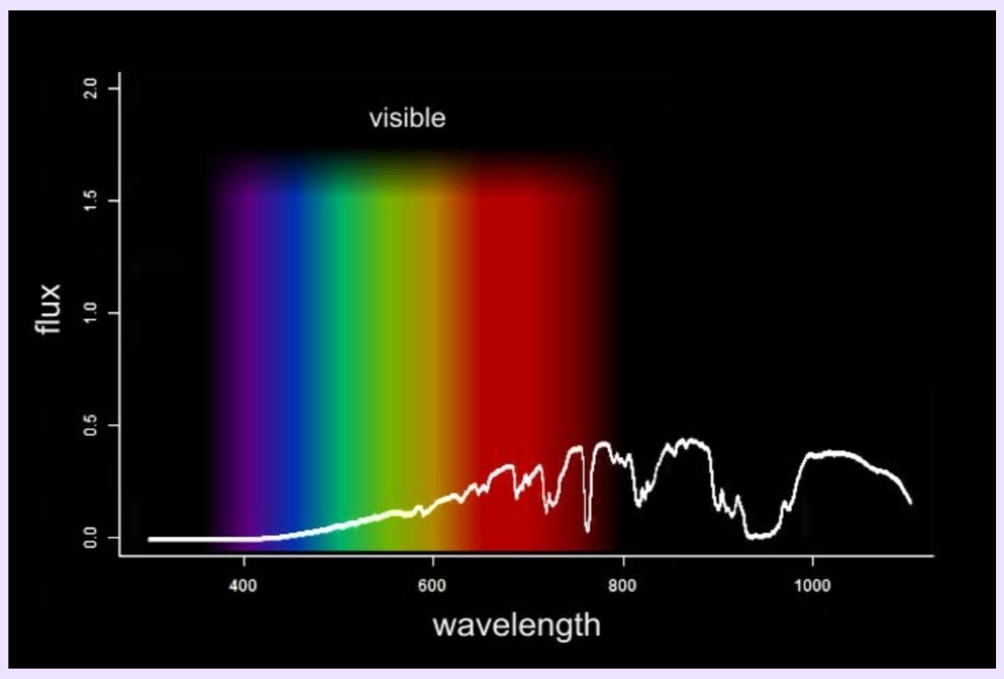
SpectraML

Estimating Solar Spectra with Machine Learning

Bo Jung, Tyler Shakibai, Brian McGreal, Rancy Zhang, Joslyn Chen

1) Accurate Prediction of Solar Spectra (given atmospheric parameters)

- Atmospheric data is recorded with **higher spatial frequency** than solar radiation
- **Physics-based models** predict solar spectra based on atmospheric observations, but are **computationally intensive**
- **Easy and accurate** spectral predictions would be applicable in many different fields, including:
  - Climate Science
  - Public Health
  - Photovoltaics
  - Built Environments
  - Ecological Modeling
  - Meteorology
  - Agronomy
  - Optics



The sun emits energy or "light" across a spectra of visible and invisible frequencies.

The spectra which reach the surface of the Earth are largely determined by atmospheric conditions.

2) Data Sources and Methodological Approaches

- Algorithmic models are trained and tested using data from the **NREL Solar Radiation Research Laboratory** in Golden, Colorado.
- Algorithmic modeling in **three stages:**
  - Ordinary least squares regression - predicts scalar
  - Neural network (multilayer perceptron) - predicts vector
  - Convolutional neural network - predicts vector

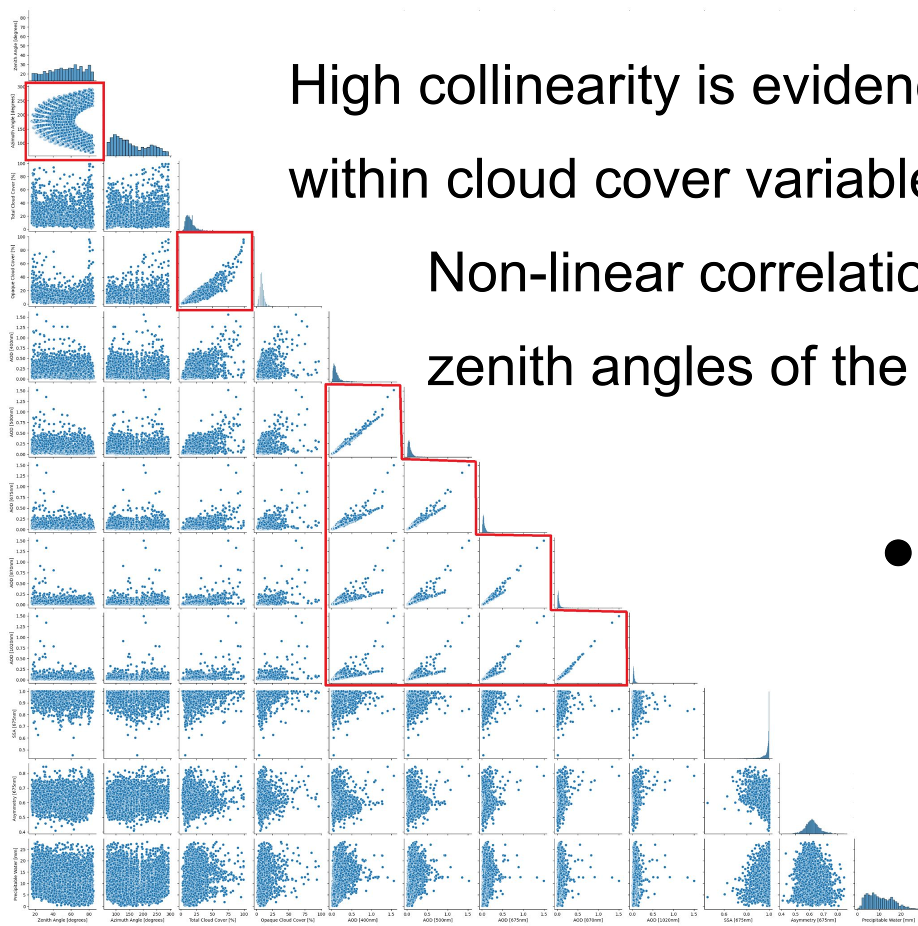
Model	Appeal	Drawbacks
Ordinary Least Squares	• Least computationally intensive	• Only able to return scalar estimates (CCT)
Neural Network (MLP)	• Returns vector (spectra) output	• Large jump in computational demand (compared to OLS)
Convolutional Neural Network (CNN)	• Generally expect better estimators • Returns vector output	• Most computationally intensive

- Choice model **estimators** are captured and used to **predict solar spectra values** from other weather data

3) Modeling Solar Spectra

Ordinary Least Squares Regression

- OLS regression returns **high measure of fit**
  - $R^2 = 0.778$
- Potential for **multicollinearity** in explanatory variables must be investigated.



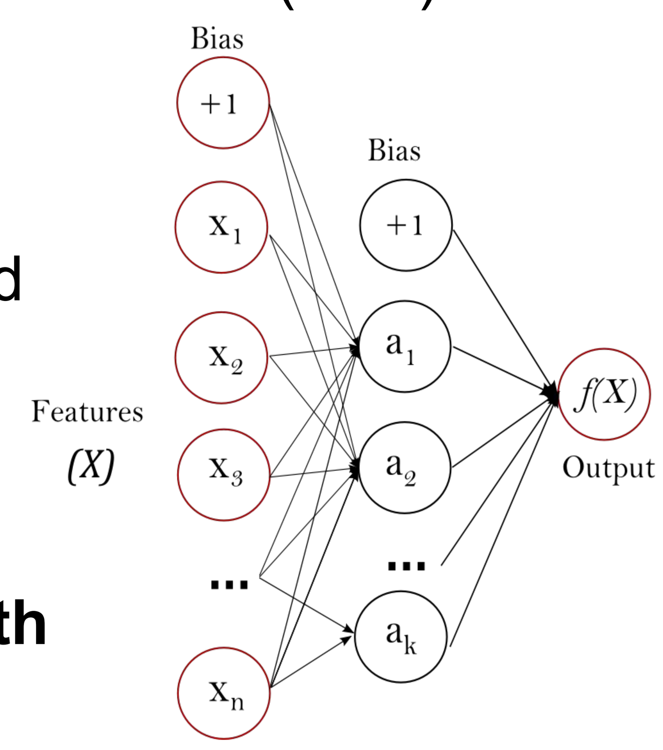
High collinearity is evidenced within aerosol measures and within cloud cover variables

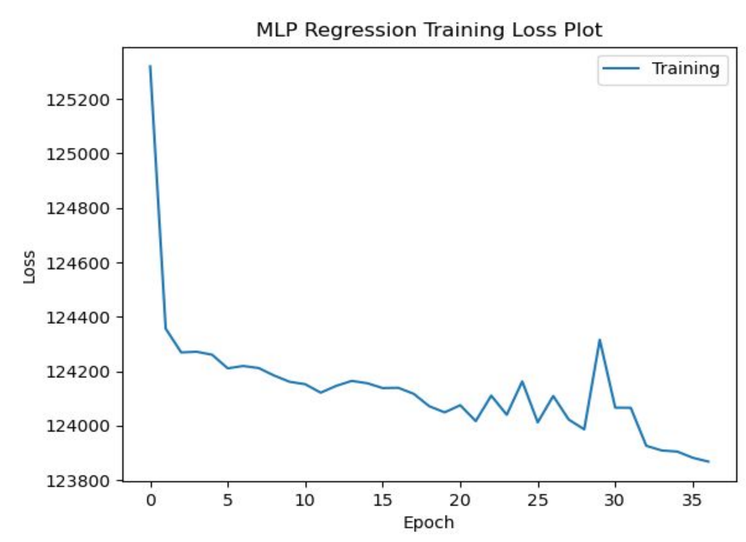
Non-linear correlation observed between azimuth and zenith angles of the sun

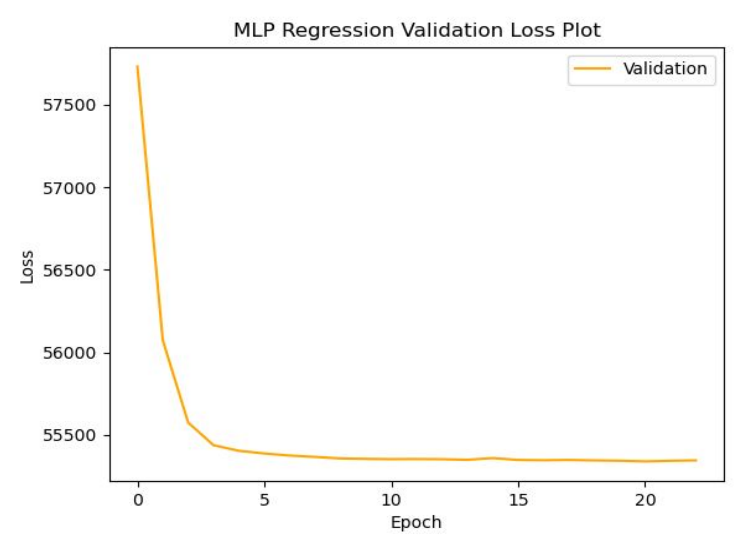
- Dropping 4 of 5 aerosol measures and choosing one cloud cover variable **reduces collinearity** with only a **small reduction in R^2** (0.72)

Neural Network

- Hyperparameters: Two 64-node hidden layers with ReLU activation function and learning rate = 0.01.
- PyTorch implementation of neural network yields an $R^2 = 0.803$.
- Considerable time to train model as **both input and output space are matrices.**







4) Limitations and Next Steps

Limitation	Next Step
Convolutional Neural Network modeling could not be completed due to time constraint.	Complete CNN analysis and incorporate parameter estimates in to implementation module.
Models trained on data collected at only one site (Golden, CO).	Future evaluation of modeling parameters using data from multiple sites to contribute to robustness.
Implementation module requires some background in python coding to run.	Development of user friendly interface for uploading input data and specifying output parameters.