Task items for machine learning inflow temperature study: **Effort Estimate (40hr/w)**

1. Develop the recurrent neural network (RNN) model **1.5m**
   1. Standardize the data: make sure the right transformations are chosen. If the model did not converge or spitted out weird results this may be the culprit.
   2. Pick the correct number of hidden nodes and epochs. The upwards and downwards paths of the predictions following the observations may be easy to tweak, while hitting the peaks and valleys may be problematic.
   3. Fit the RNN.
   4. Evaluate and iterate.
   5. The Long short-term memory (LSTM) is a recurrent neural network (RNN) but better at keeping events “in memory”. Try this if needed.
   6. Report performance on different epochs (make it a gif, could be cool to see how the model’s predictions converge the longer you let it run).
2. Develop the auto-regressive integrated moving average (ARIMA) model **1.5m**
   1. Decompose the seasonality, trends and cyclical components of the data: Use decompose() or stl() to examine and possibly remove components of the time series data.
   2. Stationarity: Use adf.test(), ACF, PACF plots to determine order of differencing needed.
   3. Autocorrelations: examine plots and determine the model order.
   4. Fit ARIMA: read up on the Box-Jenkins method of fitting so you know what’s going on.
   5. Evaluate and iterate.
3. Develop the ungauged basin problem **3m**
   1. Gathering data for CDEC basins (Don may have all of this cleaned up already).
   2. Processing spatial data.
   3. Cleaning up dataframe and prep for ml.
   4. Using the LOOCV method to perform the ml for each basin.
   5. Compare the performance to a process based model (probably something that Don already has (predicted datasets)).
4. Comparison of the different machine learning methods **2w**
   1. Based on relevant measures of fit (find this in literature, Moriasi/EPA may have laid this out already).
   2. Based on # of outliers only (this will give a different flavor to the problem; here, the performance metric is just for the predictions to not be too far off, which is useful when complying with regulations is concerned). Also, can try to find a performance metric that is sensitive to outliers (consult with the statistics department at ucd).
   3. Can include the process-based model in comparisons too.