

# MATTHEW WILDE, PHD

## Data Scientist

@ mattcwilde@gmail.com

🔗 <https://mattcwilde.github.io>

in [www.linkedin.com/in/mattcwilde](http://www.linkedin.com/in/mattcwilde)

🐙 [github.com/mattcwilde](https://github.com/mattcwilde)

## RELEVANT SKILLS

**Python** (11 years, Pandas, Numpy, Matplotlib, Seaborn, SciPy, Scikit-Learn, Pytorch, JAX, Multiprocessing, Xarray)

**Bayesian Statistics** (5 years) **Machine Learning** (3 years)

**Data Visualization** (10 years, Plotly, D3.js, Tableau) **Cloud Computing** (4 years, kubernetes, AWS, Azure)

**SQL** (1 years) **Git** (8 years) **Big Data** (100 GBs simulations, TBs tabulated data, 100 GBs raw image data)

**Remote Sensing** (UV/visible/IR spectral, image processing)

**Leadership** (group meetings, mentoring, collaborations)

**Scientific Communication** (seminars, group meetings, public lectures)

**Interdisciplinary Collaborations** (applied statisticians, computer scientists, 1 co-authored paper)

**Scientific Writing** (3 first authored papers, 7 coauthored) **Mentoring** (11 years)

## WORK EXPERIENCE

### Modelling the Universe with Slime Mold

#### SDSS IV collaboration

📅 March 2021 - Present

📍 Remote

- Adapted and applied novel algorithm based on *Physarum polycephalum* to a 3D map of  $>10^9$  galaxies, creating a revolutionary new public data product of the dark matter in the universe (Python, SciPy, Git, SQL, Xarray)
- Lead an interdisciplinary collaboration with a computer scientist and astronomers from UC Santa Cruz, New Mexico State, and UW
- Identified and corrected major computational errors via exploratory data analysis and optimized the statistical interpretation using advanced image processing algorithms (Python, Bayesian statistics, Git, Dask)
- Created and continue to develop custom analysis software, `pysl1me` (Python, Bayesian statistics, Git, Dask, Xarray)
- Lead a weekly collaboration meeting to apply analysis to next generation datasets to increase the impact and improve downstream users experience (Python, Bayesian statistics, Machine Learning)

### Mapping Galactic Atmospheres via Image Processing and Bayesian Analysis

#### University of Washington

📅 Jan 2017 - Present

📍 Seattle, WA

- Leader in large collaboration, wrangling raw and unstructured imaging data from the worlds premier ground and space telescopes into a public relational database for the scientific community that is  $10\times$  larger than previous datasets (Python, Pandas, Git, Scikit-learn, Machine Learning)
- Increased the detections of low signal-to-noise objects by factor of  $\sim 10$ , enabling more robust model predictions for downstream users
- Developed Bayesian models that estimate the extent of galactic atmospheres as a function of galactic properties using a data-driven priors (Python, Bayesian statistics, Pandas, Git)
- Lead weekly meetings to train undergraduate researchers in python analysis tools and inspect unstructured data with custom GUIs (Python, PyQt, git, Jupyter, Azure)

### NSF Integrative Graduate Education and Research Traineeship (IGERT) Fellow

#### eScience, University of Washington

📅 Sep 2015 - May 2018

📍 Seattle, WA

- Internship in Data Science which included courses in PhD level Machine Learning, Data Visualization and Advanced Statistics
- Created a tool `astrokriging` to approximate a computationally expensive generative model to explore parameter space using ML, finding that a boosted random forest performed as well as a more expensive Gaussian Process (Python, Scikit-learn)
- Created a web-based, interactive, multi-panel dashboard to visualize and analyze high-dimensional data of planet-hosting stars (Python, D3.js, SQL)

## EDUCATION

### PhD Astronomy

University of Washington  
2016 - 2022  
Seattle, WA

### MS Astronomy

University of Washington  
2015-2016  
Seattle, WA

### BS Physics

UT Austin  
2005-2010  
Austin, TX