Machine learning (ML) methods provide options for real-time and predictive analysis of large, complex datasets. Some well-known applications include image classification as used in image and reverse-image search, speech recognition as used in speech- to-text applications, and entertainment recommendations made through video streaming services. While all of these applications are extremely exciting both for their technological merit and either their novelty or importance to society, implementation of these methods to natural systems datasets is still in its infancy. For example, within climate datasets, ML algorithms have been used to identify the 'expert' climate model in a Multi-Model Ensemble, to analyze spatial patterns of fire danger, and identify climate extremes. The most compelling of these examples simultaneously leverages both spatial and temporal domains of datasets, enabling the use of data to its fullest extent.

Additionally, the sheer amount of available data is overwhelming. For example, NASA has planned missions that stream more than 24TB of data per day, and they project the size Climate Change data repositories alone to grow to nearly 350 Petabytes by 2030. It's this large amount of data, the need to analyze it, and the desire to use it to the maximum extent possible that drive me to bridge the fields of computer science and geological sciences by bringing ML toolsets to natural systems datasets.

Current Research

My current research applies ML methods to satellite gravity data from the Gravity Recovery and Climate Experiment (GRACE) mission to identify drought and flood in real time. Previous studies show that there are capacity limits and deficits of regional total water storage as inferred from GRACE data associated with flood and drought. However, this analysis has so far only been done *post facto*. For example, Zhao et al's 2017 work captured floods from 2002-2014, while Reager et al's 2014 work showed several months lead time in flood warnings for the Missouri River floods that occurred in 2011. While this work proves important to directly linking GRACE observations to drought and flood identification, the time lag in analysis means that these methods can not yet be used for real-time identification and warning, or in other words, the data can't be used to its fullest extent. ML methods offer the capability of real-time analysis and possible predictive capabilities by leveraging both the spatial and temporal domains of the data.

Future Research Plans

In the future I'd like to continue to apply ML methods to other natural systems, spatiotemporal datasets. Though, as I mentioned above, there are and will be numerous datasets available, I am particularly interested in those that are sparse in nature, whether that be due to limited measurements from the start of satellite missions, or due to infrequent observations. I'm interested in exploring methods for proper data augmentation with model datasets and I'm also interesting in using more complete datasets as a means to pre-train algorithms for use with infrequent observations.