Change Point Detection with State Space Models for Remote Monitoring of Forest Health

This project will aim to develop tools for assessing forest health from satellite data. The goal is to apply scalable Bayesian "change point detection" algorithms to find disturbances in forest health (as measured by abrupt change in the satellite imagery at a particular location) on large scale data. The undergraduate researcher will implement a state-space model for spatiotemporal data, perform exploratory data analyses, and experiment with various mechanisms for change point detection.

At a minimum, the student will deliver a scalable Python package for spatiotemporal change point detection as well as data analyses for our collaborators in forestry. If we are particularly successful, this project should yield a novel change point detection algorithm which could lead to a co-authored publication.

This project requires proficiency in probability and statistics (e.g. STAT 302/305) as well as strong coding capabilities in Python. Competitive candidates will also have a background in Bayesian modeling and inference (graphical models, latent variable models, state-space models, Gaussian processes, and/or MCMC sampling) and/or demonstrate the aptitude to rapidly learn these concepts.

Application procedure

Please submit a cover letter, unofficial transcript, and resume/CV to Dr. Geoff Pleiss (geoff.pleiss@stat.ubc.ca). Your application materials should highlight your relevant skills and experience.

Distributed Bayesian Inference

Bayesian inference is a tool that is widely used in all fields of science and engineering to model reality while quantifying uncertainty. Modern, cutting edge Bayesian inference problems in recent years have become complex enough to require the use of massively parallel computation on cluster computers; however, existing software packages are not designed for such large-scale problems. To address this challenge, we are developing foundational methods in large-scale distributed Bayesian inference, available as an open-source package, Pigeons (pigeons.run/dev/), which enables practitioners to harness distributed computing for performing Bayesian computation at scale. Our software is based on a method called variational parallel tempering, an exciting recent development at the intersection of Monte Carlo methods from statistical computing and variational inference from machine learning.

The summer research project has two components: (1) investigating the use of variational parallel tempering in real-world scientific problems, in collaboration with domain experts in black-hole astronomy and single-cell genomics, and (2) extending the methodology from inference problems to global optimization problems.

Students will have the opportunity to gain experience in distributed computing, Monte Carlo methods, variational methods, scientific collaboration, and optimization.

Proficiency with either Python or Julia required. Strong mathematical and coding skills are preferred. Experience (from course work or otherwise) with Monte Carlo methods, variational inference, and optimization is an asset.

Application procedure

To apply, please email your resume and unofficial UBC transcript to Alexandre Bouchard bouchard@stat.ubc.ca and Trevor Campbell trevor@stat.ubc.ca

USRA Project Proposal (Supervisors: Harry Joe and Natalia Nolde)
Climate Extremes

Background

Recent years have seen many extreme weather events around the world. These include heavy rainfall events causing floods and landslides, cold spells and snow storms, heat waves and droughts having impact on human health and agriculture, leading to property damage and disruptions of day-to-day lives. It is of importance to be able to accurately quantify frequency and severity of these extreme events and their impact. Modelling of such events should account for non-stationarity of the underlying processes due to changing climate.

Project objectives

The aim of the project is to support future research on modelling of climate extremes through data collection,

visualization and exploratory data analyses as well as literature review (e.g., from journal Weather and Climate Extremes). One ultimate goal would be estimation of multivariate upper quantile curves relating weather variables to variables that measure the human effects of extreme weather.

Below are examples of several datasets that would be of interest for this project.

- Rainfall events: the extreme events are floods due to overflowing river causing property damage. Use weather and river flow and other databases (over decades) to collate/wrangle datasets that consist of storm and non-trivial rainfall events.

  An example would be to take a location where floods have occurred recently; get daily river height/flow from a nearby river monitoring station, get daily weather variables from nearby weather stations; for each "storm" or event, process the daily data to get maximum daily precipitation, cumulative precipitation, length of event, maximum river height, average temperature, etc.
- Heat events or droughts: the extreme events consist of extreme temperatures in a drought period causing deaths and hospitalizations. Use weather databases for daily temperature in order to get average and maximum daytime temperatures, average and maximum night time temperatures, average and maximum humidity, pressure etc. Maybe air quality databases can be used to get daily air quality indices as surrogates for hospitalizations, if there are nearby wildfires causing lower air quality.
- Cold spells: the extreme event consists of extreme cold temperatures causing accidents, deaths, hospitalizations and travel cancellations. Use weather databases for daily temperature in order to get average and maximum daytime temperatures, average and minimum night time temperatures, total and maximum daily snowfall amounts, etc. What are surrogate variables for insurance claims (car and property)? Maybe total days of school closures, or flight cancellations at nearby airports?

A statistical analysis will initially consist of an exploratory data analysis which will include checks for trends in median and extreme quantiles, study (visually and quantitatively) of associations between multiple variables during extreme weather events (e.g., floods, droughts).

Requirements

- An average in coursework in Math, Statistics and Data science courses
- STAT 306, STAT 443
- Comfortable programming in statistical software R.
- Excellent English communication skills (both verbal and written).
- Knowledge on data scraping from the internet would be an asset.

Application procedure

To apply, please email your resume and unofficial UBC transcript to Harry.Joe@ubc.ca and natalia@stat.ubc.ca.