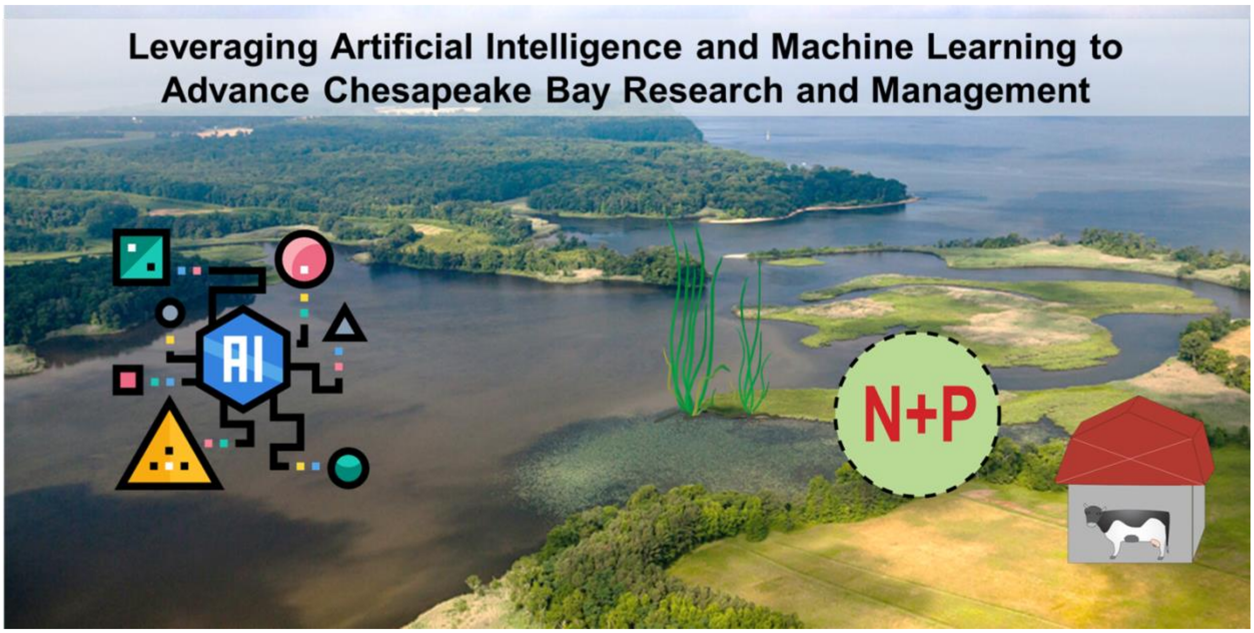
Leveraging Artificial Intelligence and Machine learning to Advance Chesapeake Bay Research and Management: A Review of Status, Challenges, and Opportunities



STAC Workshop Review
February 24-25, 2025
Edgewater, MD

STAC Publication 25-005

About the Scientific and Technical Advisory Committee

The Scientific and Technical Advisory Committee (STAC) provides scientific and technical guidance to the Chesapeake Bay Program (CBP) on measures to restore and protect the Chesapeake Bay. Since its creation in December 1984, STAC has worked to enhance scientific communication and outreach throughout the Chesapeake Bay Watershed and beyond. STAC provides scientific and technical advice in various ways, including (1) technical reports and papers, (2) discussion groups, (3) assistance in organizing merit reviews of CBP programs and projects, (4) technical workshops, and (5) interaction between STAC members and the CBP. Through professional and academic contacts and organizational networks of its members, STAC ensures close cooperation among and between the various research institutions and management agencies represented in the Watershed. For additional information about STAC, please visit the STAC website at www.chesapeake.org/stac.

Publication Date: October 2025 *(revised November 2025)*

Publication Number: 25-005

Cover graphic: Chesapeake Bay Program, Integration & Application Network, Freepik

Suggested Citation:

Zhang, Q., M. Baker, I. Bertani, B. Dennison, L. Linker, K. Maloney, R. Sabo, C. Shen, G. Shenk, K. Van Meter, and M. Cole. 2025. Leveraging Artificial Intelligence and Machine Learning to Advance Chesapeake Bay Research and Management: A Review of Status, Challenges, and Opportunities. STAC Publication Number 25-005, Edgewater, MD. 45 pp.

This project has been funded wholly or in part by the United States Environmental Protection Agency under assistance agreement CB-95304801 to the Chesapeake Research Consortium. The contents of this document do not necessarily reflect the views and policies of the Environmental Protection Agency, nor does the Environmental Protection Agency endorse trade names or recommend the use of commercial products mentioned in this document, as well as any images, video, text, or other content created by generative artificial intelligence tools, nor does any such content necessarily reflect the views and policies of the Environmental Protection Agency.

STAC Administrative Support Provided by:

Chesapeake Research Consortium, Inc.
645 Contees Wharf Road
Edgewater, MD 21037
Telephone: 410-798-1283; 301-261-4500
Fax: 410-798-0816
http://www.chesapeake.org

Workshop Steering Committee:

Qian Zhang, Workshop Chair, University of Maryland Center for Environmental Science (UMCES)
Matt Baker*, University Of Maryland, Baltimore County (UMBC)
Isabella Bertani, University of Maryland Center for Environmental Science (UMCES)
Bill Dennison*, University of Maryland Center for Environmental Science (UMCES)
Lew Linker, U.S. Environmental Protection Agency (EPA)
Kelly Maloney, U.S. Geological Survey (USGS)
Robert Sabo, U.S. Environmental Protection Agency (EPA)
Chaopeng Shen, The Pennsylvania State University (PSU)
Gary Shenk, U.S. Geological Survey (USGS)
Kim Van Meter, The Pennsylvania State University (PSU)
**STAC member*

STAC Staff:

Meg Cole, STAC Coordinator, Chesapeake Research Consortium
Tou Matthews, STAC Projects Manager, Chesapeake Research Consortium

Table of Contents

Executive Summary

The Chesapeake Bay and its watershed (hereafter "Chesapeake Bay region") have been the focus of extensive restoration efforts for several decades. These restoration efforts are guided by the Chesapeake Bay Watershed Agreement (Chesapeake Executive Council 2014) which outlines 10 goals and 31 measurable outcomes. The Chesapeake Bay is globally recognized as a model for coastal restoration due to long-term investments in monitoring, modeling, implementation and research by the Chesapeake Bay Program (CBP) partnership. These monitoring network spans tidal and non-tidal regions and provides data across multiple scales. Artificial intelligence (AI), particularly machine-learning (ML) and deep learning (DL), has emerged as a powerful tool for analyzing large, complex datasets. These techniques have gained widespread adoption across various disciplines, including ecology, hydrology, and environmental science. In the Bay context, AI/ML is increasingly being used to explore drivers of environmental change, analyze system dynamics, and predict conditions in areas with limited monitoring.

The CBP partnership, particularly its Scientific and Technical Advisory Committee (STAC), has increasingly recognized the growing role of AI/ML in watershed and estuarine management. Recent Chesapeake Community Research Symposium sessions and initiatives such as the Chesapeake Global Collaboratory highlight increasing regional momentum to apply big data and AI/ML for environmental solutions. Together, these developments underscore the timely need to explore how AI/ML can help advance Chesapeake Bay restoration and management.

This STAC workshop, titled "Leveraging Artificial Intelligence and Machine learning to Advance Chesapeake Bay Research and Management: A review of status, challenges, and opportunities," was held from February 24-25, 2025, in Edgewater, Maryland to bring together over 50 federal, state, and academic scientists and partners to synthesize the current state of AI/ML applications and identify research gaps in Chesapeake Bay research and management. The workshop focused on three main objectives:

1. Summarize recent AI/ML applications and lessons learned in both tidal and non-tidal areas of the Chesapeake Bay region.
2. Identify challenges and gaps in applying AI/ML approaches to Chesapeake Bay data. Such challenges and gaps may include data limitations, harmonization issues, ineffective communication of AI/ML insights, and a lack of coordination among research and management institutions.
3. Develop recommendations and identify opportunities for leveraging AI/ML to address issues across the Chesapeake Bay region. Key areas of focus may include generating new information to support watershed management, delivering AI/ML-generated insights to managers in a clear and actionable way, and fostering greater collaboration among stakeholders within the CBP Partnership.

Workshop participants engaged in science presentations and breakout sessions to develop recommendations for advancing the integration of AI/ML techniques into research and management across the Chesapeake Bay region. By synthesizing current applications, identifying challenges, and exploring new opportunities, the workshop has provided valuable insights and recommendations for better leveraging AI/ML approaches to support the success of Bay restoration efforts. Together, these recommendations provide a roadmap for enhancing data-driven, science-based decision making aligned with the goals and outcomes of the Chesapeake Bay Watershed Agreement.

Recommendations

1. Strengthen data infrastructure and integration for AI/ML applications

- Harmonize spatial and temporal datasets across programs and ensure consistent metadata.
- Leverage diverse datasets, including satellite, in-situ, and high-frequency data, for use in modeling and monitoring applications and filling data gaps.
- Design monitoring and data processing efforts so that resulting products are problem-relevant and can be readily incorporated into AI/ML workflows.
- Build harmonized response and predictor datasets and develop example use cases to guide widespread AI/ML applications.

2. Leverage AI/ML for restoration of Chesapeake Bay tidal and non-tidal regions and decision support

- Use AI/ML to assess effectiveness and efficiency of restoration practices, evaluate progress, and identify drivers of environmental change.
- Enhance watershed and estuarine models by integrating AI/ML outputs and insights.
- Promote integration between AI/ML and traditional monitoring, analysis, and modeling approaches to enhance scientific credibility and transparency.
- Develop accessible AI-driven tools (e.g., Chesapeake-specific large language models) for scenario planning to help identify management priorities.

3. Promote transparency and engage managers and stakeholders

- Advance explainable AI/ML and uncertainty protocols so that results are interpretable, credible, and trusted.
- Couple AI/ML with tailored data visualizations to improve interpretability and use at broader scales.
- Foster engagement of managers and decision makers at all stages of AI/ML projects to ensure products align with management priorities and can be effectively applied.
- Use tailored communication strategies to translate AI/ML insights into actionable guidance for restoration planning.

4. Build collaboration and capacity

- Establish an AI/ML network (e.g., Chesapeake Bay Research with Artificial Intelligence and Networking or "Ches-BRAIN") to foster collaboration and to provide a clear place where managers and others can easily find and connect with AI/ML experts.
- Encourage participatory events such as hackathons to spark innovation and strengthen cross-sector collaboration.
- Invest in training and literacy programs so that scientists, managers, and decision makers can effectively use and interpret AI/ML tools and outputs.

Introduction

The Chesapeake Bay and its watershed (hereafter "Chesapeake Bay region") have been the focus of extensive restoration efforts for several decades. These restoration efforts are guided by the Chesapeake Bay Watershed Agreement (Chesapeake Executive Council 2014), which outlines 10 goals and 31 measurable outcomes. The Chesapeake Bay is globally recognized as a model for coastal restoration due to long-term investments in monitoring, modeling, and research by the Chesapeake Bay Program (CBP) partnership. These extensive monitoring data span both tidal and non-tidal regions of the Chesapeake Bay and cover various temporal and spatial scales. Such data provide valuable insights into ecosystem changes and help generate hypotheses about environmental drivers. However, such data are often complex and difficult to interpret. As a result, new approaches to extract and communicate meaningful patterns could advance scientific understanding and support ongoing restoration efforts.

Artificial intelligence (AI), particularly machine-learning (ML), has become a transformative tool in environmental research, especially for extracting elusive patterns from large, complex datasets that traditional analysis methods can fail to detect. These AI techniques have gained widespread adoption across various disciplines, including ecology, hydrology, and environmental science (Shen, 2018; Xu and Liang, 2021). In the context of the Chesapeake Bay region, AI/ML techniques have been increasingly applied to analyze complex dynamics, identify environmental drivers, and predict conditions in unmonitored areas. For example, recent research has used AI/ML to study chlorophyll *a* (Yu and Shen, 2021), dissolved oxygen (DO) (Yu et al. 2020), nutrient limitation (Zhang et al. 2022), water-quality standards (Zhang et al. 2025), and biological stream health (Maloney et al. 2022b).

The CBP partnership, particularly its Scientific and Technical Advisory Committee (STAC), has increasingly recognized the potential of using AI/ML for improving decision support towards better Bay watershed management. Researchers from institutions in the Bay watershed, such as the University of Maryland Center for Environmental Science (UMCES), Virginia Institute of Marine Science (VIMS), Pennsylvania State University (PSU), and Johns Hopkins University (JHU), have proposed AI/ML-focused sessions at the 2024 Chesapeake Community Research Symposium (Chesapeake Community Research Symposium 2024), signaling growing interest in these approaches. Furthermore, UMCES has recently launched the Chesapeake Global Collaboratory (UMCES 2023), a new initiative that aims to harness big data and AI/ML tools to accelerate the process of identifying cost effective, time efficient, and robust solutions for addressing complex environmental challenges. These developments underscore how the exploration of AI/ML capabilities can support ongoing restoration and management of the Bay and its watershed.

This STAC workshop brought together federal, state, and academic partners to synthesize the current state of AI/ML applications and identify gaps in Chesapeake Bay research and management. The workshop focused on three main objectives:

1. Synthesize recent AI/ML applications and lessons learned in both tidal and non-tidal areas of the Chesapeake Bay region.

2. Identify challenges and gaps in applying AI/ML approaches to Chesapeake Bay data. Such challenges and gaps may include data limitations, harmonization issues, ineffective communication of AI/ML insights, and a lack of coordination among research and management institutions.
3. Develop recommendations and identify opportunities for leveraging AI/ML approaches to address critical management issues and foster successful restoration across the Chesapeake Bay region. Key areas of focus may include generating new information to support watershed management, delivering AI/ML-generated insights to managers in a clear and actionable way, and fostering greater collaboration among stakeholders within the CBP partnership.

This STAC workshop serves as a critical step in advancing the integration of AI/ML techniques into research and management across the Chesapeake Bay region. By synthesizing current applications, identifying challenges, and exploring new opportunities, the workshop provided valuable insights and recommendations for better leveraging AI/ML to support restoration efforts in the Bay. These insights will help inform data-driven, science-based decision making aligned with the goals and outcomes of the Chesapeake Bay Watershed Agreement (Chesapeake Executive Council 2014).

Presentation Summaries

This section summarizes presentations at the workshop. Slides for the presentations are available on the STAC *Leveraging Artificial Intelligence and Machine-learning to Advance Chesapeake Bay Research and Management: A review of status, challenges, and opportunities* workshop webpage, accessible using the following link.

The workshop was organized into three main sessions: (1) recent AI/ML applications and lessons learned, (2) challenges and gaps in applying these approaches to Chesapeake Bay data, and (3) opportunities and recommendations for advancing the use of these approaches. Descriptions of each session, including presentation topics and speakers, are provided below.

Session I: Lessons Learned from AI/ML Applications in the Chesapeake Bay Watershed

This session synthesized recent applications of AI/ML to Chesapeake Bay research, spanning both tidal and non-tidal systems. Presenters described the objectives of their studies, the reasoning behind the choice of AI/ML methods, and the novel insights these approaches provided. Discussion also emphasized how these findings have already contributed (or could contribute) to advancing restoration of Chesapeake Bay tidal and non-tidal regions in support of the Watershed Agreement's goals and outcomes.
Invited presentations in this session included:

- Gary Shenk (USGS) – Overview of Chesapeake Bay Restoration: CBP Goals & Outcomes
- Alison Appling (USGS) – Introductory Overview of AI and ML
- Kelly Maloney (USGS) – Literature Summary of Watershed and Living Resources Studies Involving AI/ML
- Jian Shen (VIMS) – Literature Summary of Estuarine and Living Resources Studies Involving AI/ML
- Stephanie Schollaert Uz (NASA) – AI/ML Integration of Satellite Remote-sensing: Data Harmonization Challenges and Gaps

Overview of Chesapeake Bay Restoration: CBP Goals & Outcomes – Gary Shenk (USGS)

The CBP is a collaborative partnership among various stakeholders – including government agencies, environmental professionals, and scientists – aimed at restoring and maintaining the health of the Chesapeake Bay. It is guided by agreements between state and federal partners, including the most recent 2014 Chesapeake Bay Watershed Agreement (Chesapeake Executive Council 2014) which established five major themes: Abundant Life, Clean Water, Conserved Lands, Engaged Communities and Climate Change. Ten total goals were established under the themes, each with specific outcomes and totaling 31 separate outcomes to be used to determine progress. Presently, the CBP has identified 230 science needs that aim toward developing necessary support to underpin successful management achievement of the themes, goals and outcomes.

Although AI and ML technologies are being explored to support the CBP's restoration efforts, particularly in areas like land-use mapping, habitat mapping, nutrient transport analysis, and submerged aquatic vegetation detection from satellite imagery, only two of the 230 science needs explicitly call for AI/ML methods. However, the CBP is still in the early stages of understanding how AI can be fully leveraged for improving decision-support on increased effectiveness with targeting, planning, and implementation towards addressing its restoration priorities. In general, many of the scientific problems involve understanding relationships between human systems, environmental resources and external factors, and generating relevant data describing those systems.

The CBP is focused on determining and implementing appropriate management strategies; therefore, AI/ML tools that make short-term predictions or are not interpretable may be of limited use. Management questions require accurate mapping of the past and models that can confidently predict the effects of different management options. It is expected that the participants in this workshop would identify areas to successfully apply AI/ML techniques that could be used as a tool to address management priorities.

Introductory Overview of AI and ML – Alison Appling (USGS)

Alison Appling presented a broad overview of AI and ML as they apply to CBP research needs. The presentation included definitions and relationships among AI, ML, and major classes of methods including tree-based machine-learning methods, neural networks, deep learning, generative AI, automatic ML selection, and eXplainable AI. A taxonomy of ML-suitable tasks was presented, including but not limited to the four major categories of classification, clustering, regression, and dimensionality reduction.

A light analysis of the STAC literature compilation of Chesapeake Bay ML applications was then used to direct a deeper dive into several major ML methods: tree-based ML (decision trees, random forests, and gradient boosting); neural networks (neurons, backpropagation, gradient descent, influential neural network architectures); differentiable modeling that hybridizes neural networks with process-based components; and SHapley Additive exPlanations (SHAP) as a popular Explainable AI method. An overview of terminology and scope is provided in Figure 1, which uses a Venn diagram to show how AI artificial intelligence encompasses ML, neural networks, DL, and generative AI.

The presentation addressed strategies for selecting an ML method, which could include following published guides, employing automatic ML selection tools, or prioritizing between accuracy and interpretability to fit the given research need. Lastly, the presentation covered some opportunities to leverage generative AI – either directly, in environmental modeling, or indirectly, by applying large language modeling tools to support the processes of developing new ML applications or exploring the scientific literature.

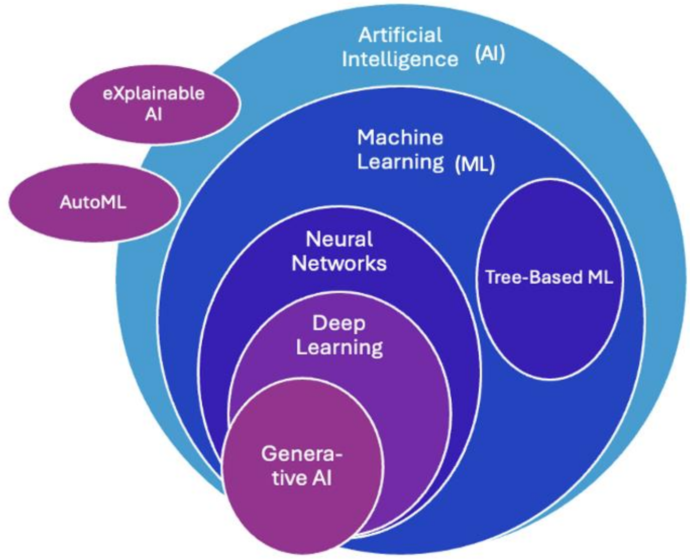


Figure 1. Conceptual Venn diagram illustrating the relationship among artificial intelligence (AI), machine-learning, neural networks, deep learning, and generative AI.

Literature Summary of Watershed and Living Resources Studies Involving AI/ML – Kelly Maloney (USGS)

Synthesizing AI/ML literature as it pertains to living resources in the watershed is fraught with many obstacles including different terminology, different ecological disciplines, studies often being published in non-ecological outlets, differing extent of studies, and a rapidly evolving field. Given these obstacles, the presentation opened by focusing on three key methodological papers: the Phillips et al. (2006) paper on maximum entropy, the Cutler et al. (2007) paper on random forests, and the Elith et al. (2008) paper on boosted regression trees. All three methods have seen a dramatic increase in application since the date of publication together totaling, as of February 5, 2025 (Scopus query), over 2,000 citations per year. A word cloud analysis for each of these papers indicated that “species distribution”, “habitat”, “climate”, “spatial”, and “prediction” were routinely mentioned in titles of the most recent 2,000 publications citing each paper.

The presentation then highlighted studies focused on species distribution and habitat assessments at either a regional/continental or Chesapeake Bay watershed scale. Random forest machine-learning algorithms were used for predicting firefly presence/absence and relative abundance in the eastern United States (McNeill et al. 2024), presence/absence of fish species in the Chesapeake Bay watershed (Maloney et al. 2022b), and stream health as measured by benthic macroinvertebrates within the contiguous United States (Hill et al. 2017) and Chesapeake Bay region (Maloney et al. 2022a). This section ended with a case study that used double ML (from the field of causal inference) to predict bird species abundance and population trends across North America and how such a method can reduce confounding bias (Fink et al. 2023).

Next, the presentation highlighted a timeline of key AI/ML cases studies implementing AI/ML techniques within the Chesapeake Bay watershed (Table 1). The presentation ended with highlighting several studies between ecologists and statisticians (e.g., Weinhold et al. 2020 and

Schmid et al. 2011) emphasizing that building such cross-disciplinary collaborations to leverage both field's expert knowledge can strengthen our understanding of the system and with a list of some key review paper citations.

Reference	Target	Methods
Goetz et al. 2007	Bird species richness and abundance	Regression tree vs traditional approaches
Maloney et al. 2009	Stream health from benthic macroinvertebrates	Comparison of regression trees, random forest, conditional regression trees and conditional random forest
McCabe 2019	Presence/absence of blue catfish	Boosted regression tree
Merriam et al. 2019	Brook trout occupancy	Boosted regression tree
Woods et al. 2023	Fish community change with changing environmental conditions	Random forest

Table 1. Timeline of Chesapeake Bay watershed studies using AI/ML approaches and involving living resource and habitat endpoints.

In conclusion the presentation provided numerous examples of how AI/ML has been used to explore living resources both within and outside the Chesapeake Bay watershed. The majority of example have used AI/ML in a prediction framework, but recent work is incorporating interpretable AI and causal inference techniques.

Literature Summary of Estuarine and Living Resources Studies Involving AI/ML – Jian Shen (VIMS)

ML has emerged as a valuable modeling tool for time-series forecasting of environmental state variables. Unlike traditional deterministic modeling approaches, ML offers a cost-effective alternative that leverages the full potential of observational data. In the Chesapeake Bay region, ML has been applied across a wide range of studies, encompassing supervised and unsupervised learning, neural networks, and deep learning (DL) techniques. Applications include forecasting storm surge, surface waves, and saltwater intrusion, as well as ecosystem-related modeling such as predicting harmful algal blooms, estimating primary production, forecasting DO levels in the Bay's main channel, and estimating hypoxia volume. These efforts have utilized observational data, model-generated outputs, satellite imagery, and hybrid approaches that integrate numerical models with empirical observations. These studies have demonstrated and expanded the practical potential of ML in coastal and estuarine science.

ML models have shown strong performance in forecasting storm surge and surface waves in the Bay, which are primarily driven by wind and governed by relatively well-understood physical dynamics. However, ecological applications pose greater challenges. Many ML models rely heavily on in situ observations, such as salinity, temperature, nutrient concentrations, and stratification, to predict ecological variables at fixed locations. This limits their utility for scenario-based analyses that aim to answer "what-if" questions. Applying ML to simulate daily variations in two or three dimensions remains a significant challenge. Although some models can reproduce observed state variables with reasonable accuracy, they may fail to capture the underlying biogeochemical processes. For example, a model might simulate DO levels but fail to reflect changes resulting from nutrient load reductions. Nonetheless, integrating numerical model outputs with observational data during training offers a promising path forward. The rapid advancement of ML technologies is creating new opportunities to simulate the environment and ecological state variables, such as salinity, temperature, and DO in two and even three dimensions, improving both spatial coverage and predictive capability.

AI/ML Integration of Satellite Remote-sensing: Data Harmonization Challenges and Gaps – Stephanie Schollaert Uz (NASA)

This presentation reviewed applications of satellite remote-sensing with AI/ML artificial intelligence and machine-learning for Chesapeake Bay living resources, highlighting both opportunities and limitations. The studies referenced demonstrated fused radar-optical classification of tidal wetlands (Lamb et al. 2021), SAV mapping using WorldView-2 imagery and deep convolutional neural networks (Coffer et al. 2023), chlorophyll *a* prediction from Visible Infrared Imaging Radiometer Suite (VIIRS, Yu et al. 2022), satellite data combined with machine-learning models (Yu et al. 2022), and simultaneous retrieval of chlorophyll *a*, turbidity (Pahlevan et al. 2022), and colored dissolved organic matter (CDOM) from Landsat, Sentinel-2, and Sentinel-3 using mixture density networks (Pahlevan et al. 2022). For water clarity, the DEEP-VIEW framework integrates Moderate Resolution Imaging Spectroradiometer (MODIS), Ocean and Land Color Instrument (OLCI), and VIIRS (Schollaert Uz et al. 2024), and additional approaches apply non-Euclidean water distance interpolation for mapping diffuse attenuation (Schollaert Uz et al. 2024 & Clark et al. 2024). Hypoxia forecasting was conducted with convolutional and long short-term memory (LSTM) networks trained on satellite-derived reflectance and hydrodynamic model fields (Zheng et al. 2024). An example SAV classification result is shown in Figure 2.

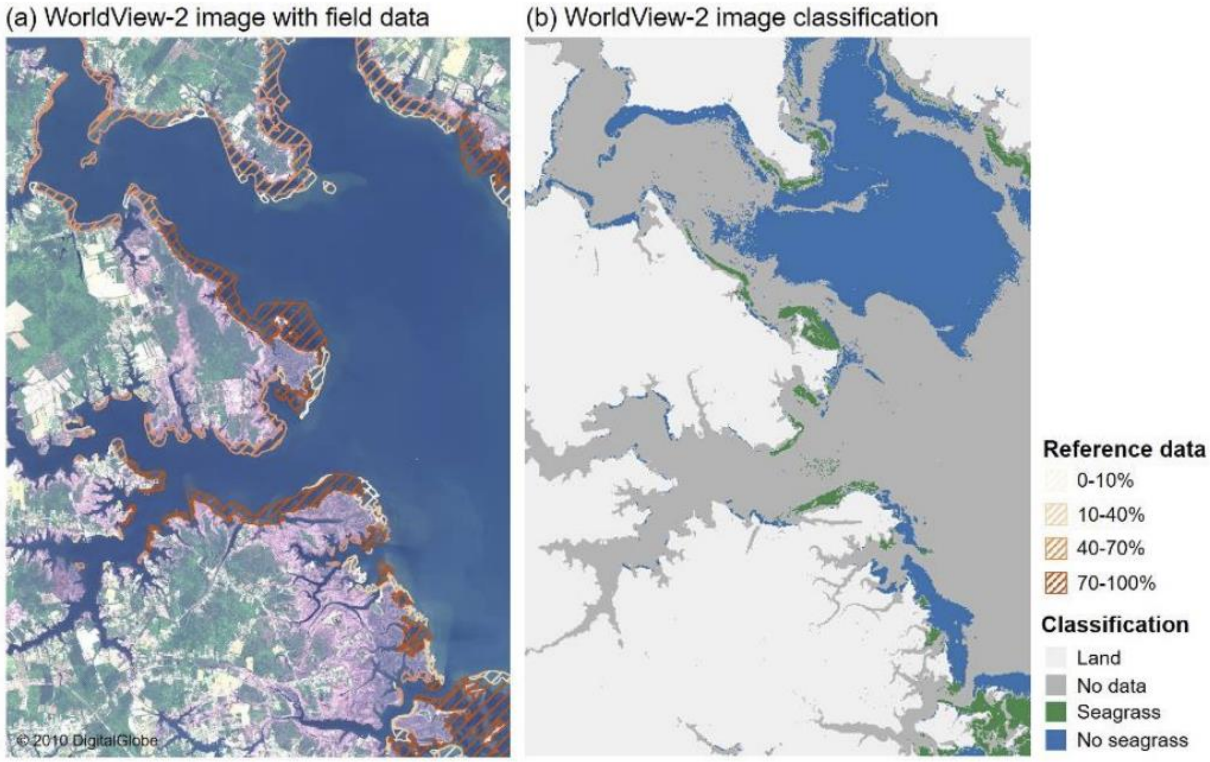


Figure 2. (a) Submerged aquatic vegetation (SAV) classification and extent using WorldView-2 imagery (1.84-m resolution, six visible bands) for Mobjack Bay, VA, on May 4, 2015, overlaid with reference data delineating seagrass percent cover obtained from Virginia Institute of Marine Science (VIMS) in May through November 2025. (b) Results of an image classification with classes for land, no data, SAV, and no SAV. Source: Coffer et al. 9(2023).

The presentation identified key satellite remote-sensing opportunities and challenges for the Chesapeake Bay, including consideration of sun glint, and atmospheric interference; land adjacency effects from narrow waterways; optically complex waters; and tradeoffs between spatial resolution and revisit frequency. Although aquatic sensors provide high signal-to-noise ratios (SNR) and daily coverage at coarse resolution (300 m–1 km), terrestrial sensors offer finer spatial resolution (10–30 m) at lower SNR and longer revisit intervals. Validation remains limited by the small number of above-water radiance measurements.

Limitations for submerged aquatic vegetation (SAV) classification include requirements for image acquisition at consistent tidal stage; frequent missing data due to clouds; signal attenuation with water depth causing mischaracterization of deep edges; and multispectral imagery being insufficient for seagrass species identification. Hyperspectral imagery can distinguish plant types through pigment discrimination but is not yet routinely available at high spatial resolution.

Ongoing interagency work connects satellite providers and users to address runoff, water-quality, algal blooms, carbon fluxes, and flooding. Recent deployments include AErosol RObotic NETwork- Ocean Color (AERONET-OC) (2021) and hyperspectral spatial-spectral understanding network (HyperNet) (2023) for calibration/validation. Priorities include improving atmospheric correction, spectral libraries, and phytoplankton classification to fill monitoring gaps. Despite limitations, satellite data fill a critical data gap and provide an important source of wide-spatial data coverage. Ground-truth data will always be needed to calibrate satellite data, the combination of both offers better data coverage for improved monitoring and prediction.

Session II: Identify the Challenges and Gaps in Applying AI/ML Approaches to Chesapeake Bay Data

This session examined the major challenges and gaps that hinder the application of AI/ML approaches to Chesapeake Bay tidal and non-tidal data. Presenters highlighted issues such as limited or inconsistent datasets, barriers to harmonizing information across sources, and the need for greater expertise in algorithm design and implementation. The discussion also raised concerns about the lack of accessible software code for replication or adaptation, the difficulty of communicating and interpreting AI/ML outputs, and the need for stronger coordination among researchers and managers within the CBP partnership. Invited presentations in this session included:

- Patrick Bitterman (Kent State University) – GeoAI and Social Systems Modeling
- Mike Evans (Chesapeake Conservancy) – Integrated AI Models to Forecast Land-use Change
- Shuyu Chang (PSU) – Advances in Water-quality Predictions: Datasets and Learning
- David Parrish (VIMS) – Modeling Light Conditions in the York River Estuary by Anchoring Satellite Imagery with High-Frequency In-Situ Observations
- Matthew Cashman (USGS) – Physical Habitat is More Than a Sediment Issue: A Multi-dimensional Habitat Assessment Indicates New Approaches for River Management
- Taylor Woods (USGS) – Observed and Projected Functional Reorganization of Riverine Fish Assemblages from Global Change
- Jenn Fair (USGS) – Images to Info: the USGS Flow Photo Explorer
- Sean Emmons (USGS) – Leveraging Machine-learning and Expert Knowledge to Unravel the Complexities of Multiple Freshwater Ecosystem Stressors

GeoAI and Social Systems Modeling – Patrick Bitterman (Kent State University)

This presentation detailed ongoing research that integrates spatially explicit machine-learning approaches with agent-based models of human decision making to better represent feedbacks in social-ecological systems. Drawing from ongoing National Science Foundation-funded research (CNH2-L: #2009248), Bitterman (Kent State University) presented research combining Geospatial Artificial Intelligence (GeoAI methods) (e.g., random forest, eXtreme Gradient Boosting or XGBoost) with structured data on land-use planning and best management practice (BMP) implementation to model scenario-based outcomes under Chesapeake Bay Phase 3 Watershed Implementation Plan inputs. Model outputs suggest strong spatial and scalar path dependencies in management trajectories, highlighting how past implementation patterns shape future land-use and conservation outcomes at county and local scales. These model results are supported by results of qualitative interview results and document analysis.

A key contribution of this work is the development of a hybrid modeling framework that couples GeoAI methods with more traditional social systems modeling approaches to reflect regulatory feedbacks and decision making processes. The presentation emphasized the importance of incorporating fine-scale social data (e.g., planning documents, local physical and social context) to improve predictive accuracy and system understanding. The results of this work demonstrate that management decisions are not only reactive to environmental targets and local conditions, but are also strongly conditioned by prior choices, institutional arrangements, and socio-political context – insights that could advance the CBP modeling system's treatment of human dynamics.

Integrated Deep-Learning Models to Forecast Land-use Change – Mike Evans (Chesapeake Conservancy)

DL models are being applied to forecast land-use change in the Chesapeake Bay watershed, with objectives to account for variation in types of places, accommodate dynamic growth trajectories, and improve the surface of transition probabilities. The aim is to generate spatially and temporally accurate allocations of projected population and employment. Long Short-Term Memory (LSTM) networks capture temporal dynamics, and convolutional LSTMs incorporate spatial information to represent growth processes and refine transition probabilities.

Socio-economic data are ordered through self-organizing and hierarchical self-organizing maps to classify counties and census blocks as "types of places." Housing, population, employment, and migration time-series support these classifications. The Chesapeake Bay Land Change Model (CBLCM) integrates these data and methods to project residential and commercial development, along with farmland and forest conversion (Claggett et al. 2023; Figure 3).

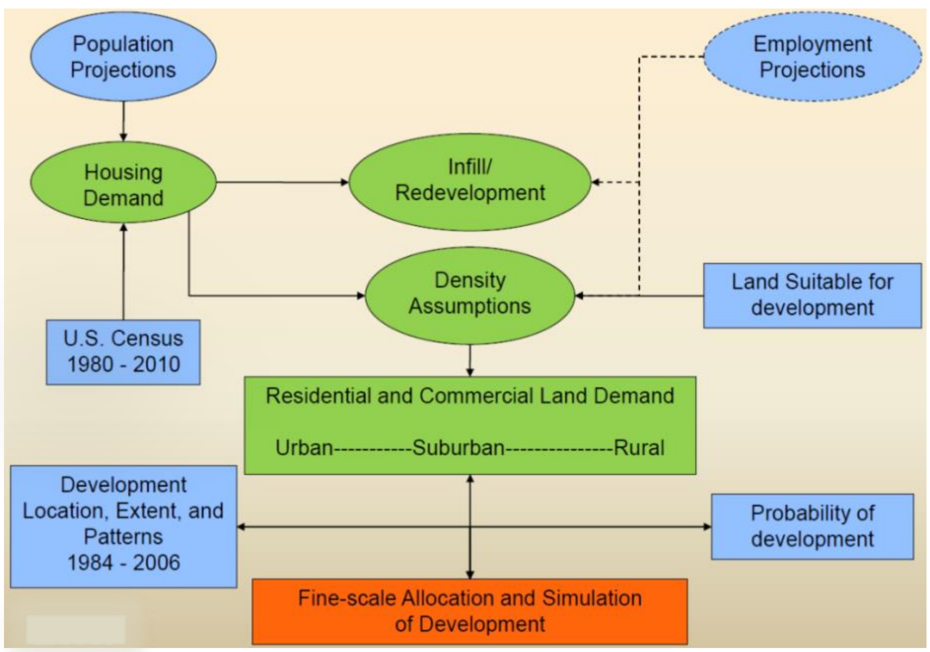


Figure 3. Predicted 2020 commercial development probabilities produced by the Chesapeake Bay Land Change Model (CBLCM), shown alongside observed 2020 development for comparison. The results illustrate the model's application of deep learning to spatially allocate growth across census blocks (Claggett et al. 2023).

Outputs include estimates of future households, population, and employment by wastewater service type (sewer vs. septic), with model performance evaluated using quantization and topographic error measures. Figure 4 shows predicted 2020 commercial development probabilities from the Chesapeake Bay Land Change Model alongside observed 2020 development, illustrating how well the model captured actual growth patterns (Claggett et al. 2023).

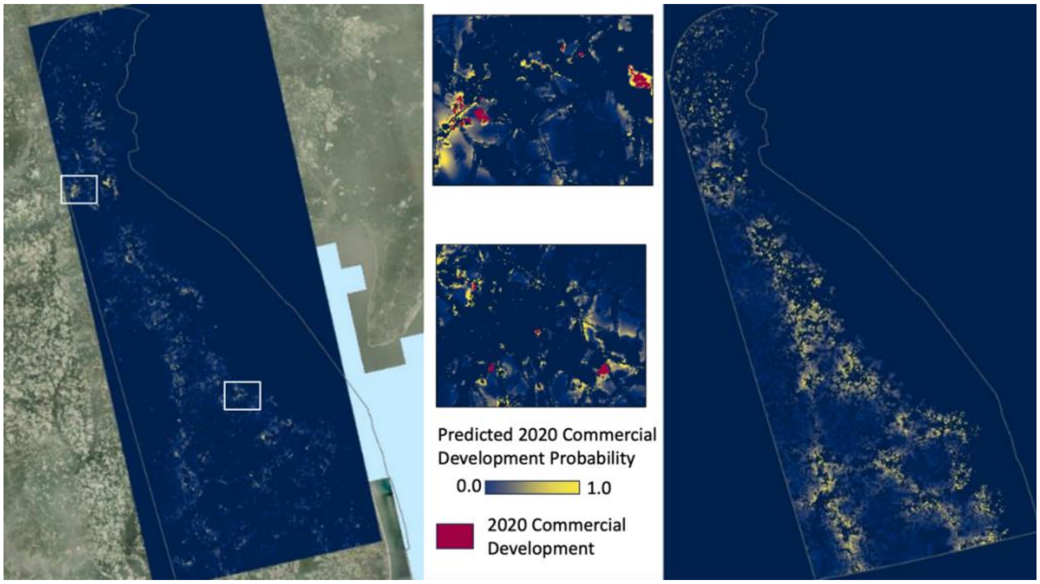


Figure 4. Predicted 2020 commercial development probabilities produced by the Chesapeake Bay Land Change Model, illustrating application of deep learning to allocate growth across census blocks (Claggett et al. 2023).

Advances in Water-quality Predictions: Datasets and Learning Frameworks – Shuyu Chang (PSU)

The Integrated Watershed Attributes and Nutrient Data (IWAND) dataset was introduced as a new benchmark resource to support large-scale predictions of riverine nutrient concentrations (Chang et al. 2025). IWAND builds on earlier efforts such as Catchment Attributes and Meteorology for Large-sample Studies (CAMELS; Addor et al. 2017) and the Catchment Attributes and Meteorology for Large-sample Studies – Chemistry (CAMELS-Chem; (Sterle et al. 2025) dataset; CAMELS-Chem augments the original CAMELS framework by adding stream-water chemistry and atmospheric deposition data for hundreds of U.S. catchments. IWAND incorporated in-situ records, catchment attributes, nutrient inputs, and climate forcing. Compared to prior benchmarks, IWAND offers more extensive records per site, broader spatial and temporal coverage, and improved representation of human influences, making it a robust foundation for developing and testing water-quality prediction models (Chang et al. 2025).

The presentation emphasized the importance of large-sample hydrology and continental-scale (CONUS-wide) datasets for advancing water-quality modeling. By leveraging AI methods with extensive benchmark datasets, it is possible to improve predictive accuracy and enhance physical understanding of biogeochemical processes across diverse watersheds. These capabilities are particularly relevant for regional systems such as the Chesapeake Bay watershed, where advances in nutrient prediction can inform management strategies and support ongoing progress in water-quality restoration.

Modeling Light Conditions in the York River Estuary by Anchoring Satellite Imagery with High-Frequency In-Situ Observations – David Parrish (VIMS)

This presentation described efforts to model light conditions in the York River estuary by anchoring satellite imagery with high-frequency turbidity data collected from the Chesapeake Bay National Estuarine Research Reserve (CBNERR)-VA/VIMS Dataflow platform (Virginia Estuarine & Coastal Observing System, 2025). The Dataflow system provides surface observations every 2–3 seconds, producing thousands of turbidity measurements per day along vessel transects. These in-situ data were paired with PlanetScope satellite imagery (~3 m resolution, eight spectral bands, near-daily coverage since 2022). Atmospheric correction with ACOLITE was applied to generate surface reflectance inputs for modeling (Vanhellemont and Ruddick, 2016).

Random forest regression was used to estimate turbidity from the eight surface reflectance bands (Figure 5; Parrish et al. 2025). This method handles non-linear relationships and interactions between variables without distributional assumptions. Block cross-validation was applied to reduce the influence of spatial autocorrelation by creating spatially independent training and testing subsets. Results from multiple sampling dates in 2023–2024 indicated that the random forest approach can reproduce turbidity patterns when anchored to dense in-situ measurements.

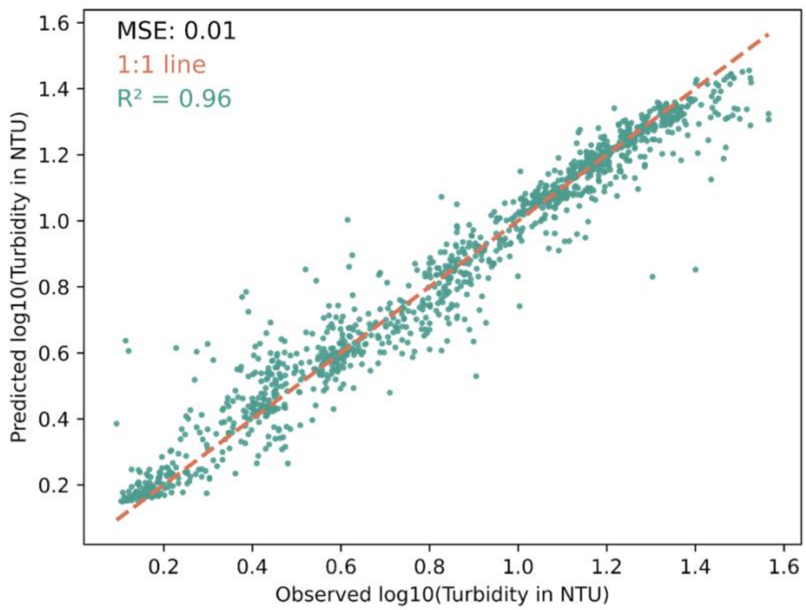


Figure 5. Random forest regression with block cross-validation applied to estimate turbidity in the York River estuary from PlanetScope surface reflectance anchored with Chesapeake Bay National Estuarine Research Reserve (CBNERR)–VA/VIMS Dataflow measurements; Dataflow: 03/29/2023, 05/21/2024, 05/22/2024, 06/20/2024, 06/21/2024 (modified from Parrish et al. 2025).

Next steps include expanding beyond turbidity to estimate the light attenuation coefficient (Kd), which forms the basis of Chesapeake Bay water clarity standards. This requires a hierarchical modeling approach linking Kd to turbidity, chlorophyll, and CDOM/salinity. Further work could include quantifying error in turbidity - and Kd-based estimates and addressing both spatial and temporal autocorrelation. Although PlanetScope imagery has only been available in the Bay since 2022, the data provide a freely available, high-resolution resource for advancing water clarity assessments.

Physical Habitat is More Than a Sediment Issue: A Multi-dimensional Habitat Assessment Indicates New Approaches for River Management – Matthew Cashman (USGS)

This presentation highlighted experiences, motivations, lessons learned, difficulties, and other considerations with ML from the recent publication, "Physical habitat is more than a sediment issue: A multi-dimensional habitat assessment indicates new approaches for river management" (Cashman et al, 2024).

Because physical habitat and sediment are major stressors for stream health in the Chesapeake Bay, the study predicted habitat quality across the watershed using metrics familiar to local stakeholders. It identified two distinct clusters and dimensions of physical habitat, highlighting related metrics and the hydrologic processes that support them. By overlaying these model outputs with models of suspended sediment and flow alteration, the study concluded that management actions focusing solely on restricting sediment - without addressing flows or in-channel hydromorphic diversity - are unlikely to improve the habitat metrics.

Challenges included the quality of source data - as the habitat metrics modeled were visually scored, semi-quantitative methods, with field-based uncertainty accounting for ~80% of the uncertainty in the final ML models.

Most important, the presenter emphasized the challenges of using machine-learning methods to answer causal, cause-effect, and counterfactual questions. Referring to Judea Pearl's causal hierarchy framework (Pearl 2009), the presenter explained that most traditional ML operates at the lowest, associative level of the hierarchy and struggles to answer causal questions accurately (Bareinboim et. al, 2022). Instead, causal inference techniques—designed specifically for causal questions such as cause-effect interventions and counterfactual thinking (e.g., outcomes under alternative scenarios)—are used in fields like public health, medicine, and econometrics and are now emerging in ecology, hydrology, and earth sciences (Kratzert et al. 2019).

The presentation highlighted several subfields of causal inference, such as causal discovery and causal ML, listing various modeling methods and introductory textbooks on the topic. Despite his engagement with the field over the past year and a half, the presenter concluded that causal ML is rapidly developing and not all methods are suitable or have 'off-the-shelf' accessibility, staying up-to-date on developments in the field can help address future management questions.

Observed and Projected Functional Reorganization of Riverine Fish Assemblages from Global Change – Taylor Woods (USGS)

This presentation focused on two forecasting projects that apply ML to assess ecological and hydrologic changes under scenarios of climate and land-use change scenarios. The ecological component used random forest models to predict habitat suitability of fish functional groups, classifying abundances as low, medium, or high across multiple future scenarios. Results suggested that species with generalist, warm-water, fine-substrate, and slow-water traits are projected "winners," whereas cold-water, clean-substrate, fast-water taxa are likely "losers" (Figure 6).

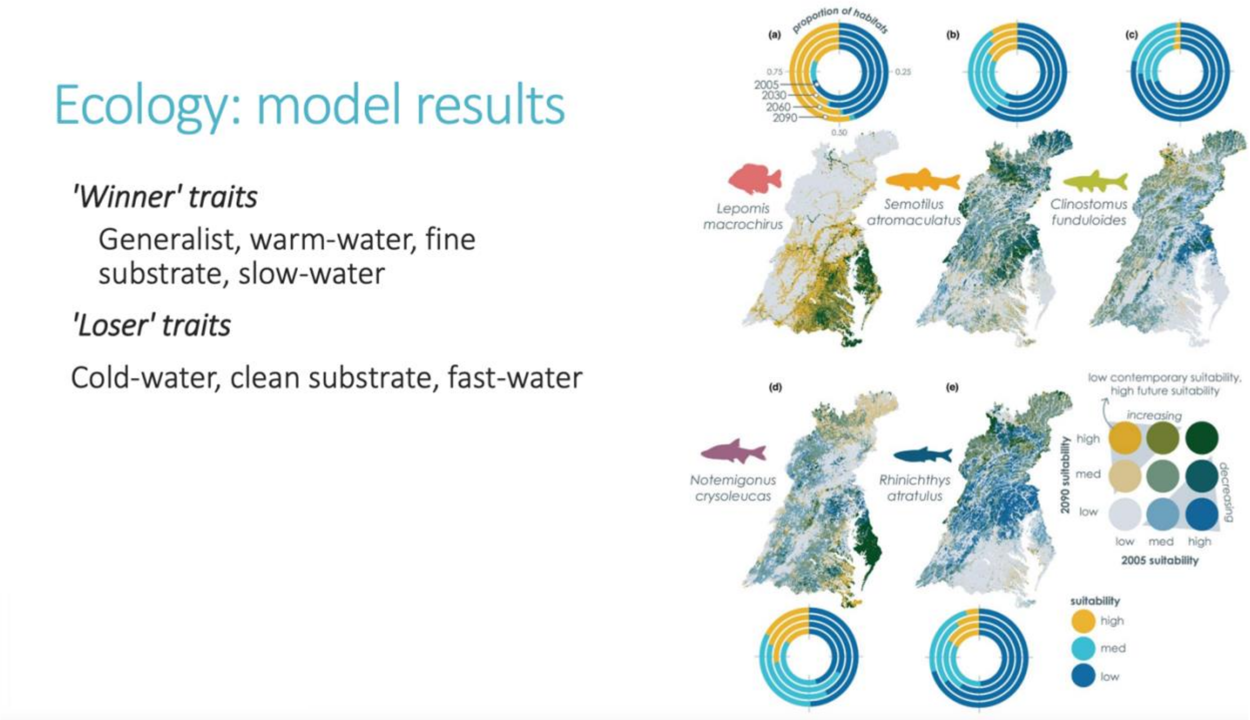


Figure 6. Ecology model results showing projected "winner" and "loser" traits of fish communities under climate and land-use change scenarios.

The hydrologic component evaluated how climate and land-use change will affect flow regimes. Random forests with temporally lagged predictors performed similarly to or better than neural networks, demonstrating predictive power while maintaining interpretability. However, data limitations, particularly small biological sample sizes, pose constraints, and uncertainty in applying ML outputs to ecological models remains an ongoing challenge. The work highlights both the promise and limitations of ML approaches for Chesapeake Bay forecasting applications.

Images to Info: the USGS Flow Photo Explorer – Jenn Fair (USGS)

The [USGS Flow Photo Explorer (FPE) Tool](https://www.usgs.gov/apps/ecosheds/fpe) is a web platform that supports a stream monitoring network for small streams (https://www.usgs.gov/apps/ecosheds/fpe). It features a data system that allows users to upload images collected with low-cost trail cameras, an annotation tool for allowing users to rank pairs of images, and a deep learning model that learns from these annotations to predict a relative streamflow hydrograph. The Flow Photo Explorer (FPE) platform currently has more than 350 users from state, federal and tribal agencies, universities, non-governmental organizations, local municipalities, and other private organizations and individuals. The FPE data system currently stores over 6 million images and over 300,000 annotations, and hosts approximately 70 models predicting streamflow dynamics. A recent evaluation of deep learning model performance indicates that the FPE data system will be useful as a low-cost, non-contact method for monitoring streamflow dynamics in under-monitored, dynamic and particularly vulnerable headwater streams (Goodling et al. 2025).

Leveraging Machine-learning and Expert Knowledge to Unravel the Complexities of Multiple Freshwater Ecosystem Stressors – Sean Emmons (USGS)

This presentation described the use of AI/ML to evaluate multiple interacting stressors on freshwater ecosystems in the Chesapeake Bay watershed. The overarching research question addressed was: What are the key stressors affecting stream health, and do these vary regionally? Goals included identifying hierarchical effects of stressors on benthic macroinvertebrate indicators, predicting biological responses under different stressor conditions, and developing a spatial prioritization framework to support watershed conservation and restoration.

The approach combines expert knowledge with Bayesian network learning to identify driver–stressor–response relationships. This involves structure learning, model averaging, and bootstrapped networks, while integrating prior ecological knowledge to constrain connections (Figure 7). By retaining only consistent relationships across bootstraps, the method improves confidence in causal links. The framework supports predictions of how biological metrics will change under stressor scenarios and enables prioritization of management actions such as resisting, directing, or accepting change.

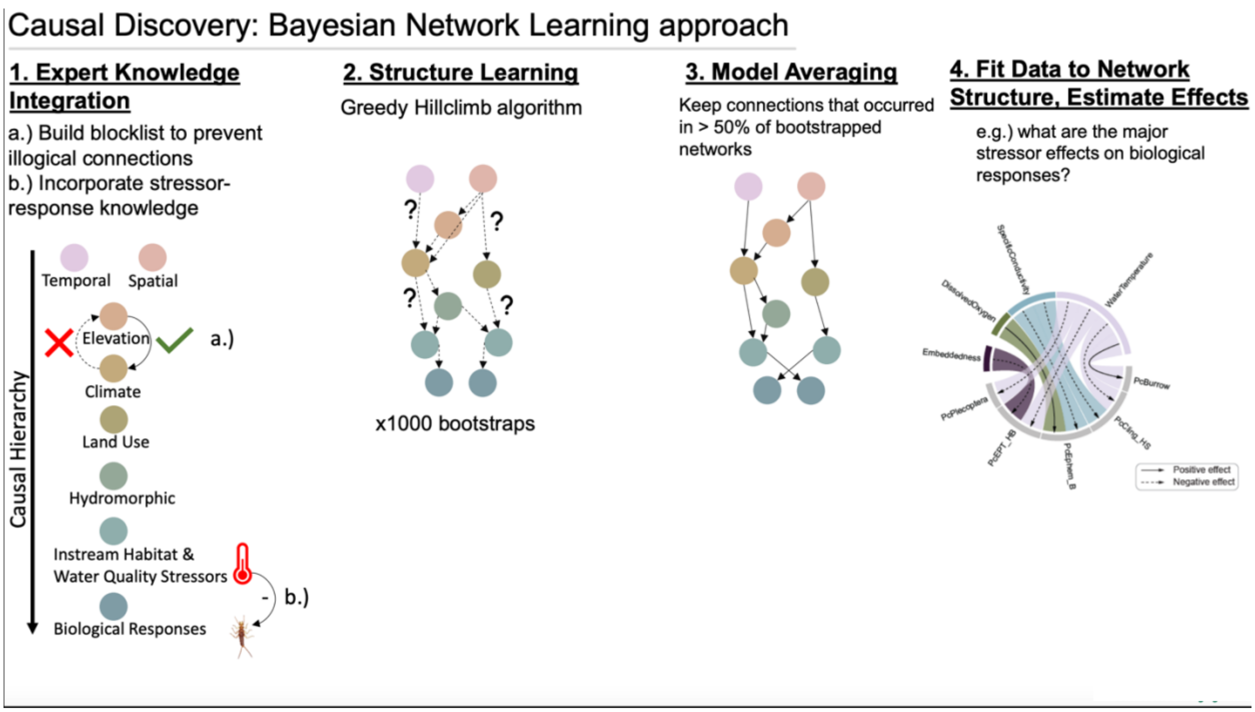


Figure 7. Conceptual hierarchy of stressors and biological responses in stream ecosystems, showing how temporal, spatial, climatic, land-use, and hydromorphic drivers influence instream habitat, water-quality stressors, and ultimately benthic macroinvertebrate responses.

This work directly addresses needs identified by the CBP's Stream Health Workgroup and illustrates the potential of causal discovery techniques to guide restoration planning under complex, interacting stressors.

Session III: Develop recommendations and identify opportunities for harnessing the power of AI/ML approaches to address Chesapeake Bay issues

This session focused on developing recommendations and identifying opportunities to expand the role of AI/ML in restoration of Chesapeake Bay tidal and non-tidal regions. Participants discussed where the CBP partnership could benefit most from these approaches, emphasizing their potential to generate new insights and support more efficient decision making. The group also considered strategies for delivering AI/ML outputs to managers in clear and actionable ways, along with guidelines for standardizing and streamlining how AI/ML methods are applied to monitoring data. Finally, the session highlighted opportunities to strengthen collaboration and build synergies among partners to accelerate the integration of AI/ML into Bay science and management. Invited presentations in this session included:

- Chaopeng Shen (PSU) – State-of-the-Art AI & Physics-Informed ML in Hydrology and Water-quality: Insights and synergies
- Dong Liang (UMCES), Chaopeng Shen (PSU), Vandana Janeja (UMBC), Kelly Maloney (USGS), Robert Sabo (EPA), Alison Appling (USGS) – AI/ML Community Development (Panel)

State-of-the-Art AI & Physics-Informed ML in Hydrology and Water-quality: Insights and Synergies – Chaopeng Shen (PSU)

Chaopeng Shen presented advances in integrating ML with process-based hydrologic and water-quality models. Traditional neural networks, such as LSTMs, have shown predictive skill for variables like dissolved oxygen (DO) and nutrients but face limitations in interpretability, transferability, and performance in data-scarce or extreme conditions. To address these gaps, Shen's group developed differentiable parameter learning (dPL), which links neural networks with governing equations to constrain learning. Their framework is illustrated in Figure 8, which shows two dPL workflows alongside a traditional calibration approach.

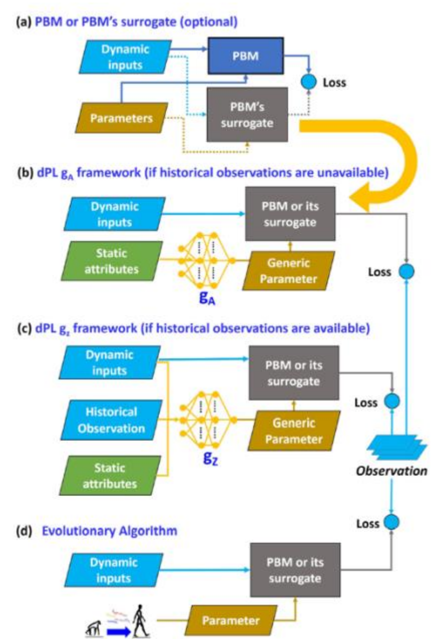


Figure 8. (a) A deep learning model is trained to mimic the outputs of a process-based model (PBM). This step is optional because one may also directly implement the model in a DL platform. (b) Workflow of the first differentiable parameter learning (dPL) option for deep neural network gA: parameters are inferred by a network (in our case, a separate LSTM network) based on auxiliary attributes. These parameters are then sent into the PBM, whose outputs are compared to the observations to calculate the loss (the difference between objective function and observation). (c) Workflow of the second dPL option for deep neural network gZ: historical observations (meteorological forcings and observed responses) are additional inputs to the parameter estimation network. (d) Traditional site-by-site parameter calibration framework. Reproduced from W. Tsai et al. (2021).

Shen emphasized the potential for these methods in Chesapeake Bay water-quality modeling, where they can provide scale-relevant predictions, link landforms and management practices to outcomes, and incorporate uncertainty quantification. He also highlighted spectral convolutional Fourier Neural Operators (SC-FNOs) as a promising tool for solving partial differential equations (PDEs) orders of magnitude faster than traditional models, with improved sensitivity for parameter inversion. These approaches offer transferable, interpretable, and efficient tools for large-scale forecasting and management support.

Panel: AI/ML Community Development – Dong Liang (UMCES), Chaopeng Shen (PSU), Vandana Janeja (UMBC), Kelly Maloney (USGS), Robert Sabo (EPA), Alison Appling (USGS)

A powerful variety of highly predictive AI/ML models can be readily applied to improve the CBP's understanding of fundamental processes that drive watershed and estuarine conditions. With the large diversity of potential applications, however, the path forward for effectively leveraging AI/ML across scales and disciplines is still being defined. One key motif that emerged when discussing a path forward was the need for meaningful partner engagement throughout the conception, development, and application of AI/ML products, a process known as co-production, in which technical tools and analyses are created collaboratively with end users to ensure their relevance and usability. By working with targeted CBP workgroups or other key stakeholders in the Bay watershed, research groups looking to inform the CBP can better define what key questions need to be addressed as well as the key research products that stakeholders need to make the research actionable.

During these discussions, questions about spatial and temporal scale, endpoint(s) of interest, the desired predictive ability vs. explainability, and others were explored. In these conversations with Bay partners, the panel discussion highlighted an opportunity to identify new and cultivate existing synergies across research disciplines and management domains. Before proceeding with AI/ML analyses, it would be helpful to determine what enhancement and contribution the AI/ML application can make to the current suite of existing models and tools (besides only being more predictive) and identify collaborators who can assure effective and accurate integration of environmental information into the models. AI/ML's looser model structure, ability to better handle some co-linearities across predictors, large information needs, and ability to accurately describe multiple environmental phenomena allow it to integrate multiple predictor datasets that span environmental disciplines. These characteristics serve as a platform for promoting inter-disciplinary collaborations and stakeholder engagement. Also, improvements in explainable AI/ML features, like the use of Shapley values, partial dependent plots, and other metrics/tools, highlight the opportunity for these models to be more wholly interpreted by inter-disciplinary teams. The final key points that was highlighted by the panel discussion were the need to the availability of model code and the production and availability of needed predictor-response variable datasets, the latter of which sounds conceptually simple but is computationally difficult.

Providing access and packaging the CBP partnership’s preferred predictor data (e.g., Chesapeake Assessment Scenario Tool (CAST) matched to Nontidal Monitoring (NTN) network stations and available at NHD+ scale) will ensure AI/ML research groups are using the best available, locally tailored data that the CBP has approved to track progress. Developing computationally efficient methodologies to adjust other non-CBP datasets to scales of interest will be key as well because AI/ML can take advantage of other ancillary environmental information from OPEN Library and other open data sources (e.g., Stream-Catchment (STREAMCAT) dataset). In closing, the panel discussion highlighted that AI/ML can be leveraged to help communities understand drivers of local water-quality and habitat while also providing inference into broader watershed and estuarine ecological conditions - conditions that can be restored through partner engagement.

Breakout Group Discussions

On the second day of the workshop, participants took part in structured breakout sessions designed to explore the workshop's three objectives in greater depth. Each group (including one that met virtually) discussed a common set of guiding questions, with the aim of identifying shared themes, key challenges, and potential paths forward. The conversations concluded with brief report-outs to the full workshop audience.

The guiding questions for these breakout sessions were:

- Objective 1
    - In what ways can AI/ML approaches be applied (or have already been applied) to issues relevant to Chesapeake Bay restoration?
    - What advantages and limitations do AI/ML methods present compared to more traditional approaches?
- Objective 2
    - What challenges or gaps have you experienced when applying AI/ML in the Chesapeake Bay context (or in related fields)?
    - What strategies have you used, or could be used, to address these challenges?
- Objective 3
    - What are the primary barriers to broader adoption of AI/ML in Chesapeake Bay research and management?
    - What kinds of forums, workshops, or working groups could help foster collaboration among AI/ML researchers, Bay scientists, and resource managers?

The following sections summarize the discussions across groups for each objective, highlighting common observations, areas of divergence, and opportunities for further research and collaboration; full breakout group responses are available to review in Appendix D.

Objective 1

Breakout groups identified a range of current and potential applications of AI/ML relevant to Chesapeake Bay restoration. Examples included tracking agricultural BMPs on croplands, recalibrating land-use layers in Pennsylvania, evaluating BMP performance, informing swimming advisories, and predicting habitat conditions. The USGS Flow Photo Explorer Tool was cited as an example of using ML to monitor streamflow dynamics in small systems, including stormwater BMPs. Groups also noted that AI/ML can be applied in intermediate steps of modeling workflows (e.g., preparing data for integration), not just in end-point predictions.

Participants emphasized that AI/ML techniques are well suited to integrating diverse datasets such as satellite imagery, in-situ monitoring, and high-frequency sensor data. These approaches are scalable and robust, with the ability to capture non-linear and complex interactions. Models trained on national-scale datasets have in some cases outperformed those built strictly on Chesapeake Bay-focused data. However, groups agreed that excessive diversity in training data can dilute meaningful signals, and that a moderate level of diversity is most effective to avoid overfitting. AI/ML can also support data harmonization and fill spatial and temporal gaps in existing monitoring. Incorporating non-Bay datasets was seen as a way to strengthen models by providing additional variability and context.

In comparison to process-based models, AI/ML methods are generally more flexible and accessible, requiring less reliance on domain-specific expertise. However, participants cautioned that these approaches should be viewed as complementary rather than as replacements for mechanistic models. Limitations identified included the lack of interpretability ("black box" issues), the potential for spurious inferences, and challenges in communicating results to managers and stakeholders. Quantification and transparent communication of uncertainty were noted as areas where AI/ML approaches currently lag behind traditional statistical methods. The rise of broadly available tools such as ChatGPT has increased public awareness of AI/ML but has also contributed to misperceptions about the capabilities and limitations of these approaches in technical and management contexts.

Groups also discussed challenges related to data scale and uniformity. Excessive heterogeneity in training data can obscure signals, and narrowly focused datasets may reduce transferability. Debates also centered on the advantages of raw versus derived data as model inputs. The consensus was that careful attention to input data, supported by collaboration between data scientists and domain experts, is essential for producing reliable outcomes.

Across discussions, participants reiterated that management decisions ultimately depend on understanding the effectiveness of practices. Although AI/ML approaches do not resolve all challenges, they were viewed as valuable tools for addressing persistent limitations in Bay restoration science and management.

Objective 2

Although the Chesapeake Bay is considered a data-rich system, groups identified several limitations that constrain the application of AI/ML. Key issues include mismatches in spatial and temporal scales between predictor and response variables, uneven temporal coverage, and gaps in problem-relevant data. Variation in data quality and ontology further reduces model reliability and transferability.

Challenges with data cleanliness and harmonization were also noted. Inconsistent metadata, unclear coding standards, and lack of standardization complicate reproducibility. Model inception (where outputs from one model are used as inputs to another) was described as a risk because errors can be compounded. Participants also highlighted the danger of models producing apparently correct results for the wrong underlying reasons, underscoring the need for careful feature selection and evaluation. Practical solutions included dropping sites with missing data or enabling users to flag problematic records.

Uncertainty was identified as one of the most difficult gaps to address. Current AI/ML methods have limited tools for quantifying and communicating uncertainty in ways that support management decisions. Visual and numerical approaches remain underdeveloped, which reduces confidence in model results.

Several remedies were suggested. Continued investment in data harmonization was emphasized, with the NTN concentration data synthesis effort cited as a promising example. Expanding the Chesapeake Bay Data Hub to include both monitoring data (e.g., temperature, point sources) and modeling outputs (e.g., CAST) was recommended. Other proposed approaches included the use of off-the-shelf image screening models, more systematic use of metadata (potentially supported by generative AI tools), and causal DL techniques to improve interpretability. A glossary of AI/ML terminology, already initiated within the CBP, was also cited as a step toward building shared understanding.

Objective 3

A central barrier to broader adoption of AI/ML in Chesapeake Bay applications is interpretability. Black-box models may undermine trust and can be difficult for managers to act upon, especially in a program that has long relied on process-based approaches. Developing interpretable methods and clearer approaches for communicating results were highlighted as important priorities.

Broader acceptance is also shaped by public perceptions. Many audiences associate AI with media narratives that emphasize risks, which can create skepticism. Education and outreach, ranging from vocational training to student programs, were seen as important for building familiarity and demystifying these methods. Ideas such as community "hackathon" challenges were proposed as ways to encourage engagement and identify key response variables.

Institutional and data-related barriers remain significant. AI/ML researchers and process-based model developers often work in disciplinary silos, limiting collaboration. Engaging CBP goal implementation teams (GITs) and workgroups early was recommended to align AI/ML efforts with management priorities. Participants also emphasized that strong national-scale models must be paired with fine-scale local monitoring to capture variability and transferability.

Avenues for advancing collaboration were identified. Pairing AI/ML with data visualization was recommended to make complex results more accessible. A dedicated informal network, tentatively called Ches-BRAIN (Chesapeake Bay Research with Artificial Intelligence and Networking), was suggested as a forum for researchers, modelers, and managers. Existing venues such as the Chesapeake Community Research Symposium (CCRS), HydroML Symposium, the Association of Mid-Atlantic Aquatic Biologists (AMAAB), and future STAC workshops were all viewed as important platforms. The Remote-sensing Workgroup was identified as a place where AI/ML applications could be introduced directly into CBP activities. Participants also discussed the potential value of a "front desk" chatbot or curated AI system, modeled after the historic Chesapeake Community Monitoring Network catalogue (Chesapeake Research Consortium, 2022), to help managers and researchers navigate the CBP's extensive data and modeling resources. Overall, participants agreed that adoption of AI/ML will depend not only on technical progress but also on improving transparency/trust and cross-disciplinary collaboration within the Bay community.

Recommendations

1. Strengthen data infrastructure and integration for AI/ML applications

- Harmonize spatial and temporal datasets across programs and ensure consistent metadata.
- Leverage diverse datasets, including satellite, in-situ, and high-frequency data, for use in modeling and monitoring applications and filling water-quality data gaps.
- Design monitoring and data processing efforts so that resulting products are problem-relevant and can be readily incorporated into AI/ML workflows.
- Build harmonized response and predictor datasets and develop example use cases to guide widespread AI/ML applications.

2. Leverage AI/ML for restoration of Chesapeake Bay tidal and non-tidal regions and decision support

- Use AI/ML to assess restoration practice effectiveness, evaluate progress, and identify drivers of change.
- Enhance watershed and estuarine models by integrating AI/ML model outputs and insights.
- Promote integration between AI/ML and traditional monitoring, analysis, and modeling approaches to enhance scientific credibility and transparency.
- Develop accessible AI-driven tools (e.g., Chesapeake-specific LLMs) for scenario planning to help identify management priorities.

3. Promote transparency and engage managers and stakeholders

- Advance explainable AI/ML and uncertainty protocols so that results are interpretable, credible, and trusted.
- Couple AI/ML with tailored data visualizations to improve interpretability and use at broader scales.
- Foster close engagement of managers and decision makers at all stages of co-production with AI/ML projects to ensure products align with management priorities and can be effectively applied.
- Use tailored communication strategies to translate AI/ML insights into actionable guidance for restoration planning.

4. Build collaboration and capacity

- Establish a Chesapeake Bay AI/ML network (e.g., Chesapeake Bay Research with Artificial Intelligence and Networking or "Ches-BRAIN") to foster collaboration and provide a clear place where managers and others can easily find and connect with AI/ML experts.
- Encourage development of participatory science events (e.g., hackathons) to spark innovation, strengthen cross-sector collaboration, and provide opportunities for students and early-career professionals to engage in applied problem solving.
- Invest in training and literacy programs so that scientists, managers, decision makers, and students/job seekers can effectively use and interpret AI/ML tools and outputs, supporting workforce development pathways across the Bay watershed.

References

Addor N, Newman AJ, Mizukami N, Clark MP. 2017. The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*. 21:5293–5313

Bareinboim, E., J.D. Correa, D. Ibeling, and T. Icard. 2022. On Pearl's hierarchy and the foundations of causal inference. In Probabilistic and causal inference: the works of Judea Pearl. *Association for Computing Machinery*. pp. 507-556

Cashman, M.J., G. Lee, L.E. Staub, M.P. Katoski, and K.O. Maloney. 2024. Physical habitat is more than a sediment issue: A multi-dimensional habitat assessment indicates new approaches for river management. *Journal of Environmental Management* 374: 123139. DOI: 10.1016/j.jenvman.2024.123139

Chang, S.Y., D. Aboelyazeed, K. Sawadekar, D. Chavada, C. Shen, and K.J. Van Meter. Integrated Watershed Attributes, and Nutrient Database (IWAND): long-term, largescale nutrient samples and paired inputs for water quality modeling for the Contiguous United States (1980-2023). Manuscript in preparation

Chesapeake Community Research Symposium. 2024. *2024 Chesapeake Community Research Symposium: Schedule of talks*. Annapolis (MD): Chesapeake Bay Program. 6 p.

Chesapeake Executive Council. 2014. *Chesapeake Bay Watershed Agreement*. Annapolis (MD): Chesapeake Bay Program

Chesapeake Research Consortium, 2022, Models & data: Chesapeake Community Modeling Program web page, accessed October 21, 2025, at https://ches.communitymodeling.org/models-data/

Claggett, P.R., L. Ahmed, F.M. Irani, S. McDonald, and R.L. Thompson. 2023. The Chesapeake Bay Land Change Model: Simulating future land use scenarios and potential impacts on water quality. *Journal of the American Water Resources Association* 59(6):1287–1312. https://doi.org/10.1111/1752-1688.13131

Clark, J. B., S. Schollaert Uz, and T. Ames. 2024. Non-Euclidean Water Distance Based Interpolation for Increased Mapping of Coastal Water Clarity, Conference: IGARSS 2024 - 2024 IEEE International Geoscience and Remote-sensing Symposium. DOI: 10.1109/IGARSS53475.2024.10642967

Coffer, M.M., D.D. Graybill, P.J. Whitman, B.A. Schaeffer, W.B. Salls, R.C. Zimmerman, V. Hill, M.C. Lebrasse, J. Li, D.J. Keith, J. Kaldy, P. Colarusso, G. Raulerson, D. Ward, and W.J. Kenworthy. 2023. Providing a framework for seagrass mapping in United States coastal ecosystems using high spatial resolution satellite imagery. *Journal of Environmental Management* 337: 117669. DOI: 10.1016/j.jenvman.2023.117669

Cutler, D. R., T.C. Edwards Jr., K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, and J.J. Lawler. 2007. Random forests for classification in ecology. *Ecology* 88:11. DOI: 10.1890/07-0539.1

Elith, J., J.R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *J. Animal Ecology* 77:4, 802-813. DOI: 10.1111/j.1365-2656.2008.01390.x

Fink, D., A. Johnston, M. Strimas-Mackey, T. Auer, W.M. Hochachka, S. Ligocki, L. Oldham Jaromczyk, O. Robinson, C. Wood, S. Kelling, and A.D. Rodewald. 2023. A double machine-learning trend model for citizen science data. *Methods in Ecology and Evolution* 14:9, 2435-2448. DOI: 10.1111/2041-210X.14186

Goetz, S., D. Steinberg, R. Dubayah, and B. Blair. 2007. Laser remote-sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote-sensing of Environment* 108:3, 254-263. DOI: 10.1016/j.rse.2006.11.016

Goodling, P. J., J.H. Fair, A. Gupta, J.D. Walker, T. Dubreuil, M. Hayden, and B.H. Letcher. Technical note: A low-cost approach to monitoring relative streamflow dynamics in small headwater streams using time lapse imagery and a deep learning model, Hydrology and Earth System Sciences, 29, 6445–6460, https://doi.org/10.5194/hess-29-6445-2025, 2025

Hill, R. A., E.W. Fox, S.G. Leibowitz, A.R. Olsen, D.J. Thornbrugh, and M.H. Weber. 2017. Predictive mapping of the biotic condition of conterminous US rivers and streams. *Ecological Applications* 27:8, 2397-2415. DOI: 10.1002/eap.1617

Kratzert, F., D. Klotz, M. Herrnegger, A.K. Sampson, S. Hochreiter, and G.S. Nearing. 2019. Toward improved predictions in ungauged basins: Exploiting the power of machine-learning. *Water Resources Research* 55: 11344–11354. DOI: 10.1029/2019WR026065

Lamb, B., M.A. Tzortziou, and K.C. McDonald. 2021. A Fused Radar–Optical Approach for Mapping Wetlands and Deepwaters of the Mid–Atlantic and Gulf Coast Regions of the United States. *Remote-sensing* 13(13), 2495. DOI: 10.3390/rs13132495

Maloney, K. O., D.E. Weller, M.J. Russell, and T. Hothorn. 2009. Classifying the biological condition of small streams: An example using benthic macroinvertebrates. *Journal of the North American Benthological Society* 28:4, 869-884. DOI: 10.1899/08-142.1

Maloney, K. O., K.P. Krause, M.J. Cashman, W.M. Daniel, B.P. Gressler, D.J. Wieferich, and J. A. Young. 2022a Using fish community and population indicators to assess the biological condition of streams and rivers of the Chesapeake Bay watershed, USA. *Ecological Indicators* 134:108488. DOI: 10.1016/j.ecolind.2021.108488

Maloney, K. O., C. Buchanan, R.D. Jepsen, K.P. Krause, M.J. Cashman, B.P. Gressler, J.A. Young, and M. Schmid. 2022b. Explainable machine-learning improves interpretability in the predictive modeling of biological stream conditions in the Chesapeake Bay Watershed, USA. *J. Environ. Manage* 322:116068. DOI: 10.1016/j.jenvman.2022.116068

McCabe, P. 2019. Habitat modeling of invasive blue catfish in the Patuxent River, Chesapeake Bay. Master's project, Duke University. https://hdl.handle.net/10161/18416

Merriam, E. R., J.T. Petty, and J. Clingerman. 2019. Conservation planning at the intersection of landscape and climate change: Brook trout in the Chesapeake Bay watershed. *Ecosphere* 10:2:e.02585. DOI: 10.1002/ecs2.2585

McNeil, D. J., S.C. Goslee, M. Kammerer, S.E. Lower, J.F. Tooker, and C.M. Grozinger. 2024. Illuminating patterns of firefly abundance using citizen science data and machine-learning models. *Science of the Total Environment* 929:172329. DOI: 10.1016/j.scitotenv.2024.172329

Pahlevan, N., B. Smith, K. Alikas, J. Anstee, C. Barbosa, C. Binding, M. Bresciani, B. Cremella, C. Giardino, D. Gurlin, V. Fernandez, C. Jamet, K. Kangro, M.K. Lehmann, H. Loisel, B. Matsushita, H. Nguyên, L. Olmanson, G. Potvin, S.G.H. Simis, A. VanderWoude, V. Vantrepotte, and A. Ruiz-Verdù. 2022. Simultaneous retrieval of selected optical water-quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. *Remote-sensing of Environment* 270: 112860. DOI: 10.1016/j.rse.2021.112860

Parrish, D., Pianca, C., Friedrichs, C., and Reay, W., 2025, Integrating in situ and remote sensing techniques for enhanced turbidity monitoring in Chesapeake Bay, Virginia: Paper presented at the 2025 National Water Quality Monitoring Conference, Green Bay, Wisconsin, March 10–14, 2025.

Pearl J. 2009. *Causality: models, reasoning, and inference*. 2nd ed. New York (NY): Cambridge University Press.

Phillips, S. J., R.P. Anderson, and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distribution. *Ecological modeling* 190:3-4, 231-259. DOI: 10.1016/j.ecolmodel.2005.03.026

Shen, C. 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour. Res.* 54, 8558-8593. DOI: 10.1029/2018wr022643

Schmid, M., T. Hothorn, K.O. Maloney, and D.E. Weller. 2011. Geoadditive regression modeling of stream biological condition. *Environmental and Ecological Statistics* 18, 709-733. DOI: 10.1007/s10651-010-0158-4

Schollaert Uz, S., T J. Ames, J.B. Clark, and S.L. Smith. 2024. DEEP-VIEW integration of coastal observations and models to inform water-quality resource managers and decisions, Conference: IGARSS 2024 - 2024 IEEE International Geoscience and Remote-sensing Symposium. DOI: 10.1109/IGARSS53475.2024.10642274

Sterle, G., A.A. Harpold, I.U. Haq, J. Perdrial D. Kincaid, and B.S. Lee. 2025. CAMELS-Chem: collection of stream-water chemistry data, 1980–2018. HydroShare. DOI: 10.4211/hs.841f5e85085c423f889ac809c1bed4ac

University of Maryland Center for Environmental Science (UMCES). 2024. *Chesapeake Global Collaboratory.* Cambridge (MD): UMCES, accessed October 21, 2025, at https://www.umces.edu/

Vanhellemont, Q., and K. Ruddick. 2016. ACOLITE for Sentinel-2: Aquatic applications of MSI imagery. In: Proceedings of the 2016 ESA Living Planet Symposium. ESA Special Publication, SP, pp. 740.

Virginia Estuarine & Coastal Observing System, 2025, VECOS (Virginia Estuarine & Coastal Observing System) home page: Virginia Institute of Marine Science, accessed October 21, 2025, at https://vecos.vims.edu/home

Woods, T., M.C. Freeman, K.P. Krause, and K.O. Maloney. 2023. Observed and projected functional reorganization of riverine fish assemblages from global change. *Global Change Biology* 29:13, 3759-3780. DOI: 10.1111/gcb.16707

Weinhold, L., M. Schmid, R. Mitchell, K.O. Maloney, M.N. Wright, and M.Berger. 2020. A random forest approach for bounded outcome variables. *J. Computational and Graphical Statistics*. DOI: 10.1080/1061800.2019.1705310

Xu, T. and F. Liang. 2021. Machine-learning for hydrologic sciences: An introductory overview. *WIREs Water* 8:e1533. DOI: 10.1002/wat2.1533

Yu, X. and J. Shen. 2021. A data-driven approach to simulate the spatiotemporal variations of chlorophyll-a in Chesapeake Bay. *Ocean Model. Online* 159:101748. DOI: 10.1016/j.ocemod.2020.101748

Yu, X., J. Shen, G. Zheng, and J. Du. 2022. Chlorophyll-a in Chesapeake Bay based on VIIRS satellite data: Spatiotemporal variability and prediction with machine-learning. *Journal of Ocean Modeling* 180: 102119. DOI: 10.1016/j.ocemod.2022.102119

Yu, X., J. Shen, and J. Du. 2020. A machine-learning-based model for water-quality in coastal waters, taking dissolved oxygen and hypoxia in Chesapeake Bay as an example. *Water Resour. Res.* 56:e2020WR027227. DOI: 10.1029/2020wr027227

Zhang, Q., T.R. Fisher, C. Buchanan, A.B. Gustafson, R.R. Karrh, R.R. Murphy, J.M. Testa, R. Tian, and P.J. Tango. 2022. Nutrient limitation of phytoplankton in three tributaries of Chesapeake Bay: Detecting responses following nutrient reductions. *Water Res.* 226:119099. DOI: 10.1016/j.watres.2022.119099

Zhang, Q., R. R. Murphy, R. Tian, and P. J. Tango. 2025. Geography, trajectories, and controls of coastal water-quality: More rapid improvement in the shallow zone of the Chesapeake Bay. *Environ. Sci. Technol.* 59, 553-564. DOI: 10.1021/acs.est.4c07368

Zheng, G., S. Schollaert Uz, P. St-Laurent, M.A.M. Friedrichs, A. Mehta, and P.M. DiGiacomo. 2024. Hypoxia Forecasting for Chesapeake Bay Using Artificial Intelligence. *Artificial Intelligence for the Earth Systems* 3(3). DOI: 10.1175/AIES-D-23-0054.1

APPENDIX A: Workshop Agenda

Participant List:	*Mentimeter:*	*Feedback Survey:*
bit.ly/STACAIparticipants	bit.ly/STACAImentimeter	bit.ly/STACAIfeedback

Chesapeake Bay Program's (CBP)
Scientific and Technical Advisory Committee (STAC)
Workshop
Leveraging Artificial Intelligence and Machine Learning to Advance Chesapeake Bay Research and Management
February 24-25, 2025
Smithsonian Environmental Research Center
Mathias 1013
Workshop webpage

****Exact Times Are Subject to Change****
This meeting will be recorded to assure the accuracy of meeting notes.

This STAC workshop is aimed at providing a unique opportunity for researchers and managers to gather and review the state of the science on AI/ML approaches and their potential to advance Chesapeake Bay research and management. The workshop will focus on three main objectives:

I. Summarize recent AI/ML applications to the Chesapeake Bay ecosystem and lessons learned;
II. Identify the challenges and gaps in applying AI/ML approaches to Chesapeake Bay data; and
III. Develop recommendations and identify opportunities for harnessing the power of AI/ML approaches to address Chesapeake Bay issues.

Monday, February 24, 2025

9:00 am **Coffee & Light Breakfast (Provided)**

9:45 am **Welcome and Introductions** – *Meg Cole (CRC)*
STAC Coordinator, Meg Cole, will outline the workshop logistics and facilitate brief introductions from all participants.

10:00 am **Workshop Overview and Motivations** – *Qian Zhang (UMCES)*
Workshop Chair, Qian Zhang, will provide context for the workshop, including its purpose, objectives, and key motivations.

I: Summarize recent AI/ML applications to the Chesapeake Bay ecosystem and lessons learned

This session will synthesize recent applications of artificial intelligence and machine learning (AI/ML) within the Chesapeake Bay region, including both tidal and nontidal areas. The discussion will focus on the objectives of each study, the rationale for selecting specific AI/ML approaches, the new insights generated through these methods, and how these findings have (or could) inform restoration efforts in alignment with the Chesapeake Bay Watershed Agreement goals and outcomes.

10:15 am **Introductory Overview of AI and ML** – *Alison Appling (USGS)*
Alison Appling will provide foundational insights into artificial intelligence (AI) and machine learning (ML), including key definitions, a taxonomy of AI approaches, and an introduction to several common methods.

10:45 am **Overview of Chesapeake Bay Restoration: CBP Goals & Outcomes** – *Gary Shenk (USGS)*
Gary Shenk will discuss the 10 key goals identified by the Chesapeake Bay Program (CBP) for ecosystem restoration. While AI is already being used to generate information supporting these goals, significant opportunities remain for further application.

11:15 am **15-minute Break**

11:30 am **Literature Summary of Watershed and Living Resources Studies Involving AI/ML** – *Kim Van Meter (PSU) and Kelly Maloney (USGS)*
Kim Van Meter and Kelly Maloney will present a summary of existing literature on AI/ML applications in watershed and living resources studies, highlighting key findings and trends.

12:00 pm **Literature Summary of Estuarine and Living Resources Studies Involving AI/ML** – *Jian Shen (VIMS) and Stephanie Schollaert Uz (NASA)*
Jian Shen will present a general overview of recent literature on estuarine and living resources studies involving AI/ML. Following this, Stephanie Schollaert Uz will discuss *AI/ML Integration of Satellite Remote Sensing: Data Harmonization Challenges and Gaps*, focusing on challenges and gaps in data harmonization.

12:30 pm **Lunch (Provided)**

II: Identify the challenges and gaps in applying AI/ML approaches to Chesapeake Bay data

This session will explore key challenges and gaps in applying AI/ML to Chesapeake Bay data, including data limitations such as lack of data or issues with harmonization, insufficient expertise in AI/ML algorithms and methodologies, unavailability of software code for replication or adaptation, communication barriers in effectively using or explaining AI/ML-generated insights (e.g., explainable ML), and coordination gaps among research and management institutions within the Chesapeake Bay Program (CBP) Partnership.

1:30 pm **Introduce Lightning Talk Speakers, Open Mentimeter, and 'Office Hours' Structure** – *Meg Cole (CRC)*
Cole will introduce the lightning talk speakers, launch the Mentimeter interactive platform for audience engagement, and outline the structure for the upcoming 'Office Hours' poster/session, designed to facilitate focused discussions and collaboration.

1:45 pm **Lightning Talks with Q&A (Round 1)**
Eight speakers have been invited to share their recent work through concise, 7-minute presentations. This session will be divided by a 15-minute break following the fourth speaker. Participants are requested to save questions for the Poster Session at 2:40pm.

- Patrick Bitterman (Kent) – GeoAI and Social Systems Modeling
- Mike Evans (Conservancy) – Integrated AI models to forecast land use change
- Shuyu Chang (PSU) – Advances in water quality predictions: datasets and learning frameworks
- David Parish (VIMS) – Modeling Light Conditions in the York River Estuary by Anchoring Satellite Imagery with High-Frequency In-Situ Observations

2:05 pm **15-minute Break**

2:20 pm **Lightning Talks with Q&A (Round 2)**

- Matthew Cashman (USGS) – Physical habitat is more than a sediment issue: A multi-dimensional habitat assessment indicates new approaches for river management
- Taylor Woods (USGS) – Observed and projected functional reorganization of riverine fish assemblages from global change
- Jenn Fair (USGS) – Images to Info: the USGS Flow Photo Explorer
- Sean Emmons (USGS) – Leveraging machine learning and expert knowledge to unravel the complexities of multiple freshwater ecosystem stressors

2:40 pm **Lightning Talk 'Office Hours' and Workshop Poster Session**
Lightning talk speakers and participants are invited to showcase their work during a poster session in the Atrium of the Mathias Lab.

4:00 pm **Wrap-Up and Objectives of Day 2** – *Qian Zhang (UMCES)*
Qian Zhang will provide a summary of the day's discussions and outline the objectives and key focus areas for Day 2 of the workshop.

4:30 pm **Happy Hour (Optional)**
Join us for refreshments and snacks in the Atrium, following the Office Hours and Poster Session, to unwind and network with fellow participants.

4:30 pm **Day 1 Recess**

6:00 pm **Dinner Off Campus (Optional)**
Participants interested in attending a group dinner are invited to meet in the Mathias Lab Atrium at 6:00pm to depart together.

Tuesday, February 25, 2025

8:30 am **Coffee & Light Breakfast (Provided)**

9:00 am **Review of Day 1; Objectives for Day 2** – *Steering Committee Members*

III: Develop recommendations and identify opportunities for harnessing the power of AI/ML approaches to address Chesapeake Bay issues

This session will focus on developing actionable recommendations and identifying opportunities to leverage AI/ML approaches for Chesapeake Bay restoration. Key areas of discussion include identifying where the Chesapeake Bay Program (CBP) Partnership can benefit most from AI/ML, brainstorming how AI/ML can generate new insights to support restoration efforts, and formulating strategies to deliver AI/ML-generated information to watershed managers in an efficient, understandable, and actionable manner. Additionally, the session will explore guidelines for standardizing and streamlining the selection and use of AI/ML approaches for analyzing monitoring data, as well as proposing ways to enhance collaboration and synergies among stakeholders within the CBP Partnership.

9:30 am **State-of-the-Art AI & Physics-Informed ML in Hydrology and Water Quality: Insights and synergies** – *Chaopeng Shen (PSU)*
Chaopeng Shen will present on the evolution of the AI/ML field in the context of

watershed and estuarine sciences, including emerging directions and opportunities, and the communication and explanation of results (e.g. xAI).

10:00 am **15-minute Break**

10:15 am **Panel: AI/ML Community Development**
– *Dong Liang (UMCES), Chaopeng Shen (PSU), Vandana Janeja (UMBC), Kelly Maloney (USGS), Robert Sabo (EPA), Alison Appling (USGS)*
Moderated by steering committee member *Matt Baker (UMBC)*, this panel will explore strategies for fostering synergies, breaking down barriers, and establishing pathways for ongoing dialogue and collaboration within the AI/ML community.

11:15 am **Breakout Sessions**
Participants will move into small groups for further discussion. Group assignments are random, but each breakout will include a facilitating steering committee member(s), a notetaker, and individuals involved in the CBP partnership.

12:30 pm **Lunch (Provided)**

1:30 pm **Breakout Groups Report-out**
Assigned steering committee members will present key insights and outcomes from their respective breakout group discussions.

2:15 pm **Plenary: Prioritization of High-Level Recommendations**
The steering committee will lead a discussion where workshop participants will refine and identify the highest-priority recommendations emerging from their small group discussions.

3:15 pm **Workshop Adjourns**
Steering Committee Meets

APPENDIX B: Workshop Participants

Name	Affiliation
Alison Appling	USGS
Matt Baker	UMBC, STAC
Isabella Bertani	UMCES
Gopal Bhatt	PSU
Patrick Bitterman	Kent State Uni
Jun Suk Byun	UMCES
Matthew Cashman	USGS
Shuyu Chang	PSU
Peter Claggett	USGS
Joseph Delesantro	EPA CBPO
Bill Dennison	UMCES, STAC
Gabriel Duran	CRC
Andrew Elmore	UMCES
Sean Emmons	USGS
Michael Evans	Chesapeake Conservancy
Jenn Fair	USGS
KC Filippino	HRPDC, STAC
Burch Fisher	UMCES
Sophia Grossweiler	MDE
Xiaoxu Guo	UMCES
Scott Heidel	PA DEP
Admin Husic	VT
Vandana Janeja	UMBC
Jared Kroh	HRG, Inc.
Brooke Landry	MD DNR
Erin Letavic	HRG, Inc., STAC
Dong Liang	UMCES
Lew Linker	EPA CBPO
Vyacheslav Lyubchich	UMCES

Name	Affiliation
Kelly Maloney	USGS
Sarah McDonald	USGS
Bob Murphy	Tetra Tech
Rebecca Murphy	UMCES
George Onyullo	DC DOEE
David Parrish	VIMS
Xueting Pu	PSU
Julie Reichert-Nguyen	NOAA
Robert Sabo	EPA
Sheila Saia	Tetra Tech
Stephanie Schollaert Uz	NASA Goddard
Chaopeng Shen	PSU
Jian Shen	VIMS
Gary Shenk	USGS
Yalan Song	PSU
Melissa Stefun	MDE
Breck Sullivan	USGS
Peter Tango	USGS
Richard Tian	UMCES
Kim Van Meter	PSU
Denice Wardrop	CRC, STAC
Jimmy Webber	USGS
Allison Welch	CRC
Joe Wood	CBF, STAC
Ryan Woodland	UMCES
Taylor Woods	USGS
Qian Zhang	UMCES
Jian Zhao	UMCES

APPENDIX C: List of Figures and Tables

APPENDIX D: Breakout Group Responses by Group

During the workshop, breakout groups addressed two key questions under each of the three objectives below. These discussions helped identify opportunities, challenges, and actionable steps for leveraging AI/ML in Chesapeake Bay restoration efforts. The responses informed the workshop's summary of recommendations for advancing AI/ML integration in Chesapeake Bay initiatives. Participants were split into in-person breakout groups and one virtual breakout group.

Below are all responses under each breakout question, across all breakout groups.

1. **Objective 1: AI/ML Applications and Lessons Learned**
    a. How can AI/ML approaches be leveraged (or have been used) to address issues in the context of the Chesapeake Bay restoration?
        - Track BMPs on croplands in PA, ID area to recalibrate LU layer (PA); evaluate performance of BMPs, inform swimming advisories, USGS Flow Photo Explorer Tool for monitoring streamflow dynamics in small systems + stormwater BMPs, predictive modeling for habitat condition
        - AI/ML has already been applied across various domains relevant to the Bay Agreement
        - AI/ML is highly effective in leveraging multi-scale data sources, including satellite imagery, in-situ monitoring, and high-frequency measurements
        - AI/ML can be used to bring in non-Chesapeake Bay data
        - AI/ML can be used to prepare data for intermediate steps (rather than final step)
        - AI/ML can be used to develop large, tempo-spatial models for water-quality (N, P, S)
        - AI/ML can be used to incorporate satellite remote-sensing data for bridging monitoring gaps, particularly those related to WQS indicators
        - Multi-scale
        - As data preparation, to better synthesize and represent data - step in the process rather than final result
        - Create synergy between two or more different types of data
        - Better capture/quantify processes
        - Developing water-quality spatial-temporal model
        - Understanding best model to use
        - Broad, national-scale models can be refined for Chesapeake Bay, which helps avoid overfitting and improves learning beyond the watershed boundary
        - AI/ML benefits from exposure to diverse datasets, with base dynamics carrying over across systems. Data from outside the Bay can reinforce and refine models developed for local use
        - Smaller-scale modeling is still valuable to capture variability that gets averaged out in broader datasets
        - AI/ML can be designed to act as a user-facing tool (similar to a "front desk" or

chatbot) that retrieves answers, connects users to the best model, and synthesizes complex information into simple responses
- Lots of current and potential examples
- Effect of management practices
- Find AI fixes for some of the problems identified

b. What are some of the advantages and disadvantages of AI/ML compared to other established approaches?
    - Compared to process-based approaches, AI/ML algorithms are generally easier to learn, adopt, and implement. However, it is not to replace traditional approaches
    - AI/ML models are flexible, data-driven, and do not require much domain knowledge, and are gaining momentum and public perception/awareness
    - Explainability of AI/ML results to management/stakeholders is a major issue
    - Uncertainty quantification of AI/ML is lagging traditional statistical approaches
    - Advantages: flexible at learning, high performance and predictability public popularity, communication tool, better able to handle multiple variables, unconventional uses, identifying oddballs, point sources
    - Disadvantages: uncertainty quantification, model hallucinations, being able to justify, address management/stakeholder concerns
    - AI/ML can handle large, diverse datasets and create connections across different domains, offering flexibility and predictive power
    - Too much diversity in training data can dilute outputs; moderate diversity is often more effective
    - Small-scale heterogeneity may be lost in broader datasets, raising questions about when uniformity is important versus when finer detail adds value
    - AI/ML outputs may be shallow compared to domain expertise, with limited depth in specialized questions
    - Advantage - take large data and boil down to make conclusion, scalable, portable (robust), ability to capture dynamics (pattern/trend detection), and non-linear complicated interactions
    - Disadvantage - demand can be high in set-up, at the local-level need more improved data sets; black box; can infer information about variables that isn't there; may behave in ways that make it difficult to evaluate; high concept but may not be understandable at local levels

2. **Objective 2: Challenges and Gaps in AI/ML Implementation**

a. What challenges or gaps have you encountered when applying AI/ML in the context of Chesapeake Bay (or elsewhere)?
    - Data cleanliness; model inception - incorporate models into models, can create compounded errors (not unique to AI); calculating and articulating uncertainty - not

sure how to *show uncertainty; estimation under different scenarios - will be dependent on variables used in model training and can be unstable. Learn right thing for wrong reason, teasing out 'why'

- Availability of data, even for the Chesapeake which is considered a data-rich system
- Availability of features data matched to response data at consistent spatial and temporal scales
- Lack of standardization in selecting AI/ML methods, requiring a balance between model accuracy and interpretability
- Response variables are not matched on a regular basis
- Need more monitoring/data collection to run models
- Outside datasets are not always "AI-ready," requiring additional preparation and alignment
- Hydrologic data can vary by scale; small-system heterogeneity may not be captured in regional datasets
- Watershed divides and boundaries shift in newer datasets, creating mismatches with existing models
- Missing predictor and response variables limit model performance, while inconsistencies in coding and uniformity remain an issue
- Social systems behave unpredictably compared to physical processes, making it harder to define predictors and responses

b. What have you done (or may be done) to address the challenges and gaps?

- Allocate resources for data harmonization of input data (e.g., the USGS NTN concentration data) and make those data available with data releases (including metadata)
- Advance the Data Hub for Chesapeake Bay researchers and managers by including both monitoring data and modeling data (e.g., CAST, temperature, point source), and making those data available to users
- Promote/allocate resources to data harmonization
- Curating and unifying datasets can improve both predictor and response variables, with potential for additional requirements to standardize inputs
- Tools can be designed to generate watershed boundaries directly from maps, simplifying data extraction for users
- Customized AI systems curated to Chesapeake Bay assets could summarize existing information and guide users to the right datasets or models
- Dialogue between data scientists and domain experts is essential for defining necessary predictor variables and ensuring models have depth
- Gaps in predictors: can drop the sites, enable users to flag data
- Use off-the-shelf pre-built image screening models

- Effective ways to use meta-data (potentially through generative AI)
- (maybe) Causal deep learning techniques

3. **Objective 3: Recommendations and Future Opportunities**
   a. What are the biggest barriers preventing broader AI/ML adoption in Chesapeake research and management?
      - ML should be paired with data visualization
      - Takes time and expertise to effectively communicate across groups; silos across fields (i.e., medical models)
      - What is meant by AI can be unclear and scary! Need to define what AI is and how it is being used in the context; should popularize its use (i.e., educate students, vocational training programs); social/human side to implementation
      - Data availability for AI/ML
      - Disciplinary barriers between AI/ML researchers and process-based model developers hinder integration and collaboration
      - Engage with CBP workgroups and the Water Quality GIT early and consistently to align AI/ML applications with their priorities and goals, and keep the decision makers in mind
      - Enhance communication of AI/ML findings to managers and decision makers to support informed policy and management actions (e.g., xAI approaches)
      - Data availability and utility
      - Communication between people (modelers, statisticians, decision makers, etc.)
      - Expertise
      - Engagement strategies
      - High ambitions for the models
      - Lack of alignment between available data and model needs (predictors vs. responses, spatial vs. temporal scales)
      - Limited ability of generalized AI tools to provide deep, domain-specific insights
      - Social science variables are harder to model because of unpredictable human behavior and rapid changes in system representation
      - Data preparation and labeling need to improve for AI to be fully effective
      - Acceptability to stakeholders
        - Interpretability => use interpretable methods
        - Newness => speed of development facilities stakeholder interaction
      - Data
        - Uneven temporal and spatial distribution
        - Lack of problem-relevant data
        - Unclear ontology and varying quality
        - May be an AI solution to these problems

    - Human language queries
    - Database standardization monitoring design

b. What forums, workshops, or working groups may be established to foster collaborations and discussions among different groups of AI/ML researchers as well as between them and Bay scientists and managers?
    - Establish an informal workgroup (perhaps similar to the Integrated Trends Analysis Team (ITAT) and Factors Team) to continue the conversation (e.g., Ches-BRAIN: Chesapeake Bay Research with Artificial Intelligence and Networking), which may be led by CBP or USGS
    - Strengthen collaborations across AI/ML research groups and resource managers (including HydroML, CGC, USGS, etc), to drive interdisciplinary advancements
    - Promote community engagement of AI/ML (e.g., community challenge on identifying response variables, etc)
    - Collaborate with Factors Team (USGS)
    - Create Bay Program Machine-learning working group (Ches-BRAIN)
    - Advertise in journals
    - Create forum for community engagement (e.g., suggest response variable)
    - A Chesapeake-focused catalogue or "librarian" system could help managers and researchers quickly identify relevant datasets and models
    - Curated AI/ML interfaces could provide both simple answers for managers and technical detail for researchers, serving as a shared resource
    - Expanding data cataloguing efforts to include metadata, attribution, and raw data would improve usability for AI/ML applications
    - Community engagement efforts could help prioritize response variables and strengthen collaboration between model developers and CBP stakeholders
    - Look for scientific discussion opportunities: CCRS, HydroML, Ches-BRAIN, STAC workshops
    - CBP GITs/WGs - e.g., introduce USGS Flow Photo Explorer to relevant groups
    - Conferences: CCMP (Chesapeake Community Research Symposium), Association of Mid-Atlantic Aquatic Biologists, HydroML

APPENDIX E: Literature Review Annotated Bibliography

The literature review compiled by the steering committee, consisting of 74 entries, focuses on the application of artificial intelligence (AI) and machine-learning (ML) techniques to study and manage environmental and ecological aspects of the Chesapeake Bay and its surrounding watershed. The studies span a range of topics, including water-quality, coastal ecosystems, microbial communities, and species populations, with a strong emphasis on leveraging remote-sensing, satellite data, and in-situ measurements. Techniques such as neural networks, random forests, clustering, and regression models are commonly employed to predict variables like DO, nutrient concentrations, chlorophyll *a* levels, salinity, and total suspended solids. Many of these works aim to enhance monitoring and forecasting capabilities, offering insights into how environmental factors (i.e., land-use, climate, and hydrology) interact with biological and physical processes in the Bay.

A significant portion of the review highlights efforts to address practical challenges in the Chesapeake Bay region, such as hypoxia, coastal flooding, and habitat conservation. For instance, studies explore the use of deep learning to model wave spectra, predict storm surges, and map wetlands or seagrass density, often integrating diverse data sources like Sentinel-2 imagery, MODIS satellite data, and lidar surveys. Other research focuses on biological conditions, using AI to assess fish assemblages, microbial populations, and phytoplankton production, or to identify drivers of species abundance and biodiversity. Collectively, the literature review demonstrates a growing reliance on data-driven approaches to inform resource management, conservation planning, and policy decisions, reflecting the interdisciplinary nature of tackling environmental issues in this critical estuarine system.

A screenshot of the Literature Review Sheet is on the following page, and can be downloaded here. Each article was cataloged by reference information, author affiliations, year of publication, journal, study focus, AI/ML methods used, and a short description drawn from the abstract. The compiled dataset provides a baseline resource for understanding the scope of AI/ML applications to water-quality, ecology, and related environmental challenges in the Chesapeake Bay watershed.

No	Reference	First Author	First-author Institution on	Coauthors	Year	Journal	Title	Category	AI/ML Methods	Description (based on abstract)
1	Ackleson, S. G.	Ackleson, S. G.	Bigelow Laboratory for Ocean	V. Klemas	1987	Remote Sens. Environ.	Remote sensing of submerged aquatic vegetation in	SAV	Clustering	Detecting SAV using remote sensing data
2	Kim, G. E., M. S	Kim, G. E.	Booz Allen Hamilton	M. Steller, S. Olson and	2020	10th International	Modeling watershed nutrient concentrations with	Watershed nutrient	H2O AutoML	Implementing an AutoML pipeline to find the best model and associated parameters for
3	Ioannou, I., A. (	Ioannou, I.	City College of New York	A. Gilerson, M.	2014	Conference on Ocean	Algorithms for the remote estimation of chlorophyll-a in	Coastal water quality	Neural networks	Remote estimation of chlorophyll-a concentration [Chl-a] in the Chesapeake Bay from
4	Ioannou, I., A. (	Ioannou, I.	City College of New York	A. Gilerson, M.	2014	Conference on Remote	Remote estimation of in water constituents in coastal	Coastal water quality	Neural networks	Remote estimations of oceanic constituents from optical reflectance spectra in coastal
5	Yang, C., D. Mil	Yang, C.	Florida International	D. Mills, K. Mathee, Y.	2006	J. Microbiol. Methods	An eco-informatics tool for microbial community	Microbial community	Support vector machines (SVM); K-	Profiling microbial community patterns - both SVM and KNN classifiers were effective
6	Meister, M. and	Meister, M.	George Mason University	J. J. Qu	2024	Remote Sensing	Quantifying Seagrass Density Using Sentinel-2 Data and	Seagrass	Naive Bayes (NB); CART; Support	Quantifying seagrass density using Sentinel-2 satellite data and machine learning in
7	Evans, M. J., K.	Evans, M. J.	George Mason University /	K. Mainali, R. Soobitsky,	2023	Biol. Conserv.	Predicting patterns of solar energy buildout to identify	Solar arrays	Neural networks	Identifying drivers of solar siting and quantified patterns of buildout in states
8	DeLuca, N., B. Z	DeLuca, N.	Johns Hopkins University	B. Zaitchik Zaitchik and	2018	Remote Sensing	Can multispectral information improve remotely sensed	Coastal TSS	Random forest	Investigating whether utilizing additional wavelengths from the Moderate Resolution
9	DeLuca, N. M.,	DeLuca, N. M.	Johns Hopkins University	B. F. Zaitchik, S. D.	2020	Remote Sens. Environ.	Evaluation of remotely sensed prediction and forecast	Vibrio parahaemolyticus	GLM, GAM, etc	Using satellite ocean color remote sensing and sea surface temperature (SST) from the
10	Arora-Williams,	Arora-Williams,	Johns Hopkins University	C. Holder, M. Secor, H.	2022	Environ. Microbiol.	Abundant and persistent sulfur-oxidizing microbial	Microbial populations	Clustering	Classifying distribution of microbial populations in low-oxygen environments of the Bay
11	Urquhart, E. A.,	Urquhart, E. A.	Johns Hopkins University	B. F. Zaitchik, M. J.	2012	Remote Sens. Environ.	Remotely sensed estimates of surface salinity in the	Coastal salinity	CART; GLM; GAM; Random forest;	Using multiple statistical models to predict daily, gridded surface salinity at 1 km
12	Scardi, M., 199	Scardi, M.	Laboratorio di Oceanografia	None	1996	Mar. Ecol. Prog. Ser.	Artificial neural networks as empirical models for	Phytoplankton	Neural networks; regression	Estimating phytoplankton production using regression approaches and artificial neural
13	He, Y., B. Bond-	He, Y.	Lawrence Berkeley National	B. Bond-Lamberty, A. N.	2024	Heliyon	Effects of spatial variability in vegetation phenology,	Soil respiration	Clustering; Random forest; Shapley	Understanding the primary driving factors affecting soil respiration within sub-
14	Balasubramani	Balasubramania	NASA	N. Pahlevan, B. Smith,	2020	Remote Sens. Environ.	Robust algorithm for estimating total suspended solids	Coastal TSS	Mixture Density Network (MDN)	Presenting a Statistical, inherent Optical property (IOP)-based, and muLti-conditional
15	Memarsadeghi,	Memarsadeghi,	NASA	S. Schollaert Uz, J. R.	2022	International Geoscience	In situ water quality data for the Chesapeake Bay	Coastal water quality	Deep learning (no access to full text)	Developing new methods for water quality monitoring in the Chesapeake Bay from
16	Schollaert Uz, S	Schollaert Uz, S.	NASA	T. J. Ames, N.	2020	IEEE International	Supporting aquaculture in the Chesapeake Bay using	Coastal water quality	Neural networks	Artificial Intelligence trained with simultaneous in situ and satellite observations is
17	Ross, A. C. and	Ross, A. C.	NOAA	C. A. Stock	2019	Estuar. Coast. Shelf Sci.	An assessment of the predictability of column minimum	Coastal DO	Regression tree	Using a machine learning model to assess the predictability of column minimum
18	Daniels, W., T. A	Daniels, W.	Northwestern University /	T. Ames, J. B. Clark, S.	2023	IEEE International	Improving extreme value prediction for water clarity using	Water clarity	Weighted regression (no access to full	Using a recently published weighted regression approach to alleviate "mean-centric" bias
19	Wang, N., Q. Ch	Wang, N.	Northeastern University	Q. Chen and L. Zhu	2023	Applied Ocean Research	Data-driven modeling of Bay-Ocean wave spectra at	Bay-Ocean wave spectra	Neural networks	Providing accurate forecasts and hindcasts of wave conditions for the Bay Bridge-Tunnel
20	Shahabi, A. and	Shahabi, A.	Old Dominion University	N. Tahvildari	2024	Coastal Engineering	A deep-learning model for rapid spatiotemporal	Coastal water level	LSTM; CNN	Developing a deep neural network for spatiotemporal prediction of water levels in
21	Wells, B. K., S.	Wells, B. K.	Old Dominion University	S. R. Thorrold and C. M.	2000	Trans. Am. Fish. Soc	Geographic Variation in Trace Element Composition of	Trace Element	Neural networks; Linear discriminant	Determining trace metal levels in juvenile weakfish from five estuarine locations
22	Bangley, C. W.,	Bangley, C. W.	SERC	M. L. Edwards, C.	2021	Ecosphere	Environmental associations of cownose ray (Rhinoptera	Cownose ray	Boosted regression trees (BRTs)	Better understanding the drivers and timing of cownose ray seasonal migration in order
23	Byrd, K. B., L. B	Byrd, K. B.	SERC	L. Ballanti, N. Thomas,	2018	ISPRS J. Photogramm.	A remote sensing-based model of tidal marsh	Tidal marsh	Random forest	Generating a single remote sensing model of tidal marsh aboveground biomass and
24	Scardi, M. and L	Scardi, M.	Stazione Zoologica 'A.	L. W. Harding	1999	Ecol. Model.	Developing an empirical model of phytoplankton	Phytoplankton primary	Neural networks; regression	Estimating phytoplankton primary production using regression and artificial neural
25	Bailey, H., A. D.	Bailey, H.	UMCES	A. D. Fandel, K. Silva, E.	2021	Ecosphere	Identifying and predicting occurrence and abundance of	Vocal animal species	GAM	Analyzing acoustic recordings from 2016 to 2018 for signature whistles of bottlenose
26	Mulligan, T. J. a	Mulligan, T. J.	UMCES	R. W. Chapman	1989	Copeia	Mitochondrial DNA Analysis of Chesapeake Bay White	Mitochondrial DN	Clustering	Understanding white perch population and migration
27	Testa, J. M., V. L	Testa, J. M.	UMCES	V. Lyubchich and Q.	2019	Estuaries Coasts	Patterns and trends in secchi disk depth over three	Coastal Secchi	Clustering	Understanding the factors driving long-term variability and trends in water clarity (i.e.,
28	Testa, J. M., W.	Testa, J. M.	UMCES	W. Liu, W. R. Boynton,	2024	Estuaries Coasts	Physical and Biological Controls on Short-Term	Shallow water DO	CART	Using a combination of time-series analysis, harmonic analysis, and machine learning
29	Windle, A. E., L	Windle, A. E.	UMCES	L. W. Staver, A. J.	2023	Frontiers in Remote	Multi-temporal high-resolution marsh vegetation	Marsh vegetation	Random forest	Classifying species-specific marsh vegetation using UAS remote sensing and random
30	Zhang, Q., T. R.	Zhang, Q.	UMCES / CBP	T. R. Fisher, E. M.	2021	Water Res.	Nutrient limitation of phytoplankton in Chesapeake Bay:	Coastal nutrient	CART	Reproducing bioassay-based nutrient limitation patterns in the mainstem of the Bay and
31	Zhang, Q., T. R.	Zhang, Q.	UMCES / CBP	T. R. Fisher, C.	2022	Water Res.	Nutrient limitation of phytoplankton in three tributaries	Coastal nutrient	CART	Reproducing bioassay-based nutrient limitation patterns in three tidal tributaries and
32	Zhang, Q., J. T. I	Zhang, Q.	UMCES / CBP	J. T. Bostic and R. D.	2022	Water Res.	Regional patterns and drivers of total nitrogen trends in	Riverine nitrogen	Clustering; Random forest	Understanding regional patterns and drivers of total nitrogen trends in the Chesapeake
33	Zhang, Q., J. T. I	Zhang, Q.	UMCES / CBP	J. T. Bostic and R. D.	2023	Environ. Res. Lett.	Effects of point and nonpoint source controls on total	Riverine phosphorus	Clustering; Random forest	Understanding regional patterns and drivers of total phosphorus trends in the
34	Pluchino, A., A.	Pluchino, A.	University of Catania,	A. Rapisarda and V.	2008	European Physical	Communities recognition in the Chesapeake Bay	Communities	Dynamical clustering	Applying the dynamical clustering to the identification of communities of marine
35	Langendorf, R.	Langendorf, R.	University of Colorado,	V. Lyubchich, J. M. Testa	2021	ACS ES&T Water	Inferring controls on dissolved oxygen criterion	Coastal DO	Clustering; Structural equation	Understanding the long-term fluctuations of DO in the Bay in response to external
36	Austin, B., D. A.	Austin, B.	University of Maryland,	D. A. Allen, A. L. Mills	1977	Can. J. Microbiol.	Numerical taxonomy of heavy metal-tolerant bacteria	Microbial taxonomy	Clustering	Classifying metal-tolerant bacteria from Chesapeake Bay samples
37	Austin, B., J. J. C	Austin, B.	University of Maryland,	J. J. Calomiris, J. D.	1977	Appl. Environ. Microbiol.	Numerical taxonomy and ecology of petroleum-	Microbial taxonomy	Clustering	Classifying petroleum-degrading bacteria from Chesapeake Bay samples
38	DeSilet, L., B. G	DeSilet, L.	University of Maryland,	B. Golden, Q. Wang and	1992	Computers & Operations	Predicting salinity in the Chesapeake Bay using	Coastal salinity	Neural networks; regression	Predicting salinity in different parts of the Bay using two approaches -- neural network
39	Mallory, L. M., E	Mallory, L. M.	University of Maryland,	B. Austin and R. R.	1977	Can. J. Microbiol.	Numerical taxonomy and ecology of oligotrophic	Microbial taxonomy	Clustering	Classifying slow-growing bacteria from Chesapeake Bay samples
40	West, P. A., G. C	West, P. A.	University of Maryland,	G. C. Okpokwasili, P. R.	1984	Appl. Environ. Microbiol.	Numerical taxonomy of phenanthrene-degrading	Microbial taxonomy	Clustering	Classifying phenanthrene-degrading bacteria from Chesapeake Bay samples
41	Islambekov, U.	Islambekov, U.	University of Texas at Dallas	Y. R. Gel	2018	Environmetrics	Unsupervised space–time clustering using persistent	Water quality	Clustering	Presenting a new clustering algorithm for space-time data based on the concepts of
42	Huang, X., I. R.	Huang, X.	University of Texas at Dallas /	I. R. Iliev, V. Lyubchich	2017	Environmetrics	Riding down the Bay: Space-time clustering of ecological	Water quality	Clustering	Developing a new data-driven procedure for optimal selection of tuning parameters in
43	Kerns, B. W. an	Kerns, B. W.	University of Washington	S. S. Chen	2022	Nat. Hazards	Compound effects of rain, storm surge, and river	Coastal flooding	K-Nearest Neighbor (KNN)?	Better understanding and improving prediction of the compound effects of rain, storm
44	Jellen, C., J. Bu	Jellen, C.	US Naval Academy	J. Burkhardt, C.	2020	Appl. Opt.	Machine learning informed predictor importance	Maritime optical	Random forest	Deriving new insights into the physical relationships affecting optical turbulence in the
45	Kim, J.-W. and Y	Kim, J.-W.	USDA-ARS	Y. A. Pachepsky	2010	J. Hydrology	Reconstructing missing daily precipitation data using	Precipitation	Neural networks	Developing a new technique to reconstruct missing daily precipitation data in the central
46	Lee, J., A. Abbas	Lee, J.	Ulsan National Institute of	A. Abbas, G. W.	2023	J. Hydrology	Estimation of base and surface flow using deep neural	Flow discharge	LSTM	Developing a deep learning model which is capable of simulating both base and surface
47	Maloney, K. O.,	Maloney, K. O.	USGS	Z. M. Smith, C.	2018	Freshwater Science	Predicting biological conditions for small headwater	Stream biological	Random forest	Developing a random forests model to predict biological condition of small streams
48	Maloney, K. O.,	Maloney, K. O.	USGS	D. M. Carlisle, C.	2021	Environ. Manage.	Linking altered flow regimes to biological condition: An	Stream biological	Random forest	Developing random forests to model biological conditions using a benthic
49	Maloney, K. O.,	Maloney, K. O.	USGS	C. Buchanan, R. D.	2022	J. Environ. Manage.	Explainable machine learning improves interpretability in	Stream biological	Random forest	Developing separate random-forest models to predict flow status (inflated, diminished,
50	Maloney, K. O.,	Maloney, K. O.	USGS	K. P. Krause, M. J.	2022	Ecol. Indicators	Using fish community and population indicators to	Stream biological	Random forest	Using community and species-level analyses concurrently to provide a more holistic
51	Wherry, S. A., A	Wherry, S. A.	USGS	A. J. Tesoriero and S.	2021	Environ. Sci. Technol.	Factors Affecting Nitrate Concentrations in Stream Base	Baseflow nitrate	Boosted regression trees (BRTs)	Linking nitrate concentrations in base flow in the Bay watershed to explanatory variables
52	Loftis, J. D., S. K	Loftis, J. D.	VIMS	S. Katragadda and Ieee	2022	OCEANS Hampton Roads	A Deep Learning Algorithmic Approach to Develop a	Inundation	Deep learning (no access to full text)	Developing Video Inundation Monitoring Systems in tidal tributaries of Chesapeake Bay
53	Shen, J., Z. Wan	Shen, J.	VIMS	Z. Wang, J. Du, Y. J.	2024	Earth and Space Science	Machine Learning-Based Wave Model With High Spatial	Wave	LSTM	Developing a machine learning model using long short-term memory to simulate large-
54	Yu, X., J. Shen a	Yu, X.	VIMS	J. Shen and J. Du	2020	Water Resour. Res.	A machine-learning-based model for water quality in	Coastal DO	Artificial Neural networks; Empirical	Understanding temporal-spatial variations of DO and hypoxic condition in Chesapeake
55	Yu, X. and J. Sh	Yu, X.	VIMS	J. Shen	2021	Ocean Model.	A data-driven approach to simulate the spatiotemporal	Coastal Chl-a	Artificial Neural networks; Empirical	Understanding spatiotemporal variability of Chl-a in the Chesapeake Bay using satellite
56	Yu, X., J. Shen, (	Yu, X.	VIMS	J. Shen, G. Zheng and J.	2022	Ocean Model.	Chlorophyll-a in Chesapeake Bay based on VIIRS	Coastal Chl-a	Artificial Neural networks; Empirical	Understanding spatiotemporal variability of Chl-a in the Chesapeake Bay using satellite
57	Lee, J.-W., J. L. I	Lee, J.-W.	Virginia Tech	J. L. Irish, M. T. Bensi	2021	Coastal Engineering	Rapid prediction of peak storm surge from tropical	Storm surge	Neural networks; PCA; K-means	A new one-dimensional convolutional neural network model combined with principal
58	Merriam, E. R., J	Merriam, E. R.	West Virginia University	J. T. Petty and J.	2019	Ecosphere	Conservation planning at the intersection of landscape	Brook trout	Boosted regression trees (BRTs)	Using boosted regression tree to predict brook trout occurrence at the stream reach scale
59	Zhang, Z., J. Hu	Zhang, Z.	Xiamen University, China /	J. Huang, S. Duan, Y.	2022	Ecol. Indicators	Use of interpretable machine learning to identify the	Riverine water quality	Random forest; Shapley Additive	Understanding how riverine nutrient export responds to the land use gradient in the
* The following were added based on participant feedback after 11/2024 .										
60	Zheng, G., S. Sc	Zheng, G.	NOAA / University of	S. Schollaert Uz, P. St-	2024	Artificial Intelligence for	Hypoxia Forecasting for Chesapeake Bay Using Artificial	Coastal DO	Deep neural network	Predicting oxygen concentrations in the Chesapeake Bay using deep neural networks
61	Gordon, S., D. I	Gordon, S.	USGS	D. K. Jones, V. S. Blazer,	2021	Environ. Monit. Assess.	Modeling estrogenic activity in streams throughout the	Endocrine-disrupting	Random forest	Using random forest regression models to predict estrogenic activity at unsampled
62	Cashman, M. J.	Cashman, M. J.	USGS	G. Lee, L. E. Staub, M. P.	2024	J. Environ. Manage.	Physical habitat is more than a sediment issue: A multi-	Physical habitat metrics	Random forest	Using rapid habitat monitoring data to train a machine-learning (i.e., random forest)
63	Fanelli, R. M., J.	Fanelli, R. M.	USGS	J. Moore, C. C. Stillwell,	2024	ACS ES&T Water	Predictive Modeling Reveals Elevated Conductivity	Riverine conductivity	Random forest	Predicting conductivity in streams across the Chesapeake Bay watershed using random
64	Woods, T., M. C	Woods, T.	USGS	M. C. Freeman, K. P.	2023	Glob. Chang. Biol.	Observed and projected functional reorganization of	Fish biodiversity	Random forest; clustering	Assessing ongoing and future fish biodiversity responses to climate and LULC in the
65	Emmons, S., T.	Emmons, S.	USGS	T. Woods, M. Cashman,	2024	J. Environ. Manage	Causal inference approaches reveal both positive and	Stream biological	Propensity score matching (PSM);	Using two causal inference approaches to identify potential management practice (MP)
66	Wang, N., Q. Ch	Wang, N.	Northeastern University	Q. Chen, H. Wang, W.	2023	Applied Ocean Research	Field observations and long short-term memory	Wave energy	LSTM	Focusing on the prediction of wave energy spectra in shallow water using winds and tides
67	Lee, J., D. Kim, S	Lee, J.	University of Maryland,	D. Kim, S. Hong, D. Yun,	2024	Sci. Total Environ.	Comparative efficiency of the SWAT model and a deep	Riverine nitrogen	LSTM; Convolutional neural network	Performance of SWAT and deep learning model was evaluated in cases when deep
68	King, R. S., A. H	King, R. S.	SERC	A. H. Hines, F. D. Craige	2005	J. Exp. Mar. Biol. Ecol.	Regional, watershed and local correlates of blue crab	Blue crab and bivalve	CART	Testing hypothesized relationships of blue crabs and bivalves with salinity (a regional
69	Fisher, G. B., A.	Fisher, G. B.	UCSB	A. J. Elmore, M. C. Fitzpat	2024	GIScience & Remote Sensi	Mapping recent timber harvest activity in a temperate fores	Timber harvest	XGBoost	Providing a framework and validation for combining approachable machine-learning
70	Mainali, K., M. E	Mainali, K.	Chesapeake Conservancy	M. Evans, D. Saavedra,	2023	Sci. Total Environ.	Convolutional neural network for high-resolution wetland	Wetland mapping	CNN	Developing a deep learning model to automatically map wetlands at landscape scale in
71	Kulkarni, C., B.	Kulkarni, C.	University of Maryland	B. A. Tama, N. -J.	2025	IEEE Journal of Selected	Anomaly Detection Using Graph Deviation Networks With	Anomaly detection	Graph Deviation Networks	Providing a framework to analyze disparate datasets by forming spatial neighborhoods to
72	Devnath, M. K.,	Devnath, M. K.	University of Maryland	K. S. Chakraborty, and	2024	Proceedings of the 32nd	CMAD: Advancing Understanding of Geospatial Clusters	Anomaly detection	CMAD	Presenting a novel method named Convolution Matrix Anomaly Detection (CMAD),
73	McGuire, M. P.,	McGuire, M.	University of Maryland	V. P. Janeja, and A.	2010	Knowledge Discovery	Spatiotemporal Neighborhood Discovery for Sensor Data	Spatiotemporal	Graph-based neighborhood	Focusing on the discovery of spatiotemporal neighborhoods in sensor datasets where a
74	Zhang, Q., R. R.	Zhang, Q.	UMCES / CBP	R. R. Murphy, R. Tian,	2025	Environ. Sci. Technol.	Geography, Trajectories, and Controls of Coastal Water Q	Water quality	Random forest	Combining long-term monitoring data, science-based assessment methods, and novel