Chesapeake Bay Program
Science. Restoration. Partnership.

Integrated Trends Analysis Team (ITAT)

Workgroup Meeting

Wednesday, May 20th, 2025
10:00 – 12:00 PM

Visit the meeting webpage for meeting materials and additional information.

This meeting was recorded for internal use only to assure the accuracy of meeting notes.

Purpose: This meeting aims to explores how artificial intelligence (AI) and machine learning (ML) are being applied across the Chesapeake Bay watershed to better understand long-term patterns in water quality, ecosystem health, nutrient pollution, and hydrodynamic processes. Presentations will highlight emerging AI/ML tools and findings, and we will discuss key challenges such as model transparency, explainability, stakeholder trust, and translating scientific insights into practical decision-making tools.

Minutes

I. Welcome, Introductions & Announcements (10:00 - 10:05 AM)

Lead: ***Breck Sullivan*** *(U.S. Geological Survey, USGS) ITAT Co-coordinator, and* ***Kaylyn Gootman*** *(U.S. Environmental Protection Agency, EPA) ITAT Co-coordinator.*

- The DataHub will be updated on June 8th. Email eyoung@chesapeakebay.net to receive a summary of the updates, and/or to be added to a DataHub users mailing list.

Upcoming Conferences, Meetings, Workshops and Webinars

- Chesapeake Community Research Symposium – June 1-3, 2026. Annapolis, Maryland. ***Registration is now open here!***
- Restore America's Estuaries' 2026 Coastal & Estuarine Summit – September 22-25, 2026. San Francisco, California.

II. AI-Driven Insights into Shifting Water, Sediment, Nutrient, and Salt Dynamics in the Chesapeake Bay and Beyond (10:05 – 10:30 AM)

Lead: **Admin Husic** (Virginia Tech, VT).

Wei Zhi's paper *From Hydrometeorology to River Water Quality: Can a Deep Learning Model Predict Dissolved Oxygen at the Continental Scale?* (2021; link) was pivotal in understanding the application of machine learning (ML) to water quality science. The authors set out to answer a fundamental question: can neural network models accurately predict complex water quality variables at large scales? By coupling continental-scale hydrometeorologic data with in-stream observational records, they utilized a Long Short-Term Memory (LSTM) recurrent neural network to predict river dynamics. The model has an exceptional ability to fill data gaps, hindcast historical periods before observational data existed, and forecast future trends. This demonstrated that while traditional process-based and statistical models have known weaknesses, deep learning can effectively overcome data scarcity and aid in testing new scientific hypotheses.

While hydrology has historically led the way in adopting ML, water quality science typically lags a few years behind, primarily due to a severe lack of data volume. Hydrologists enjoy extensive global networks with decades of continuous discharge records, whereas water quality data barely registers on the same scale, often covering a meager 1-5% of the year. This data scarcity is a massive hurdle because traditional grab samples are highly sparse, meaning we routinely miss the critical, non-linear, and short-lived transport bursts of sediment and nutrients during storm events. To bridge this gap, our group relies heavily on high-frequency water quality sensors. These instruments function as our definitive bridge to big data, capturing rapid storm responses and providing neural networks with the dense patterns they need to learn effectively across thousands of diverse watersheds simultaneously.

Our research group has actively leveraged these sensors and deep learning to model water quality variables nationwide, starting with a study on stream nitrate. By training a single neural network on flow data from about 100 USGS nitrate sensors, our model successfully simulated entirely opposing watershed behaviors: the agricultural flushing of nitrate during storms in the Mississippi River Basin, and the storm-driven dilution of urban point-source nitrate in the Chesapeake Bay's Difficult Run. Expanding on this framework, a current student is modeling salinity using data from 350 specific conductance sensors across the country. Using the Kling-Gupta Efficiency (KGE) metric, our model achieved strong results, performing best in pristine, forested basins where atmospheric deposition dominates, and showing intermediate performance in complex urban basins where unmapped, human-driven salinity sources complicate predictions.

These deep learning approaches were also applied to sediment transport to investigate an old adage: that 90% of sediment is transported in just 10% of the time. Because sparse grab samples easily miss the massive, single-day storm events that carry the bulk of annual sediment, we utilized high-frequency sensor data to calculate a metric we call "burstiness." Our model successfully captured vast regional differences, revealing low-burstiness watersheds that took 126 days to export 90% of their annual sediment load, compared to highly volatile, high-burstiness systems in Pennsylvania that exported that same 90% fraction in a staggering six days. Our model accurately recreated these step-function changes across 175 sites.

The final core application of our ML framework focuses on identifying hydrologic pathways and more specifically, determining how much streamflow is composed of "event water" (recent precipitation) versus pre-existing catchment storage. While traditional hydrograph separation using specific conductance as a tracer works well, it is often limited by data gaps or violated underlying assumptions. To circumvent this, we trained a deep learning model to learn event water generation mechanisms when tracer data is present, allowing it to seamlessly predict pathway dynamics during data gaps and into unmonitored periods. The model has proven highly adaptable, accurately capturing the flashy, event-water-dominated runoff of heavily paved urban systems as well as the heavily attenuated, storage-dominated baseflow of forested watersheds.

We utilize the high-fidelity models to reconstruct 30-year historical records across the Conterminous United States (CONUS) to see exactly how these systems are shifting under climate and land-use stressors. Our salinity reconstruction revealed prominent regional increases, particularly in the Northeast and the Chesapeake Bay watershed. More alarmingly, our sediment models show that annual sediment flux is increasing at roughly 25-27% of our study sites, and that this flux is becoming significantly more "bursty" over time. Similarly, we found that about 30% of our 750 hydrologic sites show rising event water fractions, indicating that watersheds are converting precipitation into direct runoff at a much higher rate than in decades past.

When we looked closer at the Chesapeake Bay watershed, we discovered that urban landscapes are the most sensitive to these changes, though even forested areas are experiencing rising event water fractions. Using explainable AI, counterfactual Shapley value analysis, we determined that landscape modifications like urbanization and deforestation drive 70% of the increases in sediment burstiness, while intensifying precipitation regimes are the dominant driver behind shifting hydrologic pathways. Furthermore, by modeling sediment fingerprinting literature globally, we mapped sediment provenance in the Chesapeake Bay, revealing that basin-wide sediment originates primarily from subsurface bank erosion (63%), followed by cropland (20%), and that highly urbanized zones act as incredibly efficient conveyors of this material. Ultimately, these findings show that as climate and land-use pressures compress sediment transport into smaller windows of time, our management options shrink; as we move forward, we must continue to critically evaluate whether our ML models are offering truly causal, novel insights or simply confirming what we already know.

<u>Discussion:</u>
Q: *Robert Sabo:* What is neat about these models is that we are able to extract model insights, inform our management, and then create a natural experiment where we can implement those changes. I am just wondering, how can we move from correlative insights into causative insights? Could this management dynamic properly inform that process in terms of developing natural experiments?

- ***A:*** *Admin Husic:* While highly accurate machine learning models are initially impressive, their truer value lies in their ability to uncover new insights where rigid, process-based models fail due to flawed or overconfident physical assumptions. For practical land management, a powerful data-driven prediction can be just as effective as a traditional physical model if it enables proactive action during critical environmental events. Bridging this gap, the emerging fusion of both approaches through differentiable modeling offers an exciting framework that maintains strong predictive power while addressing structural physical concerns.

- ***Comment from chat:*** *Kelly Maloney:* to assign causality we here are looking at causal inference techniques some of which involve AI/ML.

Q from chat: *Qian Zhang:* A related question: Could you offer some general comments on AI/ML model trustworthiness and explainability based on your experiences?

- ***A:** Admin Husic:* To build trustworthy and easily interpretable ML models, you should keep them as light as possible by limiting them to a few high-quality, fundamentally distinct predictors. Because model explanations shift depending on the features included, domain expertise and professional judgment are far more critical than raw code optimization or extra data inputs for properly teaching and configuring the model. Ultimately, success relies on leaning on your expert opinion to ensure the selected variables genuinely reflect the reality of the system you are trying to predict.

III. Using machine learning to examine long-term trends in stream biological condition (10:30 – 11:00 AM)

Lead: **Kelly Maloney** (USGS).

To understand long-term trends in stream biological health, my collaborative research team leverages ML to track how macroinvertebrate communities respond to a rapidly changing landscape. This is a complex undertaking across the Chesapeake Bay watershed because we are constantly battling land-use legacies, i.e. enduring historical impacts from turn-of-the-century logging and mining that continuously dictate modern ecological structures. When we layer recent disruptions like unconventional oil and gas development on top of decades of conservation interventions, establishing an accurate environmental baseline becomes incredibly difficult. Yet, having a clear baseline is absolutely essential if we want to measure our progress toward regional management targets, such as achieving a 3% improvement in stream conditions every six years.

To overcome these data limitations, we heavily upgraded our predictive framework by compiling over 7,000 high-resolution environmental summaries. After rigorous quality control, we narrowed these down to 89 highly optimized, time-varying climate and annual land-cover predictors. We trained our random forest (RF) algorithms to isolate and predict four core biological responses: the overall Chessie BIBI score (Basin-wide Index of Biotic Integrity), pollution-sensitive EPT taxa (Ephemeroptera, Plecoptera and Trichoptera), salinity-sensitive mayflies, and sedimentation-sensitive clinger organisms. In ecological modeling, hitting an R^2 value above 0.50 or 0.60 is a major milestone, and our diagnostics comfortably cleared these thresholds. This gave us the confidence to generate robust, high-fidelity annual predictions for roughly 361,000 small, non-tidal streams across the entire Chesapeake Bay watershed spanning from 1985 to 2023.

When we mapped these results across all 361,000 streams, clear geographic and environmental degradation patterns emerged. Our models confirmed that upstream impervious surface cover exerts a devastating negative impact on biology, with severe ecological decline radiating directly across major metropolitan areas like the Baltimore-Washington corridor and Richmond due to rapid urban sprawl. Conversely, steep, protected forest headwaters in the northern and western zones consistently mapped as high-quality habitats. Most notably, we detected widespread, sweeping declines in sensitive mayfly populations across the entire basin. This distinct biological signature points directly to regional freshwater salinization, strongly validating independent physical models showing that two-thirds of the Bay's stream reaches are actively salinizing.

To make these massive datasets actionable for managers, we binned our raw stream scores into standardized health categories. Our annual predictions initially exhibited immense year-to-year volatility due to climate noise, but implementing a two-year moving average successfully smoothed out the data. This revealed a real, measurable net improvement of roughly 0.9% in acceptable stream miles since 2008. While a 0.9% regional improvement falls short of idealistic 10% historical management goals, it represents a profound victory for conservation. Over this exact same 39-year window, developed land expanded by 3.1%; the fact that stream biology improved at all proves that our targeted preservation and landscape management efforts are actively buffering urban expansion.

Looking ahead, we are refining our models by integrating upcoming one-meter land-cover sets and high-resolution stream canopy data to eliminate remaining training biases. However, we firmly believe that holistic stream health cannot be measured through a single biological indicator alone. This project is just one piece of a massive, interdisciplinary USGS initiative I am co-leading to build parallel ML models for fish communities, sediment yields, flow alterations, stream temperatures, and toxic contaminants. Our goal is to synthesize all of these moving parts into highly visual, web-based decision-support portals, transforming 40 years of complex environmental data into directly actionable intelligence for resource managers across the country.

Discussion:

Q: *Robert Sabo:* what are some of your ways that you communicate uncertainty while still maintaining confidence in the model? I'm just wondering how stakeholders have reacted when they see an R^2 of 0.4 or 0.5, but then you walk them through it. Are they still open to it? Do they see that you're capturing the signal and that there's just a lot of noise because of a variety of factors?

- ***A:*** *Kelly Maloney:* When working with skeptical stakeholders who favor field measurements, it helps to point out that observed data is simply its own limited point-in-time model plagued by its own compounded errors. Because ML predictions are inherently bound by these combined data and predictor errors, shifting from raw numerical targets to categorical classifications can make the results much more reliable. Ultimately, stakeholders find greater comfort and trust in the system when the model focuses on binary or tiered categories, where it achieves high efficiency at successfully separating the best and worst environmental conditions.

Q from chat: *Qian Zhang:* What happens in the western Pennsylvania (PA) region that got assigned “poor” by the model? Could you share a little bit on your efforts in making the model results accessible to managers (your web tool)?

- ***A:*** *Claire Buchanan:* It's very clear that the tree regrowth in the western part of the basin is important. Kelly and I really haven't looked deeply into the idea of how the relationship with the streams changes as a forest regrows, but I do think a lot of that forest regrowth is important. However, in Western PA, you also have gas extraction and you have the salinity effects of that going on. One thing we're noticing is that the importance of water quality is sort of underestimated in these discussions. The biological response to stream habitat and the response to land use actually is very significantly affected by the water quality in the stream itself. Kelly and I are aiming to look at specific sites where the model and the monitoring data predict very different results for the Chessie BIBI, and we will try to tease out what those causes are.

- ***A from chat:*** *Kelly Maloney:* On the tool - we are having calls with key users (e.g., NFWF) to build the web tool in a collaborative fashion that will be "easy" to use and useful. Also, what indicators are important to them. We have a team dedicated to the development of this platform.

IV. Net declines in nonpoint source pollution into one of the world's largest estuaries (11:00 – 11:30 AM)

Lead: **Robert Sabo** (EPA, Office of Water).

To identify the long-term drivers of total nitrogen (TN) loading trends throughout the Chesapeake Bay, our research team utilized a neural network integrated with the CAST-based (Chesapeake Assessment Scenario Tool) Chesapeake Bay Nutrient Inventory. Published in 2022, this inventory holistically links diverse nutrient sources and riverine discharge data to quantify nitrogen exports across 121 monitoring sites. The overarching modeling results deliver a mix of promising historical milestones alongside clear future obstacles. Specifically, the data indicates that farm-level nutrient management, sweeping reductions in air pollution, and targeted wastewater treatment upgrades have significantly improved regional water quality, though accelerating urbanization simultaneously poses a unique counter-stressor to these systemic gains.

A cornerstone finding of our underlying nutrient inventory is that agricultural nitrogen surpluses have largely declined since the mid-1980s. While immense volumes of nitrogen are still left on fields after accounting for manure, fertilizer inputs, and crop removal, farmers are leaving progressively less nitrogen over time. This trend is heavily driven by enhanced nutrient use efficiency, allowing counties to maintain or even expand livestock and crop production while reducing land-based nitrogen surpluses by 20-30%. However, this agricultural progress is not entirely uniform across the basin; distinct agricultural sub-watersheds, such as the Choptank River, continue to show expanding nutrient surpluses that heavily complicate local water quality restoration efforts.

Beyond the agricultural sector, the nutrient inventory highlights universal progress in reducing atmospheric and point-source pollution. Driven by a succession of Clean Air Act amendments targeting nitrogen oxide emission reductions since 1990, atmospheric nitrogen deposition has experienced a widespread and substantial decline across the entire Chesapeake Bay watershed. Concurrently, major metropolitan centers like Baltimore, Washington D.C., and Richmond have successfully slashed their point-source discharges through aggressive wastewater facility infrastructure upgrades. It is important to note, however, that because many of these major municipal facilities discharge directly into tidal waters downstream of our non-tidal network monitoring stations, their direct impacts are omitted from our non-tidal loading trends and instead primarily benefit the Bay's tidal estuaries.

To robustly link these shifting pollution sources to long-term nitrogen export trends, we trained a neural network using just six streamlined spatial-temporal predictors: river discharge, percent forest cover, agricultural surplus, urban inputs, agricultural fertilizer, and atmospheric deposition. To enforce strict model stability and eliminate mathematical noise, we deployed a bagged ensemble approach to train 100 unique neural network models. The global model achieved exceptional predictive diagnostics, comfortably clearing an R^2 threshold of 0.90 across both our training and independent validation datasets. Furthermore, the framework successfully passed empirical tests, reliably pairing higher annual discharge with elevated nitrogen loads, while matching increased forest coverage with significantly depressed nitrogen yields.

With a high-fidelity model established, we executed a counterfactual analysis to isolate the environmental impacts of individual pollution sources by systematically holding them at 1995 levels. Evaluating a major artery like the Susquehanna River at Marietta revealed that actual median nitrogen loads dropped from roughly 6.8 kilograms per hectare (kg/ha) in 1995 down to 5.0 kg/ha by 2020. Our counterfactual simulations demonstrated that if agricultural surpluses had remained frozen at peak 1995 levels, the Susquehanna would have seen virtually zero water quality improvement. This concrete evidence proves that optimized agricultural nutrient management was absolutely instrumental to the river's recovery, working in tandem with the parallel benefits of reduced atmospheric deposition and wastewater mitigation.

Scaling this counterfactual analysis across all 121 monitoring stations provides compelling, basin-wide evidence that non-point source management has fundamentally reshaped nitrogen export dynamics. Across the vast majority of watersheds, declining atmospheric deposition and optimized agricultural practices emerged as the dual primary drivers of long-term reductions in nitrogen export. Conversely, the models show that population growth has caused urbanization to act as a universal degradation vector that actively offsets conservation gains. Ultimately, while historical source reductions have successfully moved the needle on water quality, accelerating further reductions in agricultural surpluses remains absolutely paramount for achieving the restoration goals of the CBP states.

Discussion:

Q from chat: *Kelly Maloney:* Robert - wow R^2 of 0.9, I am so envious. Would it be possible to convert loads to concentrations to enable better linkage to non-tidal biological endpoints? For counter factual do you have any sites that underwent a large land use change over the period to test what ifs?

- ***A:*** *Robert Sabo:* One thing I want to point out is if we use a discharge-only model or an inventory-only neural network model, the model performance is terrible. Like ~0.3-0.4. So, combining the inventory with the discharge information is instrumental for having a good performance. In terms of converting to a concentration, that's something we could definitely do. I'm project lead for EPA's National Nutrient Inventory. We're also developing some pretty sophisticated LSTM and other neural network models to explain the spatial and temporal variation of nitrogen species and phosphorus species across CONUS. And we can also zoom in on the Chesapeake. So that's something of importance as well.

V. Applications of ML for Hydrodynamics and Water quality simulations in Chesapeake Bay **(11:30 – 12:00 PM)**

Lead: **Jian Shen** (Virginia Institute of Marine Science, VIMS).

As a numerical and hydrodynamic modeler leading an estuary and coastal modeling group, my perspective is deeply rooted in utilizing three-dimensional systems to understand complex aquatic environments. Our group, specifically through work spearheaded by Joseph Zhang on next-generation 3D platforms, has historically relied on computationally rich numerical models to track storm surges, salinity intrusion, waves, and water quality parameters like primary production and dissolved oxygen (DO). While these advanced physical models are incredibly reliable and essential for understanding underlying mechanisms, running them is computationally expensive and logistically challenging. Reconstructing decades of historical trends or projecting long-term future climates requires massive, high-resolution forcing datasets and open boundary conditions that are often unavailable or costly to generate.

To bridge this operational gap, we are actively transitioning from simulating isolated monitoring points to modeling entire dynamic fields by training ML algorithms on our 3D numerical model outputs. This hybrid approach allows us to create highly efficient, cost-effective digital representations that reproduce complex spatial-temporal structures, such as hourly tidal changes, using only basic wind, flow, and boundary inputs. Furthermore, ML models break the computational bottleneck inherent to traditional 3D architectures. While it is virtually impossible to run hundreds of iterations of a heavy numerical model to assess system behavior, a trained ML model can rapidly generate large-scale ensemble simulations. This computational speed enables us to systematically quantify parameter uncertainties and construct diverse scenario studies that are incredibly valuable for environmental managers.

Our application of this workflow to wave modeling across the entire Chesapeake Bay illustrates the power of simplifying highly complex spatial forecasting. Predicting wave dynamics across the Bay's intricate grid traditionally demands immense computational power, but by training a relatively simple LSTM network on our 3D model outputs, we built a highly portable and explainable forecasting tool. We validated this model using independent satellite records and raw NOAA observational data, proving it could accurately capture major storm events and fine-scale spatial resolution. Because this ML tool only requires nine wind datasets across the Bay to function, it simplifies operational forecasting and allows us to efficiently hindcast historical wave climates back through time.

Similarly, we have successfully replicated 3D numerical simulations of salinity intrusion across major East Coast tributaries using ML. Due to a lack of long-term empirical salinity records in the Chesapeake Bay, we leveraged abundant data from the Delaware Bay to rigorously verify our trained model against field observations, confirming its high reliability. This predictive tool can ingest the previous 90 days of river flow to forecast salinity intrusion 7-14 days into the future. It also allows us to reconstruct historical salinity trends stretching back to 1980, revealing fascinating, season-specific deviations in long-term behaviors. When tested against our heavy 3D numerical models under artificial high- and low-flow scenarios, the ML model responded with excellent accuracy, allowing us to run 50-100 ensemble variations in a fraction of the time to map management uncertainties.

When translating these ML workflows to water quality parameters like DO and hypoxia volume, we encountered a fundamental modeling trap: the dominating influence of water temperature. If raw temperature data is directly fed into a neural network, the model attributes 80% to 90% of the statistical variance to temperature alone, rendering it entirely insensitive to nutrient fluctuations and useless for environmental management. To bypass this, we avoid directly inputting raw temperature, opting instead to isolate space and time using Singular Value Decomposition (SVD) and Empirical Orthogonal Functions (EOF) to preserve vertical stratification. We also mathematically transformed input variables into nutrient-limiting functions and accounted for transport delays and phase shifts between the upper, middle, and lower bays, ensuring the model accurately shrinks or expands hypoxia volumes in direct response to altered nutrient loads.

We observed similar success and unique advantages when applying ML to simulate highly erratic phytoplankton blooms. While traditional 3D ecological models struggle to simulate sudden, harmful algal blooms without explicitly programming code for numerous individual species, ML models simply bypass this biological complexity by identifying hidden patterns directly within the data. In tandem with these efforts, we are harnessing high-resolution National Oceanic and Atmospheric Administration (NOAA) satellite data to enhance our spatial training grids. Although satellite imagery contains inherent measurement errors and frequent data gaps due to cloud cover, we successfully mitigated these issues by developing a seven-day averaging methodology to synthesize data patches. This enabled us to combine satellite and field observations into robust datasets capable of training models to track complex chlorophyll anomalies.

Ultimately, our research underscores that successfully applying ML to water quality management requires much more than just throwing massive datasets into a neural network. In the James River, for instance, we benchmarked our ML models against a highly accurate 3D numerical model used by the Department of Environmental Quality (DEQ) to evaluate 30-50% nutrient reduction strategies. By intentionally embedding physical and biological processes like nutrient-limiting factors, transport times, and seasonal lags, into the ML inputs, we ensured the AI natively respects the actual mechanics of the estuary. This deliberate integration of domain expertise transforms ML from a simple statistical forecasting trick into a highly responsive, physically sound management tool, bringing us closer to our ultimate goal of establishing a functional digital twin for the estuary.

VI. Adjourn **(12:00 PM)**

Next meeting: June 24th from 10 AM – 12 PM.

Attendees:

- Kelly Maloney (USGS)
- Robert Sabo (EPA)
- Admin Husic (VT)
- Jian Shen (VIMS)
- Breck Sullivan (USGS)
- Jon Harcum (Tetra Tech)
- Kaylyn Gootman (EPA)
- Elgin Perry (Consultant)
- Qian Zhang (UMCES-CBPO)
- Gabriel Duran (CRC-CBPO)
- Lewis Linker (EPA)
- Richard Tian (UMCES-CBPO)
- Emily Young (ICPRB-CBPO)
- Claire Buchanan (ICPRB)
- Klaus Huebert (MDDNR)
- Cass Klingaman (DEC)
- Michael Lane (ODU)
- Efeturi Oghenekaro (DOEE)
- George Onyullo (DOEE)
- Rebecca Murphy (UMCES-CBPO)
- Rikke Jepsen (ICPRB)
- Cynthia Johnson (VADEQ)
- David Parrish (VIMS)
- Carl Friedrichs (VIMS)