Blue Tech Internship Final Report

Vraj PankajBhai Patel

Computer Science, College of engineering

School of Marine Science and Technology (UmassD)

Coastal Systems Program (Estuary Lab)

Micheline Labrie

08/08/2025

Summary

During my summer 2025 internship at SMAST's Estuary Lab, I successfully managed and automated the analysis of 38 years of estuary water quality data for Massachusetts environmental monitoring. My primary responsibilities included digitizing historical data from 1987-1994 that previously existed only as scanned documents, standardizing data formats across multiple decades, and developing comprehensive automation tools for ongoing data analysis. Key accomplishments included creating a complete database system, implementing PowerQuery ETL automation that reduced report generation from weeks to minutes, and building four interactive dashboards including an outlier detection system for quality control. This experience provided invaluable hands-on learning in environmental data science while directly supporting Massachusetts state government coastal management decisions. The internship significantly enhanced my understanding of the intersection between computer science and environmental research, particularly within the blue economy sector focused on sustainable coastal and marine resource management.

Internship Project Overview

Project Description

The Massachusetts Estuaries Project has been monitoring water quality across seven estuaries in southeastern Massachusetts since 1987, collecting critical data that directly supports coastal management throughout the region. This long-term monitoring program generates essential information used by state agencies for ecosystem protection efforts.

My internship project centered on managing and automating the analysis of this extensive 38-year dataset to support ongoing environmental monitoring and research initiatives. The project's primary goal was to transform decades of scattered information - including early paper records, scanned documents, and various digital files - into one clear, organized system. Additionally, I worked to develop automated tools for continuous environmental assessment and reporting that would streamline the labor-intensive process of generating regular reports for state environmental agencies.

My Role & Contributions

My specific contributions to this long-term research initiative were substantial and measurable:

- **Historical Data Recovery:** Digitized and standardized non-digital data from 1987-1994, representing approximately 8 years of monitoring data that previously existed only as scanned paper documents
- **Database Standardization:** Created estuary-specific worksheets for all 7 monitoring locations, ensuring consistent data formats across 38+ years of collection
- **Automation Development:** Programmed comprehensive PowerQuery ETL tools that automate complex environmental calculations and data processing workflows

- **Scripting Implementation:** Developed VBA scripts enabling one-click report generation and automated data distribution to specific estuary data sheets
- **Dashboard Creation:** Built 4 interactive dashboards including a sophisticated outlier detection system for data quality control and 3 comprehensive analysis dashboards for historical trend visualization (Dashboards)
- **Quality Control Systems:** Implemented systematic validation protocols ensuring data integrity across multiple parameters and decades of collection

Methodology

The technical approach required systematic handling of complex environmental data through multiple stages:

Data Recovery and Standardization: I began by examining historical scanned documents from the project's earliest years, developing consistent digitization protocols to convert paper-based measurements into standardized digital formats. This process required careful attention to maintaining scientific accuracy while establishing uniform data structures. The data was extracted using OCR (Optical Character Recognition).

ETL Development: Using PowerQuery, I created Extract, Transform, and Load processes that automatically calculate derived parameters from raw measurements. The system handles complex environmental calculations including nitrogen speciation and phosphorus cycling while maintaining traceability to original measurements.

Automation Programming: VBA scripting enabled the creation of automated workflows that generate standardized reports that are later on sent to Massachusetts state environmental agencies. These scripts handle data validation and formatted output generation with minimal manual intervention.

Dashboard Architecture: I designed interactive visualization systems using advanced Excel features to create user-friendly interfaces for data analysis. The dashboards incorporate statistical outlier detection, temporal trend analysis, and comparative assessments across multiple estuary locations.

Outcomes/Deliverables

The project produced several key deliverables with significant operational impact:

Complete Digital Database: A comprehensive, standardized database containing 38 years of estuary monitoring data, transforming previously inaccessible historical information into an analyzable digital resource.

Automated Reporting System: Reduced report generation time from several weeks of manual work to approximately 10 minutes of automated processing, representing a dramatic improvement in operational efficiency.

Quality Control Infrastructure: An outlier detection dashboard that automatically identifies potentially human data entry errors, enabling rapid quality assessment and data validation.

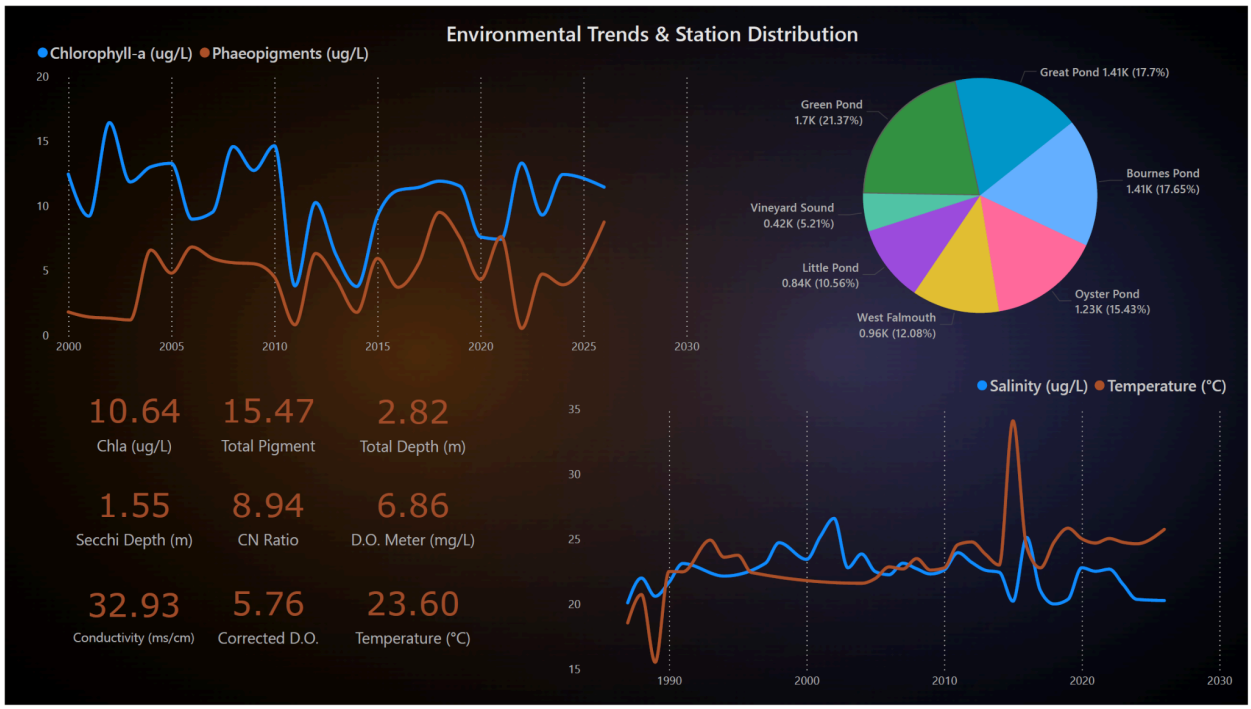
Analytical Tools: Three comprehensive dashboards providing historical trend analysis, inter-estuary comparisons, and environmental parameter correlations, directly supporting research and policy applications.

Documentation and Training Materials: Complete documentation of all automated systems, ensuring continuity and enabling future interns or staff to maintain and enhance the infrastructure. (Documentation)

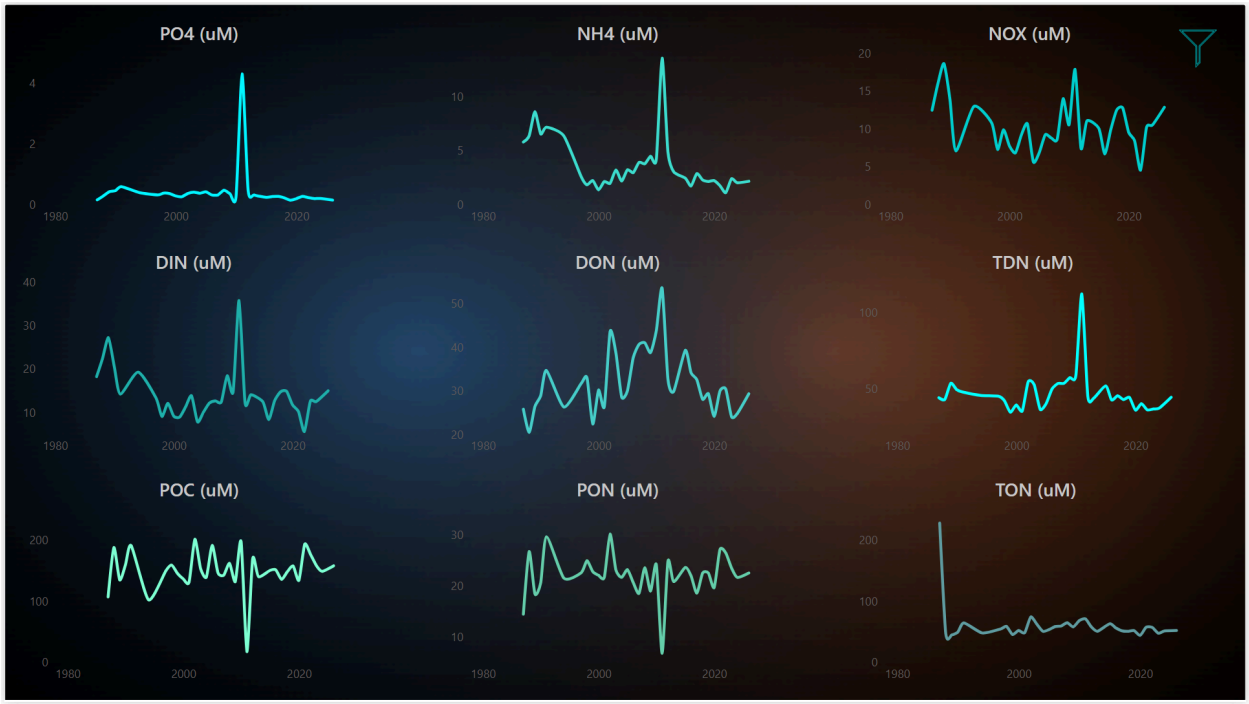
Dashboards Created:

Overall Water Quality Analysis Dashboard

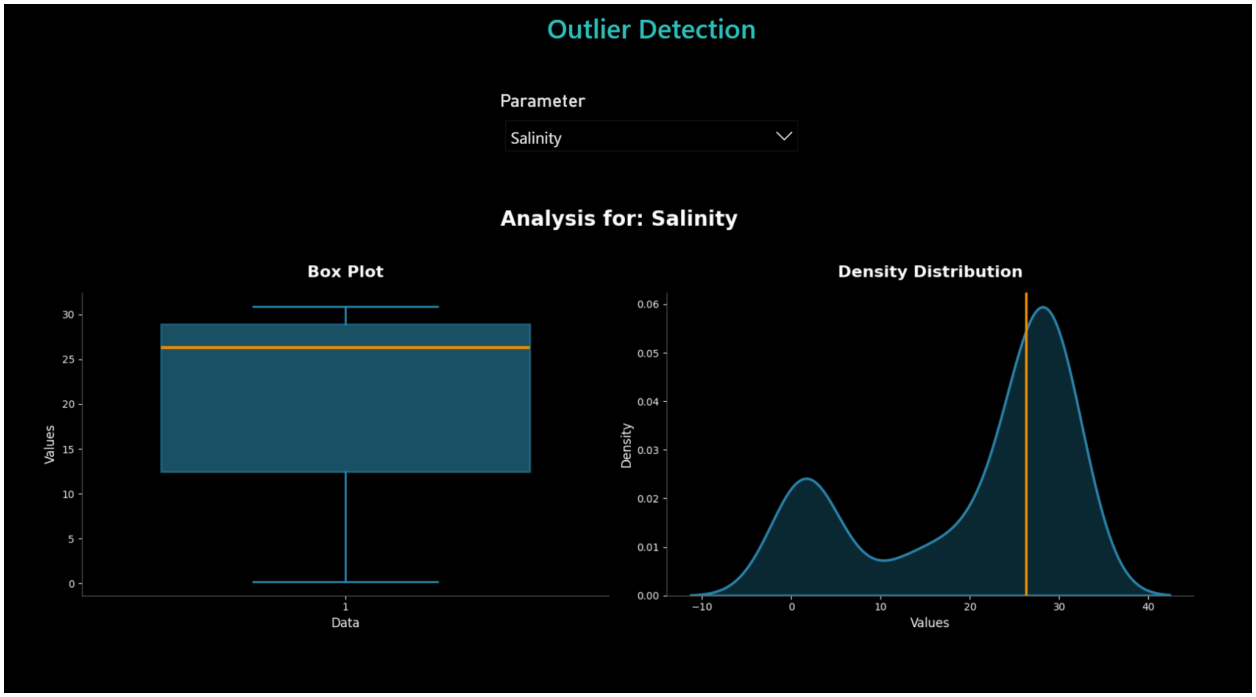
Environmental Trends and Station Distribution Dashboard



Chemical Parameter Dashboard



Outliers Detection Dashboard



Learning Outcomes and Skill Development

Specific Skills Gained

Advanced Excel Programming and Database Management: I developed sophisticated proficiency in VBA scripting and PowerQuery ETL development, learning to create complex automated systems that handle large-scale environmental datasets. This included mastering advanced Excel functions, creating user-defined functions, and implementing error-handling protocols for robust data processing.

Environmental Data Analysis: Gained comprehensive understanding of estuary chemistry parameters including dissolved inorganic nitrogen (DIN), dissolved organic nitrogen (DON), particulate organic nitrogen (PON), total nitrogen, orthophosphate, and chlorophyll-a measurements. I learned to interpret these parameters within the context of coastal ecosystem health and pollution assessment.

Data Visualization and Dashboard Development: Learning advanced PowerBI along with deploying the dashboard. This also included learning principles of effective dashboard design, statistical visualization, and user interface development for scientific applications.

Quality Control and Validation Techniques: Learned systematic approaches to environmental data validation, including statistical outlier detection, cross-parameter validation, and temporal consistency checking. These skills are crucial for maintaining data integrity in long-term environmental monitoring programs.

Connection to Academic Learning

This internship provided my first opportunity to apply the basic programming concepts from my introductory computer science courses to a real-world project. The fundamental programming principles I learned - like using loops, conditional statements, and basic functions - became essential when writing VBA scripts to automate data processing tasks. Simple concepts like iterating through data rows and implementing if-then logic that seemed abstract in class suddenly became practical tools for handling thousands of environmental measurements.

While my intro CS courses focused on general programming concepts, this internship showed me how those same building blocks could be applied to solve specific scientific problems. Learning to break down complex tasks into smaller, manageable steps - a key skill emphasized in my coursework - proved crucial when figuring out how to systematically digitize and organize 38 years of estuary data. This was my first experience seeing how the logical thinking and problem-solving approaches taught in introductory computer science translate to addressing real environmental challenges beyond traditional programming assignments.

New Knowledge

The internship provided a comprehensive understanding of estuary ecosystem dynamics and pollution cycles. I learned how human waste processing through septic systems contributes to eutrophication through groundwater contamination with ammonium (NH_4) and phosphate (PO_4), ultimately impacting estuary water quality. Understanding these environmental processes was essential for developing appropriate data analysis tools and interpreting monitoring results.

I also gained insight into the broader context of environmental monitoring, including how long-term datasets support policy decisions, coastal management strategies, and scientific

research. This knowledge of science-policy interfaces proved crucial for understanding the real-world impact of technical work and the importance of data quality and accessibility.

Personal Growth

This experience contributed significantly to my professional development by providing exposure to scientific research environments and real-world applications of computer science skills. Working with data that directly influences Massachusetts environmental policy decisions added meaningful context to technical work and emphasized the importance of accuracy and reliability in automated systems.

The internship also enhanced my ability to work independently on complex, long-term projects while collaborating effectively with research staff. Managing a project with 38 years of historical data required careful planning, systematic approach, and attention to detail that will benefit future academic and professional endeavors.

Challenges and Growth

Significant Challenges

Inconsistent Historical Data Formats: The most substantial challenge involved working with historical data from 1987-1994 that existed only as scanned paper documents with varying formats, measurement units, and recording conventions. The lack of standardization across early years meant that simple automated conversion was impossible, requiring manual verification and custom processing for each time period.

Complex Automation Requirements: Creating automation systems that could handle the intricate calculations required for environmental parameter analysis while maintaining scientific accuracy across seven different estuaries presented significant technical challenges. Each estuary has unique characteristics requiring location-specific adjustments to calculations, and the system needed to accommodate multiple data types, measurement frequencies, and quality control requirements while remaining user-friendly for laboratory staff with varying technical expertise.

Strategies for Overcoming

Iterative Development and Testing: For the automation challenges, I employed an iterative development approach, creating modular components that could be individually tested and validated before integration. I worked closely with laboratory staff to understand their workflows and requirements, incorporating their feedback throughout the development process. This collaborative approach ensured that the final system met both technical specifications and practical usability requirements.

Quality Control Integration: I embedded multiple quality control checkpoints throughout the automated system, including parameter range validation, temporal consistency checks, and statistical outlier detection. These built-in safeguards help maintain data quality while allowing the system to handle complex calculations automatically.

Areas for Improvement

Reflecting on this experience,I recognize several areas where I could have performed better and opportunities for further development. Enhanced statistical analysis capabilities would have strengthened the analytical tools, particularly in areas of trend analysis and predictive modeling. While I successfully created functional dashboards, more sophisticated visualization techniques

by asking each member of this lab about what visualizations they might be looking for and interactive features could have improved experience and analytical capabilities.

Additionally, I could have benefited from deeper understanding of environmental statistics and time series analysis methods that are specifically relevant to long-term monitoring data. Future development in these areas would enhance my ability to contribute to environmental data science applications and improve the analytical value of automated systems.

Future Impact and Career Relevance

Career Goals

This internship has significantly influenced my career aspirations by demonstrating the powerful intersection of computer science and environmental research. The experience has sparked strong interest in environmental data science and scientific computing applications, particularly within the blue economy sector. I am now considering specializing in applications that combine technological innovation with ocean and coastal research, recognizing the critical importance of data-driven approaches to environmental challenges.

The hands-on experience with real-world environmental data that directly impacts policy decisions has shown me how technical skills can contribute to meaningful environmental protection and sustainable resource management. This perspective has clarified my professional interests toward applications that have tangible environmental and social impact.

Future Plans

I plan to apply the knowledge and skills gained during this internship in several ways throughout my continued academic pursuits and future career development. The automation and

dashboard creation skills will directly support advanced coursework in computer science, particularly in areas of data management and software engineering. I might consider pursuing graduate studies in environmental informatics or marine science applications, where this foundation in environmental data analysis would prove invaluable.

Career opportunities in environmental consulting, coastal management, or marine technology development now represent attractive paths that combine my technical interests with environmental impact. The understanding of science-policy interfaces gained through this experience will be valuable regardless of specific career direction, as it has shown me how technical work translates to real-world decision-making.

Blue Economy Relevance

This internship directly supports the blue economy through its contribution to coastal and marine resource management, which is fundamental to sustainable coastal development and science-based ocean policy. The estuary monitoring data I helped organize and automate directly helps the researchers at Estuary lab, supporting sustainable development practices that balance economic activity with environmental protection.

The skills I developed are highly applicable to future blue economy opportunities including ocean data science, coastal zone management, and ocean technology applications. Understanding the intersection of data science and marine environmental research positions me well for contributing to the growing blue economy sector that depends on sustainable use of marine and coastal resources. Data science could be especially helpful since in this era, data is the key to everything. And understanding the data via the skills learned by data science courseworks would be very helpful.

Recommendations

Based on my experience, I recommend several enhancements for the SMAST estuary lab and suggestions for future interns. The laboratory would benefit from expanding automation capabilities to include statistical trend analysis and predictive modeling features that could enhance the analytical value of the long-term dataset. Also, developing a OCR model that can directly convert the hand written data on the sample data entry sheets that is used on the fields, and converting it directly to an excel format would be great. Although this might require training an entire model using the existing data which might take a long time to be developed.

For future interns, I recommend developing strong Excel programming skills before beginning the internship, as advanced spreadsheet capabilities are essential for environmental data management work. Familiarity with database concepts and statistical analysis would also be valuable preparation. The project scope should include sufficient time for learning environmental science concepts, as understanding the scientific context significantly enhances the value of technical contributions.

The internship program would also benefit from more structured interaction with state government partners who use the data, providing interns with greater understanding of how their technical work translates to policy applications and environmental management decisions.

Conclusion

This internship at SMAST's Estuary Lab provided an exceptional opportunity to apply computer science skills to critical environmental research with direct policy impact. The experience of transforming 38 years of estuary monitoring data into automated, accessible systems demonstrated the powerful potential of technology to support environmental protection and

sustainable resource management. The technical achievements in data automation and dashboard development represent significant operational improvements that will benefit ongoing research and environmental monitoring efforts.

Most significantly, this internship highlighted the vital intersection of computer science and environmental research within the blue economy context. The hands-on experience with real environmental challenges, from data digitization, provided an invaluable perspective on how technological innovation can address pressing coastal and marine environmental issues. The combination of technical skill development, environmental knowledge, and understanding of science-policy interfaces has prepared me for future contributions to the blue economy sector.

The internship exceeded my expectations for professional development and career clarification. It demonstrated that meaningful environmental impact can result from careful application of technical skills to complex real-world problems.

Acknowledgments

I would like to express sincere gratitude to Micheline Labrie for her outstanding supervision and mentorship throughout this internship. Her guidance in navigating the scientific complexities of estuary research was invaluable to my learning and success. I also thank the SMAST estuary lab staff for their generous support.

My special appreciation goes to Nishita Roy Pope (founder of tribe academy, Courage Builder), Peggy Dias, Geoff Cowles, Michael Marino.

My appreciation extends to the University of Massachusetts Dartmouth SMAST program for providing this exceptional internship opportunity.