Griffin Shelor

15-4-2024

GEOL 679: Python for Earth Sciences

Final Project Outline

For my final project, I will be working with the Snow Telemetry (SNOTEL) dataset. This is an important dataset to work with because it provides long-running, continuous, reliable data regarding the snowpack of the western United States and Alaska, and snowpack forms the majority of water resource availability throughout much of the western United States. Snowpack is critical to fueling agriculture, communities, ecosystems, and resource development. Many SNOTEL stations have data as far back in time as the 1980s, with some providing data even further back in time, continuing to do so to this day. As such, given that it is critical to gain as comprehensive an understanding of the snowpack as possible, the SNOTEL dataset serves as an ideal tool for doing so.

Given the importance of understanding the dynamics and fluxes of this resource, I will attempt to build a model which can be used to forecast future snow water equivalent (SWE) values at SNOTEL sites. I will evaluate this model by training it on data from Water Years 1990-2023 and testing it on available data for the current Water Year (2024). This is so that I do not have to wait until future SNOTEL measurements are collected to evaluate the predictive capacity of my model. The objective of this model will be to accurately describe and quantify the seasonal and annual dynamics of the system governing SWE, as well as quantify the uncertainty around SWE values at these SNOTEL sites. While the SNOTEL dataset comes with some potential covariates such as elevation, precipitation, temperature, and snow depth, some sites will occasionally have null values for these variables and I will either simply remove records that have null values at important covariates or I will potentially use datasets such as Gridmet and the Shuttle Radar Topography Mission elevation dataset to extract temperature, precipitation, and topographical variables such as slope and aspect at each SNOTEL site.

Once I have collected any relevant covariates, I will attempt to identify important variables by examining the correlation between those variables and SWE, as well as feeding them into a Random Forest model, which has the ability to provide numerical values regarding variable importance in regression models. Upon identifying the important variables, I will split my dataset into separate training and testing datasets in order to properly evaluate the predictive capabilities of my SWE model. My intention is to use a Random Forest model for the final predictions, though I am also interested in building a model using the PyStan module, which is an interface to Stan, a programming language and software that uses Markov Chain Monte Carlo (MCMC) to implement statistical models for Bayesian inference. I have experience using Stan in other classes and would be interested in utilizing it here.