

Héctor Inda Díaz

(530) 760-6198 | indahector@gmail.com | hectorindadiaz.com

Technical Skills

PROGRAMMING LANGUAGES: Python | C | C++ | Fortran | Matlab | Ferret | CUDA | Numerical Methods | SQL | R | VirtualBox | Docker | Google Cloud Run

DATA SCIENCE & MACHINE LEARNING: Data Visualization (Plotly, Seaborn, matplotlib) | AWS | Pytorch | scikit-learn | Numerical & Statistical Modelling | Quantitative Analysis

INDUSTRY KNOWLEDGE: Languages (English, Spanish, Russian) | Geospatial Data | Scientific Writing | Literature Review

Education

Data Scientist Certification - Fellowship Program, The Data Incubator | August 2023

Ph.D. in Atmospheric Science | University of California Davis | Apr 2021

M.S. in Physical Oceanography | Ensenada Center for Scientific Research and Higher Education | Apr 2015

B.S. in Physics | National Autonomous University of Mexico | Jul 2012

Experience

EAGLE ROCK ANALYTICS

July 2023 – current

Research associate

- Support developing and performing novel research on cloud-based climate and weather data platforms for multiple State of California agencies (e.g., energy and climate-adjacent).
- Support the development of the Cal-Adapt: Analytics Engine, an associated historic data platform of weather station data, and early work to develop a climate risk index with the California Air Resources Board.
- Development of an open-source historical data platform synthesizing surface meteorological observational networks across the western United States from just under 16,000 stations, including the development of robust quality assurance/quality control protocols.
- Sub-consultant for the *UCSD Decarbonization, Electrification, and Sustainability Planning Study*. Supporting the identification of climate projection data relevant to UCSD's facilities and operations, which is necessary for assessing climate vulnerabilities and potential climate impacts.

LAWRENCE BERKELEY NATIONAL LABORATORY

Jun 2022 – June 2023

Postdoctoral Scholar

- Leveraged Indiana University's BigRed200 supercomputer to port and execute the RRM-E3SM model from NERSC, successfully generating 200TB of historical and future projection model output over a period of six months.
- Implemented dask parallel computing and SLURMClustering on BigRed200 supercomputer, enabling parallelized analysis of over 200TB of data across 30 computing nodes (~3840 cores), significantly reducing data analysis time from potential months to just one week.
- Conducted monthly, seasonal, yearly, and climatological analyses of 100 years of model data (more than 20TB) and regridded output and analysis files. - Successfully established and executed the Regionally Refined Energy Exascale Earth System Model (RRM-E3SM).
- Designed and implemented a grid for RRM-E3SM at 14 km resolution, targeting Mexico and Southeast US, demonstrating attention to detail and spatial analytics expertise.
- Created a Python library, specifically tailored to analyze and visualize RRM-E3SM output, enhancing data interpretation and reporting.
- Expertly handled and analyzed an extensive dataset of over 400TB, deriving insights from large-scale model output.

Graduate Student Research Assistant

- Conducted statistical analysis on evapotranspiration and heat stress in the central US under varying climates and assessed convective self-aggregation in convection-resolving models.
- Significantly reduced uncertainty in atmospheric rivers (ARs) analysis through advanced statistical and data analysis techniques, including Kernel Density Estimation and extreme value theory, showcasing skills in complex environmental data analysis.
- Developed a Python library for self-aggregation data in CRM output and devised massively parallel data analysis routines for large climate model ensembles, demonstrating expertise in scalable programming and parallel computing.
- Implemented five statistical AR size estimation methods, including a 3D Lagrangian backtrajectory model in C/CUDA with CPU and GPU compatibility.
- Analyzed hundreds of TBs of historical and future climate simulations data from various sources including CMIP5 and CMIP6, demonstrating proficiency in handling and interpreting large datasets.
- Contributed to the scientific community by publishing peer-reviewed literature and presenting findings at international conferences.