Kevin Matthew Duyst

Austin, TX | 559.308.0931 | matthew.duyst@gmail.com

linkedin.com/in/matthew-duyst | github.com/matt-duyst | mappingwithmatt.com

RESEARCH INTERESTS

I build machine-learning models that combine data from satellites and drones, atmospheric and ground-based sensors, and field measurements to detect, measure, and predict environmental conditions. What ties the work together is not the subject but what rides on the result: each measurement feeds a real decision, so it has to be right where it counts and honest about where it is not. The difficulty is that two sources rarely agree. Each was built for its own purpose, at its own resolution and revisit time, dense in some places and sparse in others. Everything downstream, every detection, measurement, and prediction, rests on first reconciling them into one measurement you can rely on.

That problem has run through everything I have worked on, whatever the subject. It first drew me in as an undergraduate, mapping the severity of California's largest wildfire from how the burned land reflected light before and after. In my master's thesis it became the center of the work. I trained a deep network to label every pixel of the one year when satellite methane, urban extent, and rice paddies were all observed together over China, then used it to reconstruct methane across the decades before the satellites existed. At a peatland in Minnesota, I combined several models to forecast a year of carbon-dioxide and methane exchange from a single tower's record. At flare sites across the Bakken and Permian, I trained a model to find oil-and-gas equipment from the air and measured how much methane each flare failed to burn. Now, at continental scale, I predict grassland growth at any point across US rangelands with a model that pairs the known science of how grass grows with machine learning. The subject changes and the method changes, but the problem does not.

A doctorate is where I want to work on what makes these measurements hold up in the cases that matter most. The problem I most want to take further is what happens when a model leaves the ground it was trained on and has to answer for somewhere it has never seen: a different region, a later year, an ecosystem unlike anything in its training data. Foundation models change what is possible here. Trained on vast archives of imagery, they now carry a general representation of nearly every place on Earth, and I want to know how far that representation reaches: whether it can carry a reliable prediction onto unfamiliar ground, and whether a model can be honest about where it cannot. Physics is a second lever. A model that holds the behavior of an ecosystem inside it, its estimate of stored carbon bound to how sunlight, water, and growth actually move, stays reliable where a model fit to data alone quietly breaks down. Timing is the last of these. The processing that makes a measurement trustworthy (reconciling the sources, grounding it in physics, quantifying what is uncertain), is the same processing that makes it slow, so the most reliable number often arrives only after the moment it described has passed. I want to work the other side of that trade-off: measurement fast enough to act on while it still matters, without giving up what makes it worth trusting.

The future of this work is open. A measurement should earn trust by being checkable, not by the reputation of whoever produced it; the data, the code, and the models should be there for anyone to run, verify, and improve, built by a community rather than locked inside the institutions that can afford to produce them. This is a matter of fairness as much as rigor. Ground data is plentiful where sampling is easy and well funded and scarce almost everywhere else, so a model comes to know least about the places that need it most. Open, shared tools are what keep that gap from hardening into a permanent divide.

I have worked on these problems in startups, in applied research labs, and in universities, and seen where models hold up under real conditions and where they still fall short. A doctorate is the place to take on what those settings leave unfinished: making a measurement that generalizes to new ground, stays true to the physics, arrives in time, and lies open for anyone to check: that's what lets the decisions resting on it hold.

EDUCATION

Yale University, School of the Environment 2021–2023

Master of Environmental Science (M.E.Sc.). Focus: Machine Learning, Remote Sensing, and Spatial Statistics.

Thesis: *Urban-Methane Transfers (2000–2018): A Convolutional Neural Network Approach to Forecasting Historical XCH_4 Emissions Through Urban Boundaries and Paddied Rice Extents Along China's YRD*. Reconstructed multi-decadal atmospheric methane distributions across the YRD's four provinces by training a DeepLabv3+ semantic-segmentation convolutional neural network on the 2018 simultaneity of Landsat-derived urban and rice extents with Sentinel-5P TROPOMI XCH_4 observations, then projecting the learned spatial relationships back across the pre-satellite window to deliver the first integration of urban-rice land-use coupling with deep learning for multi-decadal methane reconstruction.

Advisor: Prof. Xuhui Lee, Sara Shallenberger Brown Professor of Climate Science; Director, Yale Center for Earth Observation.

Committee: Prof. Tim Gregoire (J.P. Weyerhaeuser Jr. Professor of Forest Biometrics & Statistics; Fellow, American Statistical Association); Prof. Karen Seto (Frederick C. Hixon Professor of Geography and Urbanization Science; U.S. National Academy of Sciences; IPCC Coordinating Lead Author, AR5 & AR6); Prof. Alex Wong (Computer Science; Director, Yale Vision Laboratory).

Honors: Yale Merit Research Scholar (fully funded); Hixon Center for Urban Sustainability Fellow.

Selected coursework: Machine learning and deep learning through Yale Computer Science, including Deep Learning Theory and Applications (CPSC 452), Introduction to Computer Vision (CPSC 480), Introduction to Machine Learning (CPSC 481), and Real World Environmental Data Science (ENV 617); remote sensing through graduate seminars with Karen Seto on Remote Sensing of Land-Cover and Land-Use Change (ENV 725) and Urbanization, Global Change and Sustainability (ENV 645), alongside Workshop on Remote Sensing and Photogrammetry with Drones (ENV 704) and Observing Earth from Space (ENV 726/EPS 362); spatial and applied statistics through Applied Spatial Statistics (ENV 781/S&DS 674), Multivariate Statistical Methods for Environmental Sciences (ENV 758), and Regression Modeling of Ecological and Environmental Data (ENV 753). Continued as Graduate Teaching Fellow for ENV 753.

Additional research: Co-authored *Toward UAV Decision-Support Tools for Photovoltaic Arrays: A Coastal Connecticut Thermal Mosaicking Case Study* (Anthony, Duyst, Verheyden, Zhao; Yale ENV 704), integrating thermal and RGB UAV imagery for photovoltaic panel performance assessment across Yale's West Campus. The project produced a four-band orthomosaic with radiometric calibration against ground-truth solar panel readings ($R^2 = 0.81$), correlating thermal signatures with measured energy output across 4,400 panels.

University of California, Los Angeles *2014–2018*

Bachelor of Arts, Geography and Environmental Studies. Minors in Geospatial Information Systems and Technologies, and English. GPA: 3.67/4.00.

Thesis: *Mendocino Fire Complex: Calculating Surface Vegetation Fluctuations through Spectral Indices (NDVI & NBR)*. Paired NDVI and differenced Normalized Burn Ratio (dNBR) analyses of the 2018 Mendocino Complex, California's largest wildfire in state history, which burned roughly 450,000 acres across Mendocino County and the Mendocino National Forest between July and October 2018. Processed Landsat 8 OLI/TIRS C1 Level-1 imagery from USGS Earth Explorer across two temporally-distinct scene pairs: a July 10 / October 30 pair for NDVI seasonal differencing, and a tighter July 23 / October 11 pair for the four-day-pre / one-week-post NBR pairing that resolves immediate burn severity before regrowth confounds the spectral signal. Compared a manually-staged ENVI pipeline (radiometric calibration to top-of-atmosphere reflectance, band-threshold ROI water masking against Clear Lake's NIR-low-reflectance signature, Brightness Temperature calibration of Thermal Band 1, Layer-Stacking, and Band Math differencing for dNBR) against a Google Earth Engine UN-SPIDER pipeline of algorithmic cloud-masking and composited atmospheric correction, selecting the GEE pipeline as the operational classifier after cross-platform validation against the USGS dNBR burn-severity scheme. NDVI differencing quantified a mean ΔNDVI of -0.2395 across the burn extent (pre-fire mean 0.3288 falling to post-fire mean 0.0893), while dNBR classification resolved fire severity from moderate-low through high-severity bands across the affected Mendocino National Forest interior. Advised by Dr. Nick Burkhart (UCLA Department of Geography), with additional remote sensing mentorship from Geoffrey Fricker. Presented at the 2019 Los Angeles Geospatial Summit, USC Spatial Sciences Institute.

Honors: UCLA Regents Scholar; Undergraduate Research Scholars Program (URSP); Dean's Honors List (five quarters).

Selected coursework: Geospatial methods sequence, including Geographic Information Systems (GEOG 7), Environmental Modeling (GEOG 166), Cartography (GEOG 167), Intermediate Geographic Information Systems (GEOG 168), Remote Sensing & GIS (GEOG 169), and Advanced Geographic Information Systems (GEOG 170); climate science and ecology in Biodiversity in a Changing World (GEOG 2), People & Earth Ecosystems (GEOG 5), Forest Ecosystems (GEOG 111), Applied Climatology (A&O SCI M106), and Climate Change (A&O SCI 1). Twenty-plus course-based projects across remote sensing, supervised and unsupervised classification, SAR analysis, DEM construction, spectral indices, SQL-based buffer analysis, predictive modeling, raster analysis, spatial interpolation, and cartography are documented at mappingwithmatt.com.

PUBLICATIONS & PRESENTATIONS

Publications

Duyst, K. M. et al. (2024). *Quantification of Methane Flare Destruction Removal Efficiency (CH_4 DRE%) Across the Bakken and Permian Basins Using UAV-Mounted TDLAS Sensors*. Society of Petroleum Engineers Annual Technical Conference and Exhibition (SPE ATCE 2024), New Orleans, LA. First author, peer-reviewed conference proceedings.

Presentations

Duyst, K. M. (2022). "Urban-Methane Transfers: A Spatiotemporal Analysis of Atmospheric Emissions Along the YRD." Hixon Center for Urban Sustainability Fellow Presentation, Yale School of the Environment, New Haven, CT.

Duyst, K. M. (2019). "Mendocino Fire Complex: Calculating Surface Vegetation Fluctuations through Spectral Indices (NDVI & NBR)." Poster presentation, *Los Angeles Geospatial Summit*, USC Spatial Sciences Institute, Los Angeles, CA.

RESEARCH EXPERIENCE

Grassroots Carbon *Feb 2025 – present*

Environmental Data Scientist, Machine Learning. San Antonio, TX (Hybrid).

- Develop the methodological core of a soil carbon credit verification system for continental US rangeland under the Climate Action Reserve Soil Enrichment Protocol (CAR-SEP): a hybrid mechanistic-learned biomass modeling framework feeding RothC, a process-based biogeochemical model that determines whether issued credits reflect real climate benefit or accounting artifact. The architecture pairs a deterministic Light Use Efficiency (LUE) backbone, where each site's conversion of intercepted sunlight into biomass is calibrated empirically from satellite-observed productivity over the early growing season, with a bounded machine-learning correction trained on the systematic errors the deterministic backbone leaves behind. The correction draws only on causally-constrained inputs (weather, soil, terrain, climate normals) that exclude any vegetation signal influenced by the management practice itself, preserving the counterfactual's independence from the management it is meant to measure. Comparing observed and counterfactual scenarios isolates the soil carbon change attributable to management rather than weather, with held-out ranches testing whether the framework generalizes beyond the calibration sites.
- Build and own the data infrastructure underpinning the modeling work: a multi-source covariate integration pipeline that ingests sparse ground-truth biomass observations from federal monitoring networks, ecological observatories, agricultural research stations, and NASA mission products, and harmonizes them with cloud-native extraction across the federal earth observation archive: POLARIS gridded soils, gSSURGO soil survey, 3DEP terrain, NLCD land cover, NRCS MLRA/LRR ecological zones, PRISM climate normals, GridMET meteorological forcing, and the Rangeland Analysis Platform (RAP) for vegetation cover and biomass. The pipeline integrates the Harmonized Landsat-Sentinel (HLS) spectral suite across structural and metabolic indices (NDVI, EVI, SAVI/MSAVI, TVI, NDRE, NDMI/LSWI, NDWI, NBR/NBR2) alongside Google AlphaEarth Foundations 64-band embeddings, with multi-scalar resolution reconciliation between 10m and 30m grain through nested kernel architectures that resolve fine-scale heterogeneity within coarser baseline pixels. The system delivers training-ready feature manifolds into downstream ML workflows, with ranch-level stratification designed around CAR-SEP sampling design requirements, where clustering geophysically similar areas reduces measurement uncertainty in the carbon stock estimates that anchor verification.
- Architect the engineering foundation underneath the modeling work: a cloud-native data system built on AWS S3, with virtualized streaming from Cloud Optimized GeoTIFF endpoints and OPeNDAP fetch from federal climate servers replacing bulk-download patterns, and parquet artifacts indexed under Uber H3 spatial indexing that serve both point-level coordinate extraction and the gridded surfaces downstream tools require. Above this storage layer runs the orchestrator, a modular covariate integration pipeline that scales from local development to production through the same codebase, with resilience design that recovers cleanly from the transient server failures inherent to fetching across many federal providers at once. The work is maintained on GitHub across a repository network spanning modeling, shared utilities, and adjacent pipelines, with feature-branch development, peer-reviewed pull requests, and passing test suites gating the path into main. Shared geospatial utilities migrated from project code into a team-wide library now serve modeling pipelines across data science, with data provenance documentation and parameter governance (every model constant cataloged with units and source) built into the codebase rather than maintained separately.
- Extend the methodological framework into applied research and validation work across the broader carbon market and ecological geography. Lead the validation expansion that scales the modeling framework from its original calibration set into five IPCC climate zones spanning continental gradients (Cool Temperate Dry and Moist, Warm Temperate Dry, Tropical Dry, Boreal Moist), testing whether site-specific calibration parameters generalize across ecological regimes the framework was not trained on. Coordinate a synchronized multi-source data assimilation across federal monitoring networks, ecological observatories, NASA mission products, university research stations, and program-specific grassland registries to address spectral saturation in dense humid-region canopies that standard rangeland indices systematically under-report. Translate the modeling outputs into operational decision support for the business: a county-level enrollment prospecting dashboard surfacing high-productivity unenrolled rangeland for the enrollment team, and supply-shed scoped biomass rollups feeding Scope 3 emissions reporting where predicted biomass parameterizes emissions intensity rather than sitting next to it. Active research directions include remote-sensing detection of grazing management practices from integration with longitudinal management event data, building toward a continuously-retrained detection model that identifies management state from satellite signals alone.

SeekOps Inc. (NASA JPL spin-off) *Jan 2024 – Dec 2024*

Environmental Data Scientist. Austin, TX (Hybrid).

- Conducted first-author research quantifying methane Destruction Removal Efficiency (CH_4 DRE%) across 42 unique flare systems in the Bakken and Permian basins, published in the peer-reviewed proceedings of the SPE Annual Technical Conference and Exhibition 2024 (SPE-221067-MS). The measurement campaign deployed unmanned aerial vehicles mounted with in-situ tunable diode laser absorption spectroscopy (TDLAS) sensors, comparing the mass flux of unburnt methane exiting each flare plume against the mass flowrate of methane entering the system from flowmeter telemetry, with full accounting for volumetric flowrate, vented gas temperature, and pressure. The dataset spanned 71 total observations across three oil and gas operators, including 27 same-site revisits motivated by initial measurement-quality flags such as hyperbolic efficiency rates and flowmeter discrepancies. This structural reassessment design surfaced 11 cases where revisited systems exceeded the global combustion efficiency benchmark of 92% after originally falling short, reframing flare efficiency assessment as a problem requiring continual rather than one-time measurement and surfacing the data-acquisition improvements needed for credible regional comparison across O&G methane mitigation programs.

- Engineered the integrated quality assurance architecture underneath the measurement campaign, addressing the systematic uncertainty sources that confound mass-balance methane flux estimation in operational field conditions. Terrain effects were corrected via LiDAR-derived surface models capturing topographic perturbations to airflow at flare scale, surface roughness adjustments quantified the airflow disruptions characteristic of densely-built oil and gas facility environments, and wind vector validation via AirData telemetry resolved measurement discrepancies between sensor-anchored anemometer readings and the actual three-dimensional wind field at plume altitude. Plume transport was modeled through coupled Navier-Stokes formulations describing the spatial dynamics of the dispersing flare gas, with conservation of mass anchoring the closed-form mass-balance equation that converts plume flux measurements into source emission rates. The resulting high-resolution emission source attribution maps were structured for direct alignment with Oil and Gas Methane Partnership (OGMP) reporting frameworks and US Environmental Protection Agency (EPA) inventory standards, making the methodology operationally compatible with the regulatory frameworks that govern industry methane disclosure.
- Developed a complementary computer vision research track extending the measurement program beyond direct emission quantification into autonomous oil and gas asset characterization. Fine-tuned the YOLOv8 object detection architecture on high-resolution UAV-collected aerial imagery, training the model to identify the surface equipment that defines methane emission risk profiles at oil and gas facilities: wellheads, compressors, storage tanks, and flare stacks. The end-to-end research cycle spanned imagery acquisition under operational flight conditions, custom annotation protocols for the equipment class taxonomy, transfer learning from pretrained backbones into the operator-specific equipment distribution, hyperparameter optimization across model architecture and training schedule, and held-out validation against site visits the model had not seen during training. The resulting capability enabled survey flights to characterize asset inventories in flight rather than through manual post-flight image review, an operationally relevant capability for the kind of continuous, high-frequency monitoring methane mitigation programs require.
- Translated the measurement methodology and modeling outputs into operational decision support for client operators and internal program design. Built multi-platform geospatial visualization suites in Python (Folium, Plotly Dash) and GIS environments (QGIS, ArcGIS Pro) that communicated spatial emissions distributions across surveyed sites, plume dynamics at the facility scale, and site-level emission attribution back to specific equipment classes: the operational granularity that distinguishes targeted Leak Detection and Repair (LDAR) intervention from generic emission inventory reporting. The published findings on regional flare efficiency variability positioned the work to inform LDAR program design across two major US shale basins, surfacing the data acquisition strategies and continual reassessment patterns that the measurement program demonstrated were required for credible regional comparison. Active research extension included predictive emissions modeling prototypes designed to estimate methane flux at sites and time windows outside direct UAV observation, building toward a measurement-plus-inference framework that would extend campaign coverage beyond the discrete observation windows aerial surveying alone can produce.

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

Graduate Research Scientist, Remote Sensing & Machine Learning. New Haven, CT.

- Investigated the spatial coupling between urbanization, paddied rice cultivation, and atmospheric methane emissions across China's Yangtze River Delta (YRD), a region of 163 million people across four provinces (Anhui, Zhejiang, Jiangsu, Shanghai) generating nearly a quarter of China's GDP and 70% of YRD agricultural output through rice cultivation. Classified multi-decadal land cover from the complete 30m Landsat archive (TM, ETM+, OLI) through Google Earth Engine, delineating urban extent via the Global Artificial Impervious Areas (GAIA) algorithm and mapping paddied rice extent through the Phenology and Pixel-Based Paddy Mapping (PPPM) algorithm. The rice classification integrated vegetation indices spanning structural and metabolic signal (NDVI, EVI, LSWI) with China's regional cropping calendar and biophysical flooding signatures unique to rice transplanting phases, resolving the spectral overlap with other aquatic crops that confounds standard rangeland-style classification. Validated against provincial agricultural statistics from the National Bureau of Statistics of China, the autonomous classification recovered the documented six-fold expansion of YRD urban extent across the study window (8,297 km² in 2000 to 49,725 km² in 2018) and matched provincial sown-area-of-rice records within close proximity at large scale.
- Quantified atmospheric methane patterns across the YRD using Sentinel-5P TROPOMI column-averaged dry-air mixing ratios of CH_4 (XCH_4), the satellite product whose 2017 launch first enabled high-resolution global atmospheric methane monitoring at 7 km² spatial resolution. Applied the Getis-Ord Gi* statistic to identify statistically significant spatial clusters of high and low XCH_4 concentrations and Global Moran's I to quantify spatial autocorrelation across the provincial landscape, yielding a Moran's Index of 0.46 with z-score of 276.31 ($p < 0.01$) indicating strong positive clustering structure beyond random distribution. Cross-referenced the classified emission hotspots against the GAIA-derived urban extent and PPPM-derived rice distributions, identifying paddied rice fields in northern Anhui, northern Jiangsu, and northeast Zhejiang as the dominant contributors to elevated XCH_4 concentrations, with secondary urban-driven signatures emerging in central Shanghai and southwest Zhejiang where natural gas vehicle (NGV) retrofitting infrastructure has been concentrated. The integration unified two emission source categories that the YRD literature had historically treated as separate, surfacing the spatial co-location of agricultural and urban methane drivers within a single analytical frame.
- Developed a semantic-segmentation convolutional neural network on the DeepLabV3+ architecture to hindcast XCH_4 emission distributions across the multi-decadal pre-satellite window (2000 and 2010), addressing the fundamental temporal gap in atmospheric methane research that orbital monitoring only began in 2017 while the land-use transformations driving emissions span decades prior. The architecture leveraged atrous convolution with spatial pyramid pooling (ASPP) for multi-scale contextual capture and an encoder-decoder structure with skip connections built on a ResNet50 backbone, with the model trained on 5,000 image croppings spanning three YRD provinces from the 2018 year where all three signals (urban extent, rice extent, XCH_4) were simultaneously observed, and validated against 1,000 held-out croppings from the fourth province. The trained model then projected historical XCH_4 patterns from 2000 and 2010 urban and rice distributions, demonstrating that segmentation-trained CNNs can reconstruct atmospheric emissions records from land-use proxy distributions where direct satellite observation is unavailable. This work formed the basis of the M.E.Sc. thesis (advised by Dr. Xuhui Lee), establishing methodological novelty as the first integration of urban-rice coupling with deep learning to reconstruct multi-decadal atmospheric methane patterns across the YRD.

University of Minnesota, Marcell Experimental Forest *May – Aug 2022*

Visiting Research Scientist, Dr. Timothy Griffis Lab. Minneapolis, MN.

- Conducted field and computational research at the Marcell Experimental Forest (MEF), one of the world's longest-running peatland monitoring programs (established 1962) and the site of the first year-round measurements of peatland methane emissions, where the eddy covariance technique was first applied to peatland net carbon exchange. The research focused on the Bog Lake Peatland (AmeriFlux site US-MBP), a sub-boreal peatland with continuous CH_4 and CO_2 flux records since 2007, supporting field instrumentation across three micrometeorological towers measuring CH_4 vertical concentration gradients at 1.5m and 3.5m intervals above the lake surface. Eddy diffusivity was derived from 3D sonic anemometer-thermometer data sampled at 10 Hz, with atmospheric CH_4, CO_2, and water vapor concentrations measured near-simultaneously via Ultraportable GHG Analyzers through a custom-designed manifold at 406 SLPM, calibrated every six hours against zero-air and standard calibration gases under AmeriFlux protocols. Beyond eddy covariance, the program integrated in-lake biogeochemical instrumentation including thermistor chains, water quality sondes, and a buoy-mounted sensor array (pH, alkalinity, dissolved inorganic carbon, dissolved oxygen, chlorophyll, high-frequency CO_2) to capture coupled lake-atmosphere carbon dynamics relevant to peatland CH_4 flux.

- Assembled the multi-decadal monitoring record (2009-2021) into a structured forecasting framework targeting CH_4 flux forecasting from environmental drivers. Predictor variables (soil temperature at 10 cm, atmospheric temperature, CO_2 flux, and precipitation) emerged from a correlation analysis revealing soil temperature as the strongest signal (Pearson r = 0.87 with CH_4 flux) followed by atmospheric temperature (r = 0.76) and CO_2 flux (r = -0.74). An ordinary least squares multivariate regression yielded $R^2 = 0.78$ with F-statistic of 95.7 (statistically significant across all predictors), while standardized residual diagnostics confirmed near-normal distribution, Durbin-Watson testing revealed positive autocorrelation (1.2) consistent with seasonal time-series structure, and Augmented Dickey-Fuller testing showed CH_4 flux to be non-stationary while CO_2 flux exhibited stationarity: diagnostic asymmetry that motivated the decomposition strategy and ensemble approach.
- Constructed an ensemble layer-stacking architecture for CH_4 flux forecasting that integrated LightGBM, SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous factors), Facebook Prophet, and XGBoost alongside OLS regression, with each component exploiting a different architectural strength: SARIMAX for seasonal autocorrelation structure, Prophet for trend changepoint detection with yearly seasonality and 95% confidence intervals, gradient boosting models (LightGBM and XGBoost) for nonlinear multivariate interactions, and OLS for baseline interpretability. Time-series decomposition partitioned the signal into trend, seasonality, and residual components feeding the component models in the ensemble, with the top-performing component models selected by lowest Mean Absolute Percentage Error (MAPE) and combined into a weighted layer-stacked ensemble that extended CH_4 flux forecasts to a 52-week horizon. Cross-comparison of univariate against multivariate variants compared the multi-predictor architecture incorporating soil temperature, atmospheric temperature, precipitation, and CO_2 flux against univariate baselines, with XGBoost emerging as the strongest single model under MAPE minimization for both CH_4 and CO_2 flux prediction.

HONORS, AWARDS & FELLOWSHIPS

- **Yale Merit Research Scholar / Hixon Center for Urban Sustainability Fellow**: fully funded M.E.Sc., Yale School of the Environment (2021–2023).
- **Honorable Mention, Global Stocktake Climate Datathon**: With Abhinav Bhardwaj (Yale Mathematics, PhD), built *Carbon Trading: A Cross-Sectoral Analysis of Allowance Pathways Under the EU ETS*, an entry to the inaugural datathon convened by the Data-Driven EnviroLab (UNC-Chapel Hill) and the OpenEarth Foundation ahead of UNFCCC COP 27 in support of the Paris Agreement's Global Stocktake. Drawing verified compliance emissions and free-allocation records from the European Environment Agency's EU ETS Data Viewer across EU Member States and EEA-EFTA participants, the tool regressed sectoral emissions against allocated allowances over the closed third trading phase (2013–2020) to recover the empirical relationship between cap design and compliance behavior, projecting those relationships forward into the open fourth trading phase (2021–2030) under user-parameterized linear reduction factor (LRF) decay rates – the annual cap-contraction mechanism that drives the EU ETS – across the three sectors that anchor the system: power and heat generation, energy-intensive industry, and intra-EEA aviation, together accounting for roughly 41% of EU greenhouse gas emissions. The full pipeline surfaced through an interactive Dash application where policy users vary the LRF schedule and observe the resulting sectoral pathways resolve in real time. Selected by a panel of expert judges as one of four recognized entries from a field of more than 100 participants spanning 30 countries and six continents, with our two-person graduate team awarded alongside startup-founded teams and a multi-institutional international team. Submission archived at github.com/matt-duyst/Carbon-Trading.
- **UCLA Regents Scholar**: full academic merit recognition, University of California, Los Angeles (2014–2018).

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 & Statistics: GeoPandas, Rasterio, GDAL, NumPy, Pandas, Statsmodels, Folium, Plotly Dash, Matplotlib, spatial autocorrelation (Getis-Ord Gi*, Global Moran's I), eddy covariance, GHG flux modeling.