Chesapeake Community Research Symposium 2024

Session 14: Modern Research Innovations Harnessing Big Data, Machine Learning, and Remote Sensing for Advanced Estuarine Ecosystem Modeling and Monitoring

Kim Van Meter, Shuyu Chang

"Leveraging Machine Learning for Predictive Modeling of River Temperatures across the Chesapeake Bay Watershed: Assessing the Impacts of Changing Land Cover in a Changing Climate"

Under changing land use and a warming climate, river temperatures across the Chesapeake Bay Watershed have been increasing over the last 50 years, negatively affecting aquatic ecosystems. This study introduces a novel machine learning-based approach to model and predict river temperatures in Chesapeake Bay rivers, focusing on the interplay between land cover, tree cover, and a range of geomorphic and climatic variables. Specifically, we employ an ensemble of machine learning models, including Random Forest and Neural Network approaches, based on historical river temperature data alongside a large set of predictor variables. This integrated approach allows for the capture of complex nonlinear relationships and interactions between key predictors and river temperatures. The primary objective of this work is to simulate future changes in river temperature under various scenarios of land cover change and future climate trajectories. We specifically examine the sensitivity of river temperatures to deforestation and urbanization, as well as the potential mitigating effects of reforestation and green infrastructure development. Our results highlight critical hotspots within the Chesapeake Bay watershed where river temperatures are most susceptible to land use and tree-cover change. These findings are crucial for developing targeted strategies to mitigate the impacts of climate change on river ecosystems and to preserve the biodiversity and ecological integrity of the Chesapeake Bay Watershed.

Jon Derek Loftis

Enhanced River Stage Detection Using a Deep Learning Algorithm Combining AI and Edge Detection

Enhancements in modern image edge detection and machine learning algorithms has made it technologically feasible to create an effective video inundation monitoring system using passive remote sensing web cameras. To achieve this goal, the USGS Next Generation Water Observing System has funded a collaboration with the Virginia Institute for Marine Science and the City of Virginia Beach to assess both hardware and software models capable of accurate stage detection (within 0.1 in.; 0.254 cm) from near-real time image inputs. In this project, newly deployed fixed-mounted video cameras were used to send image data to train a machine learning algorithm, and then that algorithm was successfully field-tested in tidal tributaries of Chesapeake Bay over 6 months in 2022-2023.

Video cameras were acquired and installed for the purpose of capturing static images, which were collected at 15-minute intervals over a three-month period during the summer of 2022. Through statistical analysis of these training samples, it was determined that these sensors are

highly effective in continuously measuring surface water levels, demonstrating a minimal RMSE deviation of less than 0.5 in. (1.27 cm) when positioned within 30 ft. (9.144 m) of the intended monitoring target area. The accuracy of these measurements were determined to inversely scale with distance from the monitored target, and water levels recorded via video and pictometry data were cross-verified using USGS Ka-band radar active remote sensors at each monitoring site. Several camera hardware models were evaluated, each recording in 4k resolution and infrared imaging to facilitate accurate nighttime observations.

As a result of this project, a novel passive remote sensing technology was refined and demonstrated, showcasing its ability to detect and extract water levels from time-lapsed images. Furthermore, the outcomes were seamlessly integrated into the USGS' NWISweb platform, enabling both internal and public access to the data in real-time.

John Hammond, Jeremy Diaz, Phillip Goodling, Aaron Heldmyer, Roy Sando

Forecasting the ecological impacts of hydrological drought in the Chesapeake Bay Watershed: Strategies for linking forecasted streamflow and groundwater conditions with potential biological and ecological responses

Hydrological droughts, characterized by abnormally low streamflow and/or groundwater conditions, have occurred in the Chesapeake Bay Watershed every 10 to 15 years over the past several decades, with two particularly impactful droughts occurring since 2000. Given the importance of groundwater withdrawals to meeting regional water demands, information on both streamflow and groundwater conditions are needed to properly assess and forecast drought conditions and their impacts. Hydrological droughts in the region can alter freshwater inflows into the bay, disrupting the delicate balance between saltwater and freshwater habitats. Furthermore, diminished water quality, intensified by elevated temperatures and decreased dilution, can trigger harmful algal blooms and oxygen-deprived dead zones. These ecological disruptions have cascading effects on the bay's inhabitants, including fish, crabs, and oysters, potentially impacting their reproductive success, migration patterns, and overall population dynamics. Managing and mitigating the impacts of drought on the Chesapeake Bay ecosystem requires understanding the relations between hydrological drought and ecosystem response, as these relations are necessary to forecast ecological responses based on predicted drought information for sub-seasonal to seasonal time scales. In this presentation, we demonstrate preliminary results from efforts to develop streamflow and groundwater drought onset, duration and intensity forecasts using machine learning models trained on data from across the conterminous U.S., and evaluate the performance of these models across the Chesapeake Bay Watershed and adjacent areas. Along with predictions for each drought metric, these models provide ranges of uncertainty for predictions that can be used to evaluate the likelihood of potential drought impacts. We then suggest ways that these drought forecasts could be informed by historical relations between drought and ecological responses to forecast potential ecosystem impacts to assist decision support.

Jian Shen

Machine Learning-based Wave Model with High Spatial Resolution in Chesapeake Bay

Accurately modeling sediment and organic matter transport requires using a high-resolution numerical wave model (NWM) in a water quality model. However, the computational expenses of running an NWM directly coupled with an ecosystem model pose significant challenges. The numerical wave model requires high-resolution grids, detailed wind field data, and precise boundary conditions. Conducting backward simulations without sufficient spatial forcing data is a challenge. We conducted a study using a machine learning approach based on Long Short-Term Memory (LSTM) to replicate the results of an NWM. The data-driven wave model (DWM) was trained with the NWM outputs and wind data from nine locations. The trained DWM can replicate the NWM results of daily mean significant wave height and period, and daily maximum waver in Chesapeake Bay with identical spatial resolution and the same predictive skill with a root-mean-square error of less than 6 cm and 1 second, respectively for the model domain. The model successfully forecasted daily mean significant wave height and period at NOAA wave stations, showing great forward prediction capability. Because the DWM only relied on fewer wind data from selected locations and had a short runtime, it is convenient to use to simulate wave climate changes under different scenarios. The results indicate that machine learning is promising as an alternative approach to NWM for wave forecasting. Moreover, it can serve as an alternative wave model and be coupled to ecological models for simulating sediment transport, light attenuation, and resuspension of organic materials.

Andrew Muller, Diana Lynn Muller

Creating a Long-Term Climatologically Based Forecast for Hypoxia in the Chesapeake Bay

The modeling and prediction of hypoxia in oceans, estuaries, and coastal zones are essential to understanding these habitats' long-term sustainability and resiliency for ecosystem health. Progress has been made in modeling oxygen dynamics and their short-term predictions. However, there needs to be long-term forecasting of hypoxia that incorporates climatological effects. These include El Nino Southern Oscillation (ENSO), North Atlantic Oscillation (NAO), cross-bay winds, river flows, and Gulf Stream position. Annual summer-time hypoxic volume (<2 mg/l) was determined using 29 years of dissolved oxygen measurements supplied by USEPA's Chesapeake Bay Long-Term Monitoring program. A continuous wavelet transform and coherence analysis were used to determine input variables to build a feed-forward neural network model (Artificial Intelligence Model) of hypoxic volume. The non-linear neural network model (NARX) developed could predict future hypoxic volume in the Chesapeake Bay with a means squared error of 0.5 kg/m3. Long-term future predictions of hypoxia can be used as a climatological benchmark to evaluate ecosystem health. This is the only method that uses the continuous wavelet transform to determine input variables for the Artificial Intelligence Model.

Carl Friedrichs, Dave Parrish, Chris Patrick, Willy Reay

Exploring Relationships Among and Controls on Estuarine Water Quality Parameters Using Unsupervised Clustering and Structural Equation Modeling

This study utilizes unsupervised clustering methods and exploratory Structural Equation Modeling (SEM) to investigate the relationships among external forcings and water quality parameters observed for decades at dozens of estuarine sites. Data are from ~120 National Estuarine Research Reserve (NERR) System monitoring stations, including nine stations associated with the Chesapeake Bay NERRs in Maryland and Virginia. The scale of this analysis focuses on the typical state and variability of water quality at these stations rather than anomalous events or multi-year trends. Our unsupervised clustering is driven by k-means grouping applied to the results of principle component analysis. The initial components and paths in our SEM meta-model are based on expectations from the literature. However, our general approach is driven more by data exploration than specific hypothesis testing. Components and paths in our SEM are iteratively updated and applied to clusters of monitoring stations based on our data exploration results. Statistical best-fit criteria that favor parsimony will be used to determine optimal clusters and paths. Our goal is to improve our understanding of the broad similarities and differences in the controls on water quality at NERR stations and at estuaries in general – including the Chesapeake Bay – albeit under generally shallow site conditions.

Allison Dreiss, Jeremy Testa, Vyacheslav Lyubchich, Ryan Woodland, Ryan Langendorf

Modeling Impacts of Nutrient Reduction, and Warming on Benthic Forage and Hypoxia in the Chesapeake Bay

Climate change and eutrophication are important fields of study because of their potential for widespread ecological impact, especially within estuaries which provide critical habitat for species. Hypoxia, a prominent consequence of these ecosystem stressors, has significant biological and biogeochemical implications for benthic systems. However, there are few models representing hypoxia-benthos interactions in response to future nutrient loading and warming, two key drivers of hypoxia. To address this knowledge gap, we have linked newly developed statistical models that represent the biological and chemical effects of eutrophication and climate on benthic invertebrates to numerical model scenarios of estuarine biogeochemistry, including hypoxia. We ran several future change scenarios including warming, nutrient load reductions, and changes in freshwater inflow to make quantitative predictions about the distribution and changes in benthic invertebrate communities in Chesapeake Bay. Scenarios were run singly and in combination to capture the potential for interactions. We analyzed each scenario for differences in the distributions and temporal variations in oxygen, chlorophyll, particulate organic carbon, and temperature across the entire bottom area of the mainstem Chesapeake Bay and its tributaries, as well as the associated predictions of benthic biomass. Results indicate that nutrient load reduction decreases hypoxic area and hypoxic volume, increasing oxygen levels Bay-wide despite increased warming, resulting in a positive effect on benthic biomass. Focusing on one benthic group, large nutrient reductions of 75% or over were able to decrease area of vulnerability for mollusks despite a 2 and 4 degree warming (30% and 15%). Conversely, increasing nutrient loads and increased warming have the opposite effect. Further applications of this model could be used by resource management to develop spatial maps of benthic biomass, a key component of forage for important fisheries species within Chesapeake Bay, and use these maps to predict prey availability for future fishery populations.

Olivia N. Szot, Marjorie A.M. Friedrichs, Pierre St-Laurent, Aaron J. Bever, Courtney K. Harris

Mechanisms impacting variability of hypoxia onset in the Chesapeake Bay

Seasonal hypoxia in the Chesapeake Bay is exacerbated by both global climate change and local anthropogenic stressors, and presents an environmental concern as it reduces suitable habitats, alters food webs, and degrades the overall quality of the ecosystem. In the Bay, hypoxia generally begins in May and persists for several months until it terminates in September or October. However, previous studies have revealed year-to-year variability in the temporal patterns of hypoxia, with some years experiencing earlier onset than others. Several environmental factors likely contribute to the onset and development of hypoxia, including air temperature, wind patterns, wave activity, freshwater discharge, and nutrient delivery. While these factors are known to influence oxygen dynamics, the biological and physical processes are complex and interconnected, making it a challenge to understand the relative importance of each individual factor and the underlying mechanisms that contribute to changes in the year-to-year pattern of hypoxia onset. To address this, a three-dimensional, fully coupled hydrodynamic-biogeochemical numerical modeling system was used to simulate the Chesapeake Bay during 2017 as a control. Then, a series of sensitivity experiments were conducted in which various environmental factors were modified in a manner that is expected to hasten the onset of hypoxia (i.e., reduced wind speed, increased nutrient loading, increased temperature). The onset dates of the control simulation were compared to those of the sensitivity experiments to examine the relative importance of each environmental factor. Results suggest that the wind patterns, both speed and direction, have the greatest impact on the timing of hypoxia onset, while the temperature and waves have secondary effects. Terrestrial input was shown to have little influence on the timing of hypoxia onset, but had the greatest influence on the magnitude of summer hypoxia.

Vyacheslav Lyubchich, Allison Dreiss, Ryan E. Langendorf, Jeremy M. Testa, Ryan Woodland

Predictability network of oxygen concentrations in Chesapeake Bay

Globally, the extent of oxygen-depleted marine waters is growing, increasing pressure on the integrity and productivity of marine ecosystems. Coastal areas like the Chesapeake Bay are particularly vulnerable to hypoxia because their receipt of high nutrient loads fuels extensive phytoplankton productivity, which generates abundant organic matter to fuel oxygen

consumption in waters that can be strongly stratified due to persistent freshwater inputs. Coastal hypoxia affects nutrient recycling, habitat availability, and biotic community dynamics, along with the growth, reproduction, and survival of living resources. Numerical models are a powerful tool providing continuous estimates of water quality parameters, including oxygen concentrations, that impact habitat for many organisms. Although the relationship of modeled oxygen conditions across space is a predicted outcome of numerical model simulations, it is difficult to post-process the model to understand the detailed linkages between key water quality variables across space. Here we evaluate the patterns of modeled hypoxia processes in the bay, using a spatial network of daily bottom-water oxygen concentrations derived from hindcast numerical model simulations from 1986–2015 in Chesapeake Bay. Nodes of the network represent spatial cells of the numerical model (nodes are the same across the years), while directed links connecting the nodes are defined based on Granger causality (predictability) between time series of oxygen concentrations within each year (links are reevaluated each year). The changing link structure accounted for the annual reset of oxygen concentrations during the winter mixing in the bay and allowed us to correlate network properties with environmental conditions. The strongest correlations were found between the average annual hypoxic volume and the network size (number of links), and were negative, suggesting fragmentation of bay water mass connectivity during hypoxic events. This finding may result from the emergence of several independent hypoxic zones in tributaries and habitats outside the mainstem channel. This network can further be used to adjust monitoring efforts and improve predictions of other benthic conditions across the landscape of patchy hypoxic zones during large oxygen depletion events.

Vyacheslav Lyubchich, Ryan J. Woodland, Jeremy Testa, Allison Dreiss

Using machine learning to develop models of habitat suitability for a range of benthic taxa in Chesapeake Bay

Nutrient loading can over-stimulate primary production in coastal areas, leading to seasonal hypoxia that can reduce habitat and cause stress or mortality for a diversity of organisms, including key fisheries species and the prey species that constitute their forage. However, nutrient-stimulated production at the base of the food web can also benefit organisms by enhancing food availability if water quality remains suitable, making it hard to predict the effects of nutrient enrichment on food web productivity. In this study, we aimed to link model simulations of oxygen conditions and food availability to the biomass and distribution of higher trophic level consumers by explicitly modeling the biomass of a key prey resource in the Chesapeake Bay ecosystem (benthic forage). To do this, we modeled patterns of benthic community biomass from the Chesapeake Bay Program’s Long-term Benthic Monitoring Survey (1995-2019) as a function of water quality output from the Chesapeake Bay 3-D ROMS model and a novel hypoxia network model (HyNet) that captures spatially-explicit relationships in hypoxia across the bay. In testing multiple machine learning models of benthic biomass, we found that a Random Forest Model best captured patterns of biomass, and incorporating information from HyNet improved predictions of benthic forage, but unevenly for different taxa. We then used hybrid model predictions of habitat conditions under various scenarios of nutrient loading and warming to predict subsequent changes in food availability for key fisheries species. This work highlights the importance of considering a more complete food web and of modeling hypoxia as a spatially explicit process in the management of Chesapeake Bay's living resources, particularly as climatic and nutrient inputs change in the future.

Kim Van Meter, Victor Schultz, Shuyu Chang

Quantifying Groundwater Nitrate Storage in the Upper Mississippi River Basin: Implications for Chesapeake Bay Watershed management

Increases in nitrogen (N) fertilizer application, livestock densities, and human population over the last century have led to substantial increases in nitrate contamination. While increases in riverine N loads are well-documented, the total magnitude of N accumulation in groundwater remains unknown. Here we provide a first data-driven estimate of N mass accumulation in groundwater within the Upper Mississippi River Basin (UMRB), an area of intensive row-crop agriculture and the primary contributor to Gulf of Mexico hypoxia. Using approximately 49,000 groundwater nitrate well concentration values and a suite of geospatial predictors, we developed a Random Forest model to produce gridded predictions of depth-varying nitrate concentrations. Our results suggest that approximately 15 Tg of N (328 ± 167 kg-N ha^{-1}) is currently stored in UMRB groundwater recharged over the last 50 years. For context, we compare these predictions to those from a lumped statistical model, which predicts accumulation of 387 ± 133 kg-N ha^{-1}, as well as to a simple N mass balance model of the UMRB, which puts an upper bound on accumulation of approximately 1,000 kg-N ha^{-1} (1967-2017). These findings highlight the importance of considering legacy N when forecasting future water quality. Here we also explore the applicability of our modeling approach to Chesapeake Bay Watershed groundwater. Given the CBW's similar challenges with nitrogen management and its pivotal role in the health of Chesapeake Bay ecosystems, our methodological framework will offer a valuable tool for assessing the impacts of legacy N on water quality in the CBW. This analysis serves as a crucial step towards integrated watershed management practices that can address both current and historical nutrient loading, contributing to the restoration and preservation of water quality in the Chesapeake Bay and beyond.

Alexander H. Kiser, Benjamin Gressler, Lindsey Boyle, Sean Emmons, Taylor Woods, John Young, and Kelly Maloney

Updating the Biological Assessments of Non-Tidal Streams in the Chesapeake Bay Watershed: Improvements, Challenges, and Lessons Learned

Initial assessments of fish habitat and stream condition within the Chesapeake Bay were completed and made publicly available in 2022. These assessments partly satisfied needs of the Chesapeake Bay Program Fish Habitat Action team and Stream Health Working Group to determine status and trends of non-tidal waters within the region. However, these studies were conducted separately and utilized existing predictor data at different spatial scales. We will present on our preliminary updated fish habitat and stream condition assessments that both use an updated 1:24,000 scale (NHDPlus High-Resolution) spatial framework and predictor datasets. This represents a major advancement in spatial resolution and has been updated with over 1,500 variables to explore the relationships of flow, climate, cumulative land use, pesticide use, and other important metrics that influence fish habitat and stream health. Additionally, the higher resolution framework allows for the linkage of biological records previously excluded from former assessments, primarily head water streams. This updated assessment allows for greater comparison of spatial reaches and improved management utility. Our approach employs machine learning algorithms to explore and synthesize 29,009 macro-invertebrate community samples from 19,412 sample sites and 20,674 fish community records from 13,961 sites. Initial results suggest select environmental variables are useful in determining disturbed sites vs healthy sites across regions. Metrics determined to be predictive with adequate spatial and temporal coverage will be used to predict and assess conditions across unsurveyed reaches and identify conservation and restoration needs at a management relevant spatial scale.

Shuyu Y Chang, Doaa Aboelyazeed, Kamlesh Sawadekar, Digant Chavda, Chaopeng Shen, Kimberly J Van Meter

Dams, nutrients, and water quality in the Chesapeake Bay Watershed

Human-made hydro-infrastructures such as dams and reservoirs are one of the most prevalent features across the globe. Although most rivers on earth are already damned, we are still in the midst of a major boom in global dam construction. The Chesapeake Bay Watershed is home to more than 1400 dams and reservoirs, from small mill dams to the large Conowingo Dam and Reservoir on the lower Susquehanna River. Of the more than 18 million people who live in the region, approximately 75% get their drinking water from rivers and streams, most of which pass through at least one of these thousands of reservoirs. The development of algal blooms in those human-constructed water reservoirs is of particular concern not only from the perspective of aquatic habitat but also due to the frequent use of these facilities as both drinking water sources and recreational spaces. But despite both the regional and global importance of reservoirs to water quality and aquatic ecosystems, we continue to have a limited understanding of how these hydrologic structures impact the sink and flow of legacy nutrients and the induced spatial and temporal extent of algal blooms in the reservoir systems. Quantifying the magnitude, ratio, and trend of nutrients to those reservoirs is key to studying reservoir algal blooms. In this presentation, I will first present the Chesapeake Reservoirs & Dams Navigator (CRDN) database, providing valuable information beyond the National Dam Inventory. I will then present the results of our new work, in which we use both data synthesis and a process-based modeling approach to to explicitly quantify reservoir N retention along the transport pathways, considering the effects the variety of individual reservoirs, reservoir construction history, and reservoir cascading. I will also talk about our newly developed multi-objective Long-short Term Memory (LSTM) model, trained based on the continental United States with a transfer learning, to capture the temporal dynamics of nutrients (NO_3, PO_4, TP) for ungauged basins in the Chesapeake Bay.

Larry Davis

Accessible Smart GI Health Monitoring

Pilot efforts for decentralized real-time monitoring for Green Infrastructure and the co-benefits of using data for strengthening landscaping bandwidth and operations in the Baltimore and DC Metropolitan Area.