**Statistical Modelling of Watershed Data**

This project is part of a grant from the Canadian Statistical Sciences Institute on "Statistical machine learning with functional data for assessment and prediction of landscape vulnerability to climate change and land cover development". Water flow data on over 1000 watersheds are available as time series, along with explanatory variables characterizing weather, land types, etc. The undergraduate researcher will perform exploratory data analysis, time-series analysis, and package and deploy novel R code generated by the research team for wider dissemination.

The required statistical background for the project includes regression (e.g., STAT 306), statistical/machine learning (e.g., STAT 406 or CPSC 440), and time series (e.g., STAT 443). Familiarity with R is essential, and experience creating an R package would be helpful.

Interested students please contact **Professor William Welch at will@stat.ubc.ca** and provide a cover letter and unofficial transcripts.

**Novel statistical methods for the analysis of historical forest fire data**

Along with collaborators in Forestry, we have been developing novel statistical methods for the analysis of historical forest fire data obtained from tree scars in forests throughout North America. The ultimate goal of studying the data is to develop a better understanding of how forest ecological factors, repeated burns, and climate interact over time and space (see, e.g., [1], for more background). The data present a number of challenging statistical problems tied to the "fading record" and related missing data issues inherent to historical fire data. Building on work started by a previous undergraduate researcher, the USRA will implement the methods we have developed so far as an R package, and collaborate on new models and methods for the project.

Competitive candidates will have strong coding skills, experience with R package development, and a solid background in probability and statistics (at the level of STAT 302/305). Experience with one or more of the following is an asset but not required: Bayesian modeling and inference (MCMC sampling, variational inference); latent variable models; missing data problems. Interested applicants should contact **Dr. Bloem-Reddy at ben.bloem-reddy@ubc.ca** with a cover letter indicating interest and highlighting relevant skills/experience, CV/resume, and unofficial transcripts. (All personal information sent will be treated as confidential.)

[1] https://www.landscapesinmotion.ca/updates-1/2019/4/3/a-wildfire-story-decoding-the-past-with-tree-scars

**Build an R package for comparing diagnostic machine learning models**

The main goal of this project is to turn a set of R scripts into a fully functional R package. Our existing R scripts (1) identify differences between a set of patients and healthy controls using epigenome sequencing data, and (2) build machine learning models to predict cancer risk based on these differences. Time permitting, we would also like to add functionality in the package to compare performance of several different machine learning approaches. Competitive candidates will have experience with R package development, a solid background in probability and statistics, and an interest in the analysis of high-dimensional biological data. If you'd like to learn more about how the prediction models are applied, see this paper: https://www.nature.com/articles/s41591-020-0933-1. Interested applicants should contact **Dr. Keegan Korthauer** at keegan@stat.ubc.ca with a cover letter, CV/resume, and unofficial transcripts.

**A public dashboard nowcasting COVID-19 cases in BC**

This project will apply existing R/Python code to estimate the daily number of new symptomatic COVID-19 infections in BC. This is notably different from the official "reported new cases" number appearing daily in the media. Importantly, we wish to estimate infections in real-time. The existing code works by deconvolving daily reported COVID-19 case counts using an estimated delay distribution and employs a sensor fusion layer (which fuses together predictions from models that are trained to track infections based on auxiliary surveillance streams) in order to improve accuracy and stability. We will locate appropriate auxiliary data for BC and produce a self-updating web dashboard to publish the infection estimates.

Strong coding experience in Python is essential (excellent R background will be considered), and some experience with Github is desired. Interested students please contact **Prof. McDonald** (daniel@stat.ubc.ca) and provide (1) a cover letter describing your interest in the project and background and (2) an unofficial transcript. One or more examples of (publically viewable) R/Python projects completed for course work or other purposes would be particularly helpful.