**Using random forests to determine the importance of different physicochemical soil properties on aggregate stability**

**Taylor McDowell**

## Motivation

The goal of this project is to predict aggregate stability depending on soil type and vegetation. Soil particles group together to form aggregates, which can either be stable or unstable. Stable aggregates are preferred because they are resistant to rainfall and water movement; whereas, unstable aggregates break down quickly releasing individual soil particles, which can clog pores, seal soil surfaces or result in more erosion—overall, creating a less optimal environment for plant growth. Past research has documented the short-term effects human land management and land-use changes have had on aggregation. However, not as much is known about how long term pedogenesis under natural vegetation influences aggregate stability. It is important to consider long term pedogenesis because some soils have characteristics from which aggregate stability persists for many years under changing land use; while in other soils, aggregate stability declines.

## Methods

Random Forests were used to determine the best predictor or combination of predictors to explain aggregate stability. The dataset was provided by Kasmechack and others (2016) and includes 59 soil samples with 12 possible explanatory variables: (1) vegetation, (2) soil order, (3) horizon, (4) percent clay (clay), (5) percent nitrogen (N), (6) percent organic carbon (OC), (7) pH, (8) Cation exchange capacity (CEC), (9) effective CEC (ECEC) (10) base saturation (Base), (11) effective base saturation (EBS) and (12) Calcium/Magnesium ratio (Ca).

## Results

Random forest models predicting the relative abundance of stable aggregates (FD/180< 16μm) and the rate of disintegration of aggregates ($k_1$< 16μm) both had relatively low mean sum of squares errors indicating that there was little variation between the models' predictions and the observed data. However, there were differences in most important predictors for aggregate stability and aggregate disintegration. This suggests there are complexities in the gradient between aggregate stability and instability, which are not fully captured by the models