KEVIN MATTHEW DUYST

559.308.0931 · matthew.duyst@gmail.com · linkedin.com/in/matthew-duyst · Austin, TX

Environmental data scientist specializing in production ML systems for geospatial environmental quantification, with applied experience spanning carbon markets, atmospheric methane monitoring, and land-use change science across satellite, UAV, and ground-truth data sources. Currently at Grassroots Carbon, the largest soil carbon removal company in the United States, engineering end-to-end ML and data systems that quantify soil carbon and biomass at ranch and regional scales across two million acres of US rangeland, with verification outputs calibrated to SEP v2, VM0042, and CAR-SEP crediting standards. Prior work includes a first-author publication at SPE ATCE 2024 from UAV-based methane research at SeekOps, a NASA JPL spin-off; a fully funded MESc from Yale, where graduate research developed a DeepLabV3+ CNN to hindcast pre-satellite methane emissions from classified land-use distributions; and an Honorable Mention at the 2022 Global Stocktake Climate Datathon, competing among 100+ participants from 30 countries ahead of COP 27. Technical foundation spans geospatial data engineering, ensemble and deep learning architectures, multi-source satellite image analysis, and regulatory-grade scientific reporting.

EDUCATION

Yale University

Master of Environmental Science, Machine Learning, Remote Sensing & Spatial Statistics · 2021–2023

Fully funded as a Yale Merit Research Scholar and Hixon Center for Urban Sustainability Fellow. Graduate Teaching Fellow for Regression Modeling of Ecological and Environmental Data. Coursework spanned machine learning, computer vision, spatial statistics, multivariate environmental data analysis, and remote sensing, with additional graduate-level CS through Yale's department. Cross-departmental research integrated thermal and RGB UAV imagery for photovoltaic panel performance assessment across Yale's West Campus. Honorable Mention, 2022 Global Stocktake Climate Datathon, ahead of COP 27.

University of California, Los Angeles

Bachelor of Arts, Geography and Environmental Studies · Minors: GIS and English · 2014–2018

UCLA Regents Scholar and Undergraduate Research Scholars Program recipient. Coursework covered the full geospatial stack: supervised and unsupervised image classification, spectral index analysis (NDVI, NBR), SAR interpretation, DEM construction, spatial interpolation and autocorrelation, SQL-based spatial querying, cartographic design, and predictive environmental modeling. Undergraduate thesis on post-wildfire vegetation change detection using Landsat and Google Earth Engine, presented at the 2019 Los Angeles Geospatial Summit.

PROFESSIONAL EXPERIENCE

Grassroots Carbon Feb 2025 – Present

Environmental Data Scientist, Machine Learning · San Antonio, TX · Hybrid

■ Build and maintain the data infrastructure underpinning soil carbon verification spanning two million acres of US rangeland. The system ingests and harmonizes sparse ground-truth field observations from multiple monitoring networks, fusing them with satellite imagery, spectral indices, terrain derivatives, and soil and climate context to deliver training-ready datasets into downstream ML workflows. Ranch-level SOC stratification is designed around SEP v2 compliance, where clustering geophysically similar areas directly reduces the sampling uncertainty that drives credit deductions.

■ Engineer shared covariate extraction tooling: a Python client querying five authoritative geospatial sources through a unified API, built for cross-team use following established architectural patterns. The system is designed to scale as Grassroots Carbon expands into new carbon markets and crediting standards, consolidating what were previously fragmented extraction workflows into a single maintained codebase.

■ Benchmark biomass and structural vegetation estimation architectures spanning deep learning, ensemble methods, and process-based ecological models, with outputs supporting quantification under the Regenerative Standard, VM0042, and CAR-SEP. Modeling extends into counterfactual baseline scenarios that reconstruct expected productivity under business-as-usual land management, a required verification step under each crediting methodology.

■ Lead technology planning for the company's earth observation capabilities, evaluating remote sensing platforms, sensor networks, and data sources covering structural sensing, phenological calibration, spectral imaging, and soil moisture monitoring. Analysis directly informs the long-term geospatial technology roadmap as Grassroots Carbon scales into new carbon markets and crediting standards.

SeekOps Inc. Jan 2024 – Dec 2024

Environmental Data Scientist · Austin, TX · Hybrid

■ As one of three scientists on the Science team reporting directly to the co-founder and CTO, led a quantitative assessment of methane flare destruction removal efficiency (CH4 DRE%) across the Bakken and Permian basins. Deployed UAV-mounted TDLAS sensors, ultrasonic anemometers, and RTK ground stations, calculating DRE% through mass-balance equations comparing uncombusted CH4 flux in flare plumes against measured inflow rates and examining the influence of key operational parameters (flowrate, temperature, pressure). Research was published as first author and presented at SPE ATCE 2024, contributing to enhanced LDAR programs across two major US shale basins.

■ Designed and executed an integrated QA/QC workflow for UAV-based methane detection: LiDAR terrain corrections, surface roughness adjustments for airflow disruptions near oil and gas facilities, and wind vector validation through AirData telemetry. Modeled plume dynamics via Navier-Stokes equations and mass conservation, producing high-resolution emission source maps aligned with OGMP and EPA reporting standards.

■ Trained a fine-tuned YOLOv8 object detection model to classify oil and gas surface equipment (wellheads, compressors, storage tanks, flare stacks) from high-resolution aerial imagery, enabling autonomous site characterization during survey flights. Managed the end-to-end ML research cycle: imagery collection and annotation, dataset construction, transfer learning, hyperparameter optimization, and held-out validation.

■ Built client-facing and internal geospatial visualization suites in Python (Folium, Plotly Dash) and GIS platforms (QGIS, ArcGIS Pro) to communicate spatial emissions distributions, plume dynamics, and site-level attribution across survey campaigns, and developed predictive emissions modeling prototypes to extend the team's measurement capabilities beyond direct UAV observation windows.

Yale School of the Environment – The Lee Lab Aug 2021 – May 2023

Graduate Research Scientist, Remote Sensing & ML · New Haven, CT

■ Investigated the relationship between urbanization, rice cultivation, and methane emissions throughout China's Yangtze River Delta, home to over 150 million people and some of China's most productive rice-growing regions. Classified multi-decadal land cover from 30m Landsat imagery in Google Earth Engine, delineating urban extent via the GAIA dataset and mapping rice paddies through the Phenology and Pixel-Based Paddy Mapping algorithm, integrating vegetation indices (NDVI, EVI, LSWI) with China's cropping calendar and biophysical flooding signals to resolve spectral overlap with similar crops.

■ Monitored atmospheric CH4 using Sentinel-5P TROPOMI column-averaged methane concentrations (XCH4), then applied Getis-Ord Gi* and Global Moran's I to quantify spatial clustering of emission anomalies and map their correlation with classified land-use patterns, tracing decades of urban and agricultural transformation throughout the delta.

■ Addressed a core data gap in satellite-based atmospheric science: orbital CH4 monitoring began only in 2017, yet the land-use dynamics under study spanned decades prior. Developed a convolutional neural network on DeepLabV3+ architecture to hindcast historical CH4 emission patterns from classified land-use distributions, leveraging atrous convolution and spatial pyramid pooling to extend the monitoring record into the pre-satellite era. This work formed the basis of the MESc thesis, demonstrating that segmentation-trained CNNs can reconstruct atmospheric emissions records from land-use proxy distributions.

University of Minnesota – Marcell Experimental Forest May – Aug 2022

Visiting Research Scientist · Minneapolis, MN

■ Conducted field and computational research in Dr. Timothy Griffis' lab at one of the world's longest-running peatland monitoring programs. Instrumented the field data collection using eddy covariance towers, ultrasonic anemometers, and greenhouse gas analyzers under AmeriFlux protocols, capturing CH4 concentration gradients at fine temporal resolution. Applied feature importance rankings and seasonal decomposition to isolate dominant environmental drivers (tempserature, precipitation, CO2 flux) and inform univariate and multivariate forecasting model structure.

■ Constructed an ensemble stacking model layering LightGBM, SARIMAX, FB Prophet, and XGBoost alongside OLS regression, decomposing the CH4 flux signal into trend, seasonal, and residual components with each model exploiting its respective architectural strength. The ensemble achieved 87% explained variance and extended reliable CH4 flux forecasts to 52 weeks, with statistical validity confirmed through Durbin-Watson autocorrelation and Dickey-Fuller stationarity testing.

Cydcor Jan 2019 – Jun 2020

Geospatial Analyst · Los Angeles, CA

■ Applied GIS and spatial statistics to a non-traditional domain: optimizing sales territory planning across door-to-door and B2B campaigns. Working directly with executive leadership, integrated census tract demographics with internal sales metrics in ArcGIS using kriging and spatial interpolation to identify high-propensity regions. Built predictive geospatial models using spatial regression and weighted overlay analysis for expansion into unsampled markets. Automated geocoding workflows in Python against the Google API, with batch geocoding and spatial data validation pipelines improving targeting precision and throughput for high-volume campaigns.

TECHNICAL SKILLS

Programming & Engineering
Python, R, SQL, Docker, Microsoft Azure, ParaView, OpenFOAM

Machine Learning & Deep Learning
PyTorch, Keras, TensorFlow, Scikit-learn, XGBoost, LightGBM, Random Forest, SVM, ARIMA/SARIMAX, FB Prophet, Ensemble Stacking, Transfer Learning

GIS & Earth Observation Platforms
QGIS, ArcGIS Pro, Google Earth Engine (Python & JS API), ENVI

Remote Sensing & Image Analysis
Computer Vision, Image Classification, Change Detection, LiDAR, Photogrammetry, Landsat, Sentinel-5P TROPOMI, SAR Interpretation, Spectral Index Analysis (NDVI, EVI, NBR, LSWI)

Geospatial Data Science
GeoPandas, Rasterio, GDAL, NumPy, Pandas, Statsmodels, Folium, Plotly Dash, Matplotlib, Spatial Autocorrelation, Eddy Covariance, GHG Modeling

AWARDS & RECOGNITION

SPE ATCE 2024 – First Author Publication *UAV-TDLAS CH4 Flare Destruction Removal Efficiency, Bakken & Permian · 2024*

Yale Merit Research Scholar / Dean's Scholar / Hixon Center Fellow *Fully Funded · 2021–2023*

Global Stocktake Climate Datathon – Honorable Mention *COP 27 · 2022*

Los Angeles Geospatial Summit – Guest Speaker *2019*

UCLA Undergraduate Research Scholars Program / Regents Scholar *2014–2018*