Erdem Karaköylü

University Park, MD
erdemk@protonmail.com | erdemkarakoylu.github.io | LinkedIn

Data Scientist | Bayesian Modeling & Machine Learning Specialist

Data scientist specializing in Bayesian modeling, uncertainty-aware machine learning for decision optimization. Experienced in developing and evaluating probabilistic models in high-stakes domains including defense, remote sensing, and environmental prediction. Strong track record of applying principled statistics to improve system performance and experimental design.

Core Skills

Experimentation & Inference
A/B Testing · Bayesian Decision Theory · Experimental Design · Causal Inference (actively developing incl. do-calculus in PyMC)

Bayesian & Statistical Modeling
Hierarchical Modeling · Bayesian Additive Regression Trees (BART) · Monte Carlo Simulation · Probabilistic Programming

Systems & Performance Optimization
Model Evaluation · Uncertainty Quantification · ML System Diagnostics

Languages & Tools
Python · PyMC · Pandas · ArviZ · Scikit-learn · XGBoost · PyTorch · Git · Matplotlib · Seaborn · SQL

Select Experience & Research Highlights

Freelance Data Scientist – Bayesian Modeling for Remote Sensing & Forecasting Systems
2025 – Present

- BART (on-going study) and hierarchical Bayesian models (paper in prep) to infer a variety of marine processes, with predictive uncertainty estimation.
- XGBoost models to infer phytoplankton community structure from hyperspectral data, and strived to explain prediction results using Shapley values (paper in prep).
- Probabilistic parameter estimation for nonlinear ecological ODE-based models.

Data Scientist – Research Innovations Inc. (RII)
ML/NLP Systems for Defense & Justice Applications
2019 – 2025

- Retrieval-Augmented Generation pipeline design to support mission-specific planning for military use cases.
- Bayesian A/B tests to evaluate model changes and optimize system performance in live environments.
- Active learning loop to reduce annotation costs in large-scale image classification.
- Targeted-sentiment analysis models using fine-tuned LLMs for sensitive content.

Machine Learning Researcher – NASA Goddard / SAIC
Probabilistic Modeling for Ocean Color Products
2013 – 2019

- Bayesian models for inherent optical properties and chlorophyll from satellite-derived surface reflectance.
- Monte Carlo simulations to assess uncertainty in satellite sensor outputs and downstream biogeophysical products.
- Visualizations and statistical analyses for Earth observation reports and mission deliverables.
- Tutorials on Bayesian analysis for ocean color remote sensing practitioners.

Researcher – University of Maryland
Modeling of Fish Spawning Behavior
2011 – 2013 - Coupling of circulation model with Menhaden spawning model for recruitment prediction,

Researcher – UC San Diego / Scripps Institution of Oceanography
Plankton Ecology / Circulation Models
2003 – 2011

- Adapted a planar laser-induced fluorescence imaging system to quantify real-time feeding states in individual marine zooplankton.
- Captured high-resolution time series of gut pigment dynamics to infer behavioral state transitions (feeding, digestion, resting).
- Built an individual-based model linking physiological state to vertical foraging behavior under environmental constraints.

Education

- Ph.D. – Biological Oceanography & Marine Ecology - *Scripps Institution of Oceanography, UC San Diego*
- B.Sc. – Oceanography - *Florida Institute of Technology*

Languages

- **English** – Native/Trilingual
- **French** – Native/Trilingual
- **Turkish** – Native/Trilingual
- **Spanish** – Advanced