Jesús M. Rueda-Becerril — Ph.D.

3902 S Ferdinand St – Seattle, 98118 WA
jm.ruebe@gmail.com • altjerue.github.io • jeruebe • altjerue
0000-0003-1988-1912 • jm.ruebe

Profile

Astrophysicist with industry experience, and a robust foundation in mathematics and statistics. Eager to leverage my skills in quantitative analysis, programming, and problem-solving to design, develop, and implement scalable data science and machine learning solutions.

Experience

Spacial Data Scientist **May 2025 – Present**

TealWaters

- Lead a project to develop an optimized code that modeled topographic terrains, and calculates elevation derivatives.
- Build tools in Python to process and analyze geospatial data.
- Extensively used QGIS for visualization and analysis.
- Used the Wetland Intrinsic Potential (WIP) tool that uses a random forest model to derive maps of wetland probability in the Skykomish watershed.
- Performed EDA, data preparation (buffer, clip, merge, join, mask, reproject), feature extraction, and analysis of multi-scale topographic features for ML/AI models, and provided critical feedback on issues of data integrity; performed critical reviews of data integrity issues and proposed long-term fixes.
- Processed multispectral imagery (Sentinel 1 and 2) to include them in the stack of features to classify land-cover categories using random forest.
- Shared knowledge and results to managers, and key decision makers.
- Trained team members on how to utilize the WIP tool.
- Worked closely with software engineers and scientific team to transition research prototypes into production-quality.

Software Engineer **Apr 2022 – Jan 2024**

Paychex

- Conducted data preparation, validation, and analysis in SQL from Oracle EBS datasets.
- Developed Java Kafka consumers for the streamlined transfer of large volumes of client data across databases. Developed high quality code using Java, Spring Boot, Kafka, PL/SQL, and deployed into production using Jenkins and OpenShift, following software development best practices.
- Collaborated with stakeholders, other software developers and engineers, and senior leadership to assess product needs and meet code standards for continuous integration model.
- Created Splunk dashboards and alerts for analysis of production data.
- Developed and deployed Python tests to ensure software quality and continuous integration.

Postdoctoral Research Scientist **Jan 2018 – Apr 2022**

UMSNH (Mexico), Purdue University, Rochester Institute of Technology

- Led a team of specialists on a NSF-sponsored project and successfully upgraded a scientific code to perform HPC simulations of supermassive black hole binaries. Designed the experiments and evaluated state-of-the-art mathematic and numerical algorithms implemented in a scientific code by running simulations at *Frontera* supercomputer (TACC, UT at Austin). Worked in a detail-oriented manner to successfully benchmark and identify performance optimization opportunities of the scientific code.
- Worked in a highly collaborative environment with multi-institutional, cross-functional teams developing large-scale HPC simulations of neutron star mergers. This collaboration produced 2 papers with high impact results.
- Successfully applied and obtained a $68,000 NASA grant for one year as primary researcher. Managed 3 Ph.D. researchers to study the nature of radiation from active galaxies with state-of-the-art mathematical and numerical methods. This work produced 3 papers and 2 proceedings with high impact results, and 5 presentations to technical audiences.
- Designed the experiments to apply mathematical and numerical methods to prove a hypothesis about the origin and nature of radiation from active galaxies. Used Python to run regressions on observations from NASA telescopes.
- Participated in weekly paper discussion and knowledge sharing at the Purdue Astronomy Journal Club.
- Developed a Python script that processed images for a machine learning (SVM) training framework.
- Conducted a workshop to train and share knowledge with graduate students on creating and manipulating high-volume datasets in HDF5 format.

Graduate Research Assistant **Oct 2009 – Jul 2017**

UMSNH (Mexico), Universitat de València (Spain)

- Developed Shell and Python tools for automation and pipelines for data processing and curation of large volume datasets in HDF5 format from large-scale simulations, ensuring data quality and integrity for downstream analysis.
- Implemented sophisticated numerical tools and data handling to calculate both discrete and continuous spectra from particle distribution functions with arbitrary shape, without impacting simulation runtime.
- Used Python and R for exploratory data analysis on datasets from NASA telescopes, performing linear and non-linear regression, pattern recognition, and forecasting. This work produced 2 papers and communicated my results to technical audiences in multiple international meetings.
- Successfully developed a scientific code using a fourth order Runge-Kutta solver to study the behavior of light near black holes. This work produced one scientific paper with high impact results.
- Attended workshops on Data Analysis and Machine Learning with Python.

Skills

Proficient: Python, QGIS, R, C/C++, Unix/Linux, Shell, git, HPC, LaTeX, Fortran 95, HDF5

Familiar: SQL, Julia, Docker, Slurm

Basic: MongoDB, Rust, Splunk, Jenkins, OpenShift, Java

Education

Ph.D. in Physics **Oct. 2011 – Jul. 2017**

Universitat de València, Valencia, Spain Excellent *Cum Laude*

M.Sc. in Physics **Aug. 2009 – Sep. 2011**

Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Mexico

B.Sc. in Physics **Aug. 2004 – Dec. 2008**

Universidad Autónoma del Estado de México, Toluca, Mexico *"Dr. Juan Josafat Pichardo" Award*

Projects

Wind and Spores **Aug 2024 – Present**

Open Source Code for Dispersion of Spores in Hilly Terrains *GitHub*

- Simulate the dispersion of spores in hilly terrains in R and C++.

Tleco **Jan 2024 – Sep 2024**

Open Source Code for Simulation of Relativistic Particles Radiation *GitHub*

- `Tleco` stands for both *in the fire* and *rise* in the Nahuatl language.
- Numerical code that simulates particles in relativistic plasma, and the rise of radiation from accelerating particles.
- Consists of both Rust functions and Python functions previously built in the Fortran code `Paramo`.

Paramo **Oct 2018 – Apr 2022**

Open Source Code for Radiative Transfer Simulations in Relativistic Astrophysics *GitHub*

- Independently developed this code for distributed settings to perform HPC simulations of radiative transfer in relativistic astrophysics.
- Optimized the code with OpenMP to reduce simulation time from 2 minutes to 5 seconds.
- Researched and applied mathematical concepts of machine learning (gradient descent) to adjust the parameters of the code to classify observations from NASA telescopes.
- This code has been used for at least 5 scientific publications and also for graduate pedagogical purposes.
- Developed data analysis and data visualization tools in Python.

Co-Founder, Mexican Scientists Abroad **Aug 2019 – Feb 2021**

Group of Mexican Scientists narrowing the gap between science and common knowledge *Homepage*

- Collaborated with a group of multidisciplinary scientists to develop Python scripts for statistically modeling and forecasting the COVID-19 outbreak in Mexico.
- Wrote blogposts and infographics for non-technical Spanish-speaking populations to reduce the spread of misinformation.

Certifications and Credentials

5-Day Gen AI Intensive — **Kaggle**
Generative AI

Data Analyst with Python — **DataCamp**
Data Maniputalion, Data Visualization, Importing & Cleaning Data

Data Scientist Professional with Python — **DataCamp**
Python Programming, Data Science, Data Communication, Machine Learning

Machine Learning Scientist with Python — **DataCamp**
Machine Learning, NLP, Deep Learning, Image Processing, Big Data

Mathematical Foundations of Machine Learning — **Udemy**
Python Programming, Data Science, Statistics, Machine Learning

Python for Statistical Analysis — **Udemy**
Python Programming, Data Science, Data Communication, Machine Learning