Steven Stetzler

Contact Information

3910 15th Avenue NE, Room C319
Seattle, WA 98195-0002

Email: steven.stetzler@gmail.com
Web: http://stevenstetzler.com

Education

University of Washington — **Seattle, WA**
Astronomy PhD Program — 2018 - Present
DOE Computational Science Graduate Felllow — 2018 - Present

University of Virginia — **Charlottesville, VA**
B.S. Physics and B.A. Computer Science — 2014 - 2018
Honor: Highest Distinction

Research Interests

I am interested in problems and projects that are influenced by and require **large-scale computing** resources and knowledge of **advanced statistical techniques**. Within my domain of astronomy, this has taken the form of investigations into supervised classification of time-series data, evaluating approximate methods for fast regression with Gaussian Process models, and building a science platform for performing distributed computing in the cloud. I am broadly interested in new projects that will challenge my skills in **programming**, **data modelling**, **data management**, and **machine learning**.

Research Projects

Searching for and Analyzing Solar System Small Bodies in Dark Energy Camera Deep-Drilling Fields

Advisors: Prof. Mario Juric (UW), Dr. Aren Heinze (UW)

Keywords: ***Solar system small bodies***, ***Image processing***, ***Shift-and-stack***

I am currently working on finding and analyzing solar system small bodies using multi-band imaging from the Dark Energy Camera (DECam) taken using a cadence that mimics the Deep Drilling Field (DDF) cadence of the Legacy Survey of Space and Time (LSST). I am finding and characterizing two types of solar system small bodies: Main-belt asteroids (MBAs) and Trans-Neptunian objects (TNOs). The MBAs are found using a source detection algorithm on template-subtracted difference images of the DECam CCDs, and single-epoch detections are linked to find moving objects. From there, I have constructed light curves of these objects for further characterization to extract phase function parameters and color terms. I will discover faint Trans-Neptunian objects that lie below a single-epoch detection threshold by using a shift-and-stack technique accelerated with graphics processing units (GPUs) on the difference images. This project will set the path for future solar system object characterization and faint object discovery with the LSST deep drilling fields.

Emulating Simulation Output for Fast Inference

Advisor: Dr. Michael Grosskopf (LANL)

Keywords: ***Stastical modelling***, ***Emulators***, ***Gaussian processes***

I built and analyzed the performance of a set of approximate Gaussian process regression models when predicting the output of a Density Functional Theory (DFT) simulation code. This model – an "emulator" for the simulation – is cheap to evaluate relative to running the simulation, allowing for tractable Bayesian inference of simulation parameters when comparing simulation output to experimental data.

Research products produced:

1. Stetzler. S. et al. "Fast emulation of density functional theory simulations using approximate Gaussian processes" (in prep)

Data-Driven Models of RR-Lyrae Metallicity

Advisors: Prof. Andy Connolly (UW) & Dr. Kyle Boone (UW)

Keywords: ***Time series analysis, Variable stars, RR-Lyrae, Machine Learning***

I am currently investigating methods for describing the shape of RR-Lyrae, a type of variable star, light curves, a time series of stellar brightness measurements. We are building on traditional Fourier decomposition methods while introducing models based on Principal Component Analysis and non-linear decomposition models such as Autoencoders. We use these models of light curve shape to predict the metallicity, a measure of the abundance of non-Helium elements in a star, of the RR-Lyrae.

Building Cloud-based Science Platforms for Astronomy

Advisors: Prof. Mario Juric (UW) & Prof. Andy Connolly (UW)

Keywords: ***Jupyter, Apache Spark, Docker, Kubernetes***

I developed a cloud-based data analytics platform integrating the Jupyter platform (Notebook / Lab / Hub) with a distributed computing framework (Apache Spark and Dask). This platform uses containerization software such as Docker and Kubernetes deployed on Amazon Web Services. This platform allows for user friendly distributed analysis of very large (TB-scale) astronomy data sets stored in the cloud.

Research products produced:

1. Stetzler. S. et al. "The Astronomy Commons Platform: A Deployable Cloud-Based Analysis Platform for Astronomy" (2022; in review)
2. Stetzler. S. et al. 2020. "A Scalable Cloud-Based Analysis Platform for Survey Astronomy." Paper presented at Gateways 2020, Online, USA, October 12-23, 2020. `https://osf.io/e2zwf/`.

Coursework Projects

CSE 512 - Data Visualization **University of Washington**

Keywords: ***JavaScript, Python, Interactive Tools, Web***

Web: https://cse512-19s.github.io/FP-Musical-Wayfinder/

Collaboratively designed, created, and published an interactive tool for exploring Spotify music libraries using the `d3.js` JavaScript data visualization library. This web-based tool combines a `Python`-based back-end application which performs calls to Spotify's API to scrape user data and a front-end graphical interface built using JavaScript and HTML.

CSE 547 - Machine Learning for Big Data **University of Washington**

Keywords: ***Gaussian Processes, Approximate Regression Methods, Hyperparameter Optimization***

I explored methods for performing fast multi-dimensional Gaussian Process regression on large time-series from astronomical data sets. I focused on Structured Kernel Interpolation methods which speed up regression and likelihood calculations by approximating the GP kernel and using gradient-based linear system solvers. I found that these methods were unsuitable for performing precise inference of model parameters, the common use-case for astronomers; however, I verified their promised speed up for regression and likelihood calculations.

CSE 546 - Machine Learning **University of Washington**

Keywords: ***Classification, Data Processing, Time-Series***

Jointly participated in the 2018 PLAsTiCC Kaggle competition to perform machine-learned classification of time-series data from simulations of astronomical observations with the Large Synoptic Survey Telescope. Classification techniques used include logistic regression, neural networks, and gradient-boosted random forest classifiers (XGBoost). Project was instructive in data management techniques and in performing distributed processing of GB-scale data sets with millions of objects.