Zhuchang Zhan, Ph.D.

Research Scientist, MIT

• Cambridge, MA • (217) 979 3735 • zzhan@mit.edu
• LinkedIn • Github • Google Scholar

- **Languages & Aptitudes:** Python, SQL, C++, Linux Shell, Latex, Tensorflow, Keras, Pytorch, Scikit-learn, Pymc3, nltk, pandas, h5py (HDF5), numpy, scipy, mongoDB, Git, Jupyter, Sublime, Eclipse, VScode, matplotlib, seaborn, Tableau
- **Machine Learning: Regression (**Linear, Logistic, Elastic-net, Kernel, SVR, Random Forest), **Classification** (KNN, SVM, Naive Bayes), **Clustering** (K-means, Gaussian mixture, DBSCAN, Hierarchical), **Reduction** (PCA, t-SNE), **Deep Learning** (CNN, GAN, LSTM, NLP)
- **Technical Skills:** Bioinformatics, GPU-Computing, Data Interpretation, Signal Processing (Fourier Transform), Information Retrieval, Bayesian Inference, Linear and Stochastic Modelling, Monte-Carlo (Markov Chain, Hamiltonian)

EDUCATION

Massachusetts Institute of Technology Jun. 2015 - Feb. 2021
Ph.D., Physics (Subdiscipline: Planetary Sciences)

University of Illinois at Urbana-Champaign Sep. 2011 - May 2015
Bachelor of Science., Physics and Astronomy

DOCTORAL THESIS

Expanding Plausible Biosignature Gas Candidates

Advisor: Prof. Sara Seager

*Searching for gases produced by alien life (biosignature gases) is a vital challenge of the century but only ~20 biosignature gases are identified in the last two decades. My thesis work revolutionized the industry by enabling the assessment of millions of gases as potential biosignature gas via a novel **machine learning** framework.*

- Systematically filtered through 16,000+ gases, each with 10+ categories of data, to identify the most plausible biosignature gas candidates and discovered 20+ new biosignature gases not previously studied, 2x the total.
- Enabled assessment of millions of gases with no spectral data by **predicting** spectra through a combination of maximum common substructure (**MCS**) and mapping unknown molecules to **hierarchical-clusters** of molecules with spectral data.

RESEARCH PROJECTS

NASA TESS Variable Star Detection Pipeline

Jun. 2017 - Present

The Transiting Exoplanet Survey Satellite (TESS) is a space telescope dedicated to finding planets beyond our solar system.

- Constructed an **ETL** pipeline to reduce **terabytes** of **time series data** and discover new categories of variable stars (**periodic signal**) using **Fourier transform** and stellar flares (**anomalies detection**) using **PCA.**
- Led the discovery of 40+ “Complex Rapidly Rotating Mdwarf Stars" and applied **convolutional-neural network (CNN)** to discover 10x more targets in low temporal resolution data.
- Constructed a comprehensive dust particle extinction database by merging multiple online sources in a **pandas** data frame. The Database enables accurate modeling of dusty-tail disintegrating exoplanets.
- Developed an autovetter to aggregate key parameters, which automated **80%** of the final human vetting process. The auto vetter is regularly used by the TESS research team (over 100+ members and used in 10+ journal articles).

Biochemical Informatics Research

Jun. 2017 - Present

I led the building and maintenance of the Seager Group Natural Product Database (1M+ molecules) and led the big-data and machine-learning driven analyses to study complex biological reaction networks.

- Built multiple custom web scrapers using **selenium** and **rdkit** to extract biochemical information from 10+ online biochemistry resources and merged the data in our online **SQL** databases. Used **Tableau** for database visualization.
- Identified molecules produced by life (natural products) by testing an assortment of 20+ **machine learning** methods to predict molecular properties. New results provide valuable insight for alerting side-effects in drug discovery.

Discovery of Phosphine, A Possible Sign of Alien Life, on Venus

Jun. 2018 - Jun. 2020

Working with renowned scientists in a US-UK joint-collaboration to discover phosphine (a pungent gas found in swamps) in the upper atmosphere of planet Venus and rigorously ruled out abiotic production of phosphine.

- Developed analysis & validation tools for the discovery of phosphine on Venus and as a potential biosignature gas.
- Constructed a photochemical reaction network **SQL** database to include over 100,000 biological pathways not previously studied, which form the basis for studying phosphine’s biological production in the Venusian atmosphere.

Habitable Exoplanet Atmosphere Spectra Model
Jun. 2017 - Feb. 2020

Existing exoplanet atmosphere spectra models simulate only a few dozen gases. I developed the first model that facilitates studying the 10,000+ gases life can produce (~1000x more molecules than modern methods) and is used in multiple recent peer-reviewed journal publications.

- Implemented atmosphere information retrieval with **Hamiltonian Monte Carlo**, which enables the retrieval with ~10x speed and > 10x free parameters compared to traditional **MCMC** methods.
- Constructed the most comprehensive gas spectra database (100,000+ gases) by scraping and merging multiple online resources using **beautifulsoup** and **selenium** and stored in **SQL** databases**.**
- Converted legacy **C++** based code to python by swapping grid calculation with optimized matrix operations (**scipy**) and implemented custom rapid data-block loader (**h5py)** and improved code efficiency by 26x**.**

NASA Martian Greenhouse Sustainability Research
Jun. 2018 - Sept. 2020

*The 2019 "NASA Big Idea Challenge" competition aims to establish a self-sustainable greenhouse on Mars. The MIT team (BEAVER), which I lead the biology simulations, won **2nd** place (out of 100+ universities teams nationwide)*

- Designed the spiral cone architectural structure for our greenhouse and constructed a data-driven **simulation model** to study the crop harvest cycle and crop rotation. Display simulation results using **matplotlib** and **seaborn** plots.
- Pioneered the study of how the accumulation of 100+ trace gases produced by eight varieties of crops grown in an enclosed environment may have a long-term effect on plants and inhabitants.

ACADEMIC HONORS AND AWARDS

- Ph.D. Fellowship awarded by NASA TESS Research Grants 2018-20
- Ph.D. Fellowship awarded by Amar G. Bose Research Grants 2015-18
- NASA Big Idea Challenge: Martian Greenhouse Design: Second Place Apr. 2019

SELECTED PUBLICATIONS

- **TESS Discovers a Remnant Planetary Core in the Hot Neptunian Desert** (2020) D Armstrong, *et al*., Didier Queloz (*2019 Nobel Physics Prize*), *et al*., ***Z. Zhan***. ***Nature*** (583 (7814), 39-42)
- **Complex Rotational Modulation of Rapidly Rotating M Stars Observed with TESS** (2019) ***Z. Zhan***, MN Günther, S Rappaport, K Olah, A Mann, AM Levine, et al., ***ApJ***. (876 (2), 127)
- **Stellar Flares from the First TESS Data Release: Exploring a New Sample of M-Dwarfs** (2020) MN Günther, ***Z. Zhan***, S Seager, PB Rimmer, S Ranjan, et al., ***ApJ.*** (159 (2), 60)
- **Assessment of Isoprene as a Possible Biosignature Gas in Exoplanets with Anoxic Atmospheres** (2021) ***Z. Zhan***, S Seager, J Pektowski, et al., ***Astrobiology*** (Accepted for Publication)
- **Detection of Phosphine Gas in the Cloud Decks of Venus** (2020) J Greaves, et. al., ***Z. Zhan***. ***Nature Astronomy*** 1-10
- **Persistence of Flare-Driven Atmospheric Chemistry on Rocky Habitable Zone Worlds (2021)** H Chen, **Z. Zhan**, A Youngblood, ET Wolf, AD Feinstein, DE Horton. ***Nature Astronomy***, 1-13