

# Zachary K. Smith

10391 Old Dayton Rd. New Lebanon OH 45345 | 740-405-4219 | ezke.smith@gmail.com | smthzch.github.io

## Summary

I am a data scientist/machine learning engineer with experience in end to end model development and deployment. I primarily work in Python but have released and maintained projects developed in C#, C++, and Java. Problems solved include processing large amounts of data on cloud infrastructure, deploying models on orchestration services, and optimizing data collection. I have experience working on a diverse set of data modeling problems including deep learning on remote sensing data, Bayesian models on air quality sensor data, and optimal sequential decision making.

## Experience

### Senior Machine Learning Engineer

*Aclima* | Remote | January 2023 - Present

- Develop and deploy gas leak localization models to aid utilities in region prioritization (numpyro)
- Develop spatio-temporal models for air quality prediction using GPyTorch
- Refactor vehicle route planning code into modular python package for ease of expansion/testing
- Define and implement proper cross-validation scheme for flagship spatio-temporal model to quantify model improvements

### Data Scientist

*NCX* | Remote | April 2021 - November 2022

- Build deep learning models to predict forest carbon using remote sensing data (pytorch)
- Process large scale geospatial vector and raster datasets (python, stac, xarray, rasterio)
- Create Bayesian Optimal Experimental Design package to target field data collection used to direct and optimize \$2M in data collection: ABODER
- Write custom algorithms in C++ to speed up data processing 100x

### Data Scientist / Research Engineer

*Aptima Inc.* | Fairborn, OH | November 2018 - April 2021

- Develop particle filter based interior localization algorithm for iOS app
- Develop POMDP algorithms for adaptive training systems and agent based simulations (C#)
- Create predictive and anomaly detection algorithms for automated alert systems
- Develop Bluetooth/MQTT/Arduino based cognitive load measurement device (Arduino)

### Associate Computational Biologist

*UES Inc.* | WPAFB, OH | June 2017 - November 2018

- Analyze high dimensional datasets for biomarker discovery of human physiological state
- High throughput sequencing analysis
- Provide statistical consultation and data analysis throughout the course of studies

## Hydrology Technician

*Miami Conservancy District* | Dayton, Ohio | October 2012 - June 2017

- Create desktop / web / mobile applications for the collection, formatting, and uploading of field data including the server and database processes. (C#, Android/Java)
- Lead project development of web/map based data access applications for the public. (C#, Javascript)
- <http://apps.mcdwater.org/waterapps/wellnetwork>

## Education

### M.S. Applied Statistics

*Wright State University* | Dayton, Ohio | December 2018

### B.S. Earth Science, Minor Soil Science

*California Polytechnic State University* | San Luis Obispo, California | June 2012

## Skills

- **Programming:** Python, R, C/C++, C#, SQL, bash
- **ML Frameworks:** Scikit-Learn, PyTorch, GPyTorch, Stan, NumPyro
- **Infrastructure:** Docker, GCP, AWS, Databricks