

# Zachary Hafen-Saavedra

PhD, Theoretical and Computational Astrophysics

z.hafen.saavedra@gmail.com || zhafen.github.io ||  

## Experience

### McCue Prize Postdoctoral Fellow

University of California, Irvine

July 2020 - Present

Irvine, CA

- Employed [natural language processing](#) to convert >200,000 scientific abstracts to quantitative data
- Discovered text similarity correlates with a  $\sim 1.5\times$  [increase in citations](#), using a custom C++ backend
- Performed complex [filtering of >2 TB of remote data](#) via the NASA astrophysics data system API
- Led an [eight-institution, international collaboration](#) to enable an analysis requiring expertise from observers, analysts, and simulators
- [Validated Bayesian statistical methods](#) against three increasingly-complex test cases
- Interpolated simulation outcomes by enhancing [sampling of N-dimensional probability distributions](#)
- Improved [mock-data fidelity by >200%](#) by interfacing with open source atomic spectra data
- Explored [6 variations of 23 datasets](#) by developing a tool for application-agnostic data-management
- Organized [a meeting of 20 leading galaxy-shape experts](#) to identify target measurements

### GK-12 Graduate Fellow

Northwestern University

June 2014 - July 2020

Evanston, IL

- Created [quantitative visualizations](#), educational [explanatory visualizations](#), and award-winning [artistic visualizations](#) to communicate core messages
- Processed [tens of TB of >20-dimensional data](#) using high-performance-computing resources, reducing to <100 GB of highly-interpretable data
- Employed [modern software development best practices](#) (unit testing, version control, etc.) to maintain a broad suite of essential software
- Utilized and modified a C simulation code to generate [>100,000-CPU-hour simulations](#) of entire galaxies
- Interpreted [complex time-series trajectories](#) by utilizing and modifying 3D interactive viz software
- Developed [software for non-relational data management](#), including analysis of contained relational data
- Brought opportunities to [>100 students from underrepresented backgrounds](#) by leading one of Chicago's first data-science education initiatives
- Collaborated with a multidisciplinary range of scientists to publish (to date) [35 papers, 6 as a lead author](#)

## Skills

**Techniques:** data analysis (time-series, sparse, big data, data cleaning), natural language processing, code testing (unit, integration, functional, CI), simulations, frequentist/Bayesian statistics, multiscale/hierarchical modeling, analysis pipeline development, forward modeling/mock data, technical writing

**Soft skills:** technical leadership and management, public speaking, mentoring (technical, professional)

**Tools:** Python (numpy, pandas, matplotlib), SQL, NoSQL, Unix/Bash, C/C++, parallel computing, git, nltk

## Education

### Northwestern University

PhD, MS, Physics and Astronomy

Specialization: Theoretical and Computational Astrophysics

May 2020

Evanston, IL

### University of Northern Colorado

BS, Physics, Math emphasis

May 2014

Greeley, CO