**Erdem Karaköylü**

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

---

## Data Scientist | Bayesian Modeling & Machine Learning Specialist

Practical AI expert specializing in Bayesian methods, LLM pipelines (RAG), and predictive modeling across defense, climate, and environmental domains. Builds robust, uncertainty-aware systems to enhance decision-making.

---

## Core Skills

**Bayesian & Statistical Modeling**
Hierarchical Modeling (Regression & Classification) · Bayesian Additive Regression Trees (BART) · Monte Carlo Simulation · Probabilistic Programming · A/B Testing

**Emerging Interests**
Causal Inference (incl. do-calculus in PyMC) · Bayesian Decision Theory

**Technical Stack**
Python · PyMC · Scikit-learn · XGBoost · PyTorch · Pandas · Git · Matplotlib · Seaborn · ArviZ · XArray · SQL

---

## Experience & Research Highlights

**Freelance Data Scientist – Marine Remote Sensing and Ecological Forecasting**
*Bayesian modeling and machine learning for environmental data*

- Developed Bayesian Additive Regression Tree (BART) and hierarchical models to estimate marine optical properties and chlorophyll concentrations from satellite radiance data.
- Built predictive XGBoost models to infer phytoplankton community structure, outperforming baseline approaches.
- Used probabilistic ODE parameter estimation to analyze nonlinear dynamics in marine ecological systems.
- Published reproducible Bayesian modeling workflows as open-source Jupyter notebooks for environmental science.

**Data Scientist – Research Innovations Inc. (Alexandria, VA)**
*DOD and DOJ-focused machine learning and NLP systems*

- Contributed to the development of a Retrieval-Augmented Generation (RAG) system that improved information retrieval for military planners.
- Led Bayesian A/B testing to optimize system components and refine model selection for production environments.
- Built and iteratively refined an active-learning image classification pipeline to reduce manual annotation requirements.

- Supported targeted sentiment analysis using fine-tuned large language models for sensitive domains.

**Machine Learning Researcher – NASA Goddard Ocean Biology Processing Group / SAIC**
*Earth observation and probabilistic modeling for ocean color remote sensing*

- Developed Bayesian models to predict satellite-derived ocean color products, improving chlorophyll and particulate property estimates.
- Conducted Monte Carlo simulations to quantify uncertainty and error propagation in remote sensing reflectance (Rrs) data.
- Created climate data visualizations and analysis pipelines supporting scientific reports and satellite mission deliverables.
- Advocated for probabilistic approaches and led internal discussions on Bayesian methods for biogeophysical modeling.

**Researcher – UC San Diego / Scripps Institution of Oceanography**
*Thesis: Foraging Sorties Hypothesis – Inferring Behavioral Rhythms in Marine Primary Consumers*

- Adapted a planar laser-induced fluorescence (PLIF) 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.
- Calibrated imaging measurements against chemical extraction to ensure accuracy and repeatability across individuals.
- Findings published in *Limnology and Oceanography Methods* (2009); follow-up research on temperature effects published in *Journal of Plankton Research* (2012), where a gut fluorescence image was featured on the journal cover.

---

## 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