Background
WSU and Ecology developed an ozone forecast tool for Kennewick using machine learning techniques. The tool is constantly self-updating its understanding of the relationships between meteorological variables and measured ozone concentrations. These relationships are applied to meteorological forecasts obtained from several models (known as ensemble members) from University of Washington, to create multiple ozone forecasts for the next 3 days. The forecast based on the mean of all these ensemble meteorology is thought to be an accurate prediction of ozone levels which can be expected.

Initial work conducted to develop the machine learning methods is available at http://lar.wsu.edu/nw-airquest/docs/20190207_meeting/nwaq20190207_KaiFan_MachineLearning.pdf

Machine learning forecasts
The machine learning modeling framework includes two components. The random forest (RF) classifier model predicts the AQI categories (good, moderate, unhealthy for sensitive groups, unhealthy, very unhealthy and hazardous) and the multiple linear regression (MLR) model predicts the moving 8-hour averaged O_3 mixing ratios.

The objective of the modeling framework is to predict the next three days' O_3 AQI values and D8M mixing ratios. Hourly data for six meteorological variables (temperature and relative humidity at 2m, wind components u & v, planetary boundary layer height, and sea level pressure) from 4km WRF archives, time information (month, weekday and hour), and the previous day's observed moving 8-hour averaged O_3 mixing ratios are used to train the RF classifier models. The predicted AQI category information is added to the training dataset to train the MLR model. The time period of the training dataset includes 2015-2018 May-September and all available days in the recent year.

To predict the next three days O_3 mixing ratios and AQI, we acquire the ensemble UW-WRF forecast parameters for Kennewick twice a day and apply the learned relationships from the RF and MLR models. Because the previous day's observed ozone level isn't yet available at the time this model is run, the previous day's predicted moving 8-hour averaged O_3 mixing ratios are used instead. Firstly, the RF classifier model predicts the AQI categories for each hour; secondly, the predicted AQI categories and other meteorological, time information and past O_3 data are used to predict the moving 8-hour averaged O_3 mixing ratios.

Model output
When RF and MLR based forecasts are produced for each ensemble member, their daily maximums are computed and displayed in a bar chart along with the corresponding AQI category color.

The time series plot displays the progression of the measured and forecast data, along with the spread of the ensemble forecasts (grey lines). Very divergent ensemble forecasts are indicative of an uncertain forecast whereas the forecast is more certain if ensemble members are in closer agreement.