**USRA 2021 PROJECT LIST**

**1. R Package for Gaussian Processes, With Application to Computer Experiments**

The project will add further functionality to an R package for analysis of computer experiments developed as a summer project in 2020.

The current version implements maximum likelihood (ML) estimation, with prediction and visualization based on ML point estimates.  The extensions will add further Bayesian functionality.  First, one or more versions of Markov Chain Monte Carlo (MCMC) will need to replace ML in parameter estimation.  Then prediction and visualization will be adapted to take account of the uncertainty in the parameters characterized by MCMC.  Help files and documentation of test problems need to be maintained for these tasks as the project progresses.  If time permits, further functionality based on research already carried out by graduate students in my lab will be added.

Skill coding in R is essential, and experience with C/C++ would be helpful. Interested students please contact Prof. Welch: [will@stat.ubc.ca](mailto:will@stat.ubc.ca) and provide a cover letter and an unofficial transcript.  One or more examples of R projects completed for course work or other purposes would be particularly helpful.

**2. Novel statistical methods?**

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.  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. Experience with Bayesian modeling and inference (MCMC sampling, variational inference) is an asset but not required.

**Interested applicants should contact** Dr. Bloem-Reddy at [ben.bloem-reddy@ubc.ca](mailto:ben.bloem-reddy@ubc.ca) with a cover letter, 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>

**3. Systems biology to identify biomarkers of vaccine immune responses in newborns**

More than 2 million infants die every year from infections, particularly in resource-poor settings. Moreover, due to distinct immunity compared to adults, neonates (newborns) are less able to mount protective immune responses following vaccination. Improvement of neonatal immunization, thus requires a better biological understanding of vaccine-induced immune responses that correspond to protection, which is difficult due to scarcity of vaccine-related research in neonates. This project proposes an innovative systems biological investigation to better understand vaccine-induced immunity in neonates. Novel advanced statistical and computational approaches will be used to analyze very large and unbiased datasets of molecular and cellular information measured from small samples of blood obtained from 720 neonates undergoing immunization with hepatitis B vaccine (HBV), given with or without the Bacille Calmette-Guérin (BCG) vaccine. The molecular datasets consist of precise measurements of tens of thousands of gene expression read-outs (gene transcripts [RNA] and proteins), generated using state-of-the-art methods and instruments, such as next generation sequencing (RNA-Seq) and mass spectrometry. Molecular response signatures and biomarker classifiers that predict subsequent immunogenicity, especially measurable signatures of immunity against infection (correlates of protection (CoP)), will be identified. Innovative bioinformatics-based and data-driven biomarker integration approaches will reveal patterns of gene expression, bionetworks, molecular pathways and biomarker classifiers associated with successful immunization, and/or sub-optimal immunogenicity. The knowledge gained from this project will provide fresh insights to improving current immunization protocols for neonates adopted by health organizations across the world.

The main objective of this research project is to identify blood transcriptomic and plasma proteomic signatures and biomarkers in human neonates that correlate with effective immunization, using pre- and post-vaccine whole blood RNA-Seq datasets. In this first phase of the study we aim to develop efficient visualization tools to better explore the high dimensional dataset available. In addition, we aim to use statistical and computational approaches to allow biomarker integration across the transcriptomic and proteomic datasets.

**Contact:** Gabriela Cohen Freue, Associate Professor at [gcohen@stat.ubc.ca](mailto:gcohen@stat.ubc.ca)

**4. Someone to help develop the distplyr R package** (<https://distplyr.netlify.app/>), **and to help write a corresponding journal article.**

The ultimate goal is to promote the use of probability distributions in data analysis, by providing a modern interface to the R package and displaying compelling case studies in the journal article. Currently, the R package is gaining more stability, but still has some key decisions that need to be made, and I'd like your input in this process. The paper has not been written yet. At a minimum, the student should be comfortable with their R skills, should be somewhat comfortable with the tidyverse, have a basic knowledge of probability distributions, and most importantly, should want to learn more about the modern R landscape. Some knowledge of git and GitHub would be an asset. The R package makes heavy use of S3 Object Oriented programming, and likely tidy evaluation, so knowledge of these things would be an asset, but what's really important is for you to be keen to learn about these things. I'm looking forward to hearing from you!

**Contact: Vincenzo Coia,** Assistant Professor of Teaching **at** [**vincen.coia@stat.ubc.ca**](mailto:vincen.coia@stat.ubc.ca)