We began with two choices for both the backcasting (target) dataset and the training dataset. For our backcasting dataset, we needed a long timeseries of rainfall data. To satisfy this, we looked at both the Sitka Airport station and the USCRN station. For our training dataset, we needed as much rainfall and pore pressure data as possible to input into the LSTM model. To this end, we looked at both the USGS Harbor Mountain and the UO-Verstovia stations.

When setting up the datasets for comparison, we noticed some apparent hurdles. The UO Verstovia station had more than a year less data than the USGS station, which was not insurmountable but undesirable. Of the two training datasets, USGS became a clear frontrunner.

When it came to the backcasting datasets, both the Sitka Airport and USCRN stations had rainfall stretching back to 2008 (in the case of USCRN) and 2002 (in the case of the Airport). This boded well, as the earliest of the 5 known landslide events happened in 2015, so we’d have the data to capture these events in our backcasting. When comparing these two backcasting datasets to the chosen USGS training dataset, it was noticed that the few USCRN dataset periods of null-values overlapped with a large portion of our training data, creating a bad match of overlapping training and target data.

Thus, it was concluded that the most plausible path forward included using the Sitka Airport station as the backcasting (target) dataset, while using the USGS Harbor Mountain station as the training dataset.

Further, we have decided to run the LSTM model using rainfall intervals of both 1- and 3-hour forecasts, with various forecast window widths from {12,24,36,48,72,96} hours. We will be using 1-hour rainfall intervals as that was how the LSTM model was originally designed to be run, and we want to compare our results between 1-hour and 3-hour forecasts to investigate if one is inherently more accurate than the other. We will be using 3-hour rainfall intervals because that is the interval given by the National Weather Service in their forecasts and will represent what we can incorporate into future applications. Lastly, the forecast window widths will be varied to study their effect on the results of the backcasting.