

# MATTHEW G. HOGAN

<https://www.linkedin.com/in/matthewghogan> || <https://www.github.com/hoganman>

<matthewghogan@proton.me>

Boulder, CO, USA

1-831-402-1086

---

## SUMMARY AND QUALIFICATIONS

- Detail-driven, analytical, problem-solver, and initiative-taker through grit and creative thinking.
- Effective and creative communicator by delivering conceptual insights.
- Fluent in C++ and Python for machine vision, mathematical analysis, and parameter estimation tasks.
- Over 6 years of leadership experience in feature engineering and data collection directly applicable in fundamental science and data science applications.
- Subject matter expert in physics, image processing, and data visualization.
- Experienced in machine learning (ML), exploratory data analysis, advanced statistics, and predictive analytics in Python and R.

---

## SKILLS

### Data Science

- Performing data analysis using R and Python tools like NumPy, SciPy, Pandas, and Biopython.
- Preparing presentations with non-technical stakeholders utilizing industry standards including PowerPoint, Jupyter, and Matplotlib.
- Exploiting machine learning (ML) tools with Scikit-Learn, OpenCV, PyTorch, and Tensorflow.
- Interfacing with SQL and noSQL databases with object relational models.
- Research and development using artificial intelligence (AI), regression, and predictive analytics.
- Calibrating digital sensors used in production environments.
- Utilizing high performance computation (HPC) for numerical simulations and mathematical calculations.
- Building tools interfacing with Azure and AWS.
- Building UI/UX dashboards in Python with Dash.

### Software Development

- Templating and object-oriented programming (OOP) in Python, C/C++, and Julia.
- Scripting in Linux shells, Python, and MATLAB.
- Deploying software in Docker containers.
- Building unit tests and writing thorough documentation.
- Collaborating with version control software: Git and SVN.
- Debugging in Python and C++, including GDB and Valgrind.
- Implementing multiprocessing algorithms.

---

## PROFESSIONAL EXPERIENCE

### Jan 2023-Apr 2023 – Senior Data Scientist; Contract at Apex Systems (Lightning eMotors)

- Collaborated with electric vehicle (EV) engineers to identify key component failures to devise analytics to create new predictive health metrics.

- Maintained and took ownership of time series neural network Tensorflow models, which utilize LSTM and dropout layers, designed to predict critical vehicle metrics like driving range.
- Leader of CI/CD and documentation efforts by building integration tests, adding missing documentation with examples, and inserting typing stubs to existing Python source code.
- Removed security risks like hard-coded passwords and accomplished migration tasks in the transition from on-premises databases to AWS cloud-hosted databases. by defining incomplete Python software requirements and designing unit tests using Pytest for the new database configuration.
- Utilized numerical and scientific Python libraries to load data (Pandas DataFrames), filter data by conditions (NumPy arrays), and fit data to models (SciPy). This leveraged clever algorithms to elucidate trends and correlations on multiple time scales as visualized in Matplotlib and Seaborn.
- Discovered analytics trends that would significantly reduce the likelihood of a very common electric vehicle (EV) failure on historic telemetric time series data extracted from an on-premises data lake. This required cross-referencing EV telemetric data with customer service incident case data from a warehouse management software (WMS) called NetSuite. The methods and results were carefully documented in a report prior to departure.
- Exploited best Python practices including object-oriented programming (OOP), using abstraction layers to define behaviors and simplify patterns, test-driven development to catch bugs and errors, and building object relational models with Python SQLAlchemy to have safe and robust interfaces with MySQL relational databases.

#### **Jan 2022-Oct 2022 - Data Scientist; Contract at Brilliant InfoTech, Inc. (KPMG)**

- Integrated multiple, complex data sources, including JSON, NumPy arrays, and Pandas DataFrames, to provide live, actionable information to perform business logic and analytics.
- Built a modern, cloud-based video analysis data pipeline using Microsoft Azure services, like Blob Storage and EventHub, where human activity was observed. Videos were processed using GPUs with the YOLO object recognition model in the OpenCV library.
- Collaborated in an Agile development environment and presented work at end of sprint meetings to non-technical stakeholders.
- Implemented simple, interactable frontend/UI dashboard using the Python Dash library for a video analysis pipeline that summarized object recognition and provided playback of processed videos.
- Prevented regression by implementing integration and unit tests in the Python Pytest and unittest libraries which were automated upon merge and pull requests on the cloud with GitHub Actions.
- Maintained meticulous documentation for developers including docstrings with examples and guides using the Sphinx documentation generator.
- Designed end-to-end testing and deployed releases as Docker containers through GitHub.
- Protected the pipeline backend from malicious SQL code injection with an object relational model.
- Exploited multiprocessing algorithms to parallelize complex tasks across multiple CPUs.

#### **Aug 2021-Dec 2021 - Algorithms Engineer; Contract at Insight Global, Inc. (Northrup Grumman)**

- Provided support for remote-sensing data for a defense contractor including intensive use of the C++ debuggers GDB and Valgrind in multiprocessor programs.
- Leveraged experience in C++ and Python as an Algorithms Engineer for space-based monitoring processing software.

#### **Mar 2020-Aug 2021 - Research Scientist at NorthWest Research Associates**

- Demonstrated that disadvantaged radio frequency (RF) receivers can achieve high accuracy geolocation (error less than 10 km) by utilizing skywave and over-the-horizon propagation from known transmitters.

- Converted scientific software, initially developed in MATLAB, into Python using SciPy and NumPy.
- Expanded and innovated ridge detection techniques to search for image contours in Python. This utilized convolutional filters to post-process ionogram images using SciPy and NumPy.
- Used SVN and TortoiseSVN for version control and Anaconda Python conda package manager.
- Implemented custom digital low-pass filters to upsample and smooth numerical data.
- Designed diagnostic tools and visualizations using the Python Matplotlib library to provide feedback on how input parameters affected intermediate output with scientific software.
- Implemented mathematical models to simulate RF skywave propagation in the ionosphere. The software was optimized for Docker containers by exporting Python NumPy code into compliant C source code using Cython.
- Wrote extensive software documentation with the Sphinx documentation generator, which compiles source code documentation and other files into accessible HTML pages.
- Built and maintained Docker and Docker Compose configuration files for cross-platform software.

### **May 2013-Feb 2020 - Research Assistant at Colorado State University**

- Frequented Japan over ten times to calibrate sensors and prepare presentations to an international community of stakeholders, scientists, and engineers called the T2K experiment.
- Measured a 1%/year detector calibration degradation in a critical component of the T2K experiment which required enormous teamwork among international collaborators.
- Deployed random forest classifier to improve the purity (efficiency) of innovative analysis from 50% (6%) to 61% (12%) using a C++ machine learning (ML) library called TMVA. The inputs to the classifier were modeled from aggregate geometric and spatial information from remote sensors.
- Innovator in high dimensional (500+), correlated parameter regression problems and reduced the number of needed parameters by over 1/2.
- Leader for software documentation and hardware administration in T2K and host university.
- Eliminated measurement bias from 3% to 1% for particle momentum with minimal impact in its variance using non-linear modelling and feature engineering.
- Managed a small HPC cluster for researchers and helped migrate researchers to an externally managed HPC cluster aided with Python scripts.
- Python developer to create a data pipeline, analyze data using Pandas DataFrames, to execute batch jobs among different HPC clusters, and to evaluate the performance and workflow of HPC jobs. The diagnostic tools are built upon the NumPy arrays and Matplotlib visualization libraries.
- Used Git, GitHub, and CVS for version control.

---

## **EDUCATION**

**2019 - Ph.D. in Physics at Colorado State University**

**2012 - B.S. in Physics at University of California at Santa Cruz**

---

## **CERTIFICATIONS / LICENSES**

### **Data Science: Foundations using R Specialization at Coursera**

- **Nov 2022 - R Programming:** [coursera.org/verify/W2ULZZFESUDW](https://coursera.org/verify/W2ULZZFESUDW)  
(<https://github.com/hoganman/hospital-quality>)

### **Genomic Data Science Specialization at Coursera**

- **Nov 2022 - Introduction to Genomic Technologies:** [coursera.org/verify/GTMCK4SLTY9A](https://coursera.org/verify/GTMCK4SLTY9A)

- **Nov 2022** - *Python for Genomic Data Science*: [coursera.org/verify/7U8FERCM3B5G](https://coursera.org/verify/7U8FERCM3B5G) (<https://github.com/hoganman/Python-for-Genomic-Data-Science>)
- **Dec 2022** - *Algorithms for DNA Sequencing*: [coursera.org/verify/LDBN5RM3N7AY](https://coursera.org/verify/LDBN5RM3N7AY) (<https://github.com/hoganman/algorithms-genomic-sequencing>)

### **SQL for Data Science at Coursera**

- **Dec 2019** - *SQL for Data Science*: [coursera.org/verify/E4C5BHG5ZRUL](https://coursera.org/verify/E4C5BHG5ZRUL)

---

## **COMMUNITY INVOLVEMENT**

**Nov 2021-Jun 2022 - Boulder Data Science, Machine Learning, and AI Meetup at Meetup**

**Jan 2019-May 2019 - Colorado State University Data Science Club at Colorado State University**

**Feb 2019-Jan 2020 - Fort Collins Data Science Meetup at Meetup**

**Jan 2019-Jan 2020 - Volunteer at Animal House Rescue and Grooming**

---

## **PUBLIC TALKS**

- “Using the T2K pi-zero detector to measure the near detector constraint for the neutrino oscillation analysis”, CU BOULDER HIGH ENERGY AND NUCLEAR PHYSICS SEMINAR. Boulder, CO. (2019).
- “Data Mining and Machine Learning in Physics”, FORT COLLINS DATA SCIENCE MEETUP. Fort Collins, CO. (2019).
- “Estimating the Secondary Interaction Systematic from Neutrino-Induced Pion Production in T2K”, AMERICAN PHYSICAL SOCIETY. Fort Collins, CO. (2017).
- “Rejecting Non-MIP-Like Tracks using Boosted Decision Trees with the T2K Pi-Zero Subdetector”, AMERICAN PHYSICAL SOCIETY. Salt Lake City, UT. (2016).
- “Improving Background Rejection in the Next Generation Neutrino Oscillation Parameter Measurements”, AMERICAN PHYSICAL SOCIETY. Baltimore, MD. (2015).

---

## **PUBLICATIONS**

- “The Long-Baseline Neutrino Experiment: Exploring Fundamental Symmetries of the Universe”, LBNE Collaboration. ArXiv e-prints, 1307.7335 (2013).

---

## **PERSONAL INTERESTS**

- Japanese culture
- Physics, astronomy, and mathematics
- Early modern to contemporary history
- Equity and equality for all genders, races, and sexes